

The Syntax, Semantics and Processing of Agreement and Binding Grammatical Illusions

by

Hezao Ke

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Linguistics)
in The University of Michigan
2019

Doctoral Committee:

Professor Samuel David Epstein, Co-Chair
Professor Acrisio Pires, Co-Chair
Associate Professor Jonathan Brennan
Professor Richard Lewis

“Always rejoice, unceasingly pray, in everything give thanks.”

-1 Thessalonians 5: 16-18

Hezao Ke

hezaoke@umich.edu

ORCID iD: 0000-0003-0873-0684

© Hezao Ke 2019

This thesis is dedicated to my wife and my children.

ACKNOWLEDGEMENTS

I am now with the complete dissertation in hand, which is the result of six years of work at the University of Michigan. It is time to express my gratitude to the people who have made this dissertation possible.

I should first thank my committee members: Samuel Epstein, Acrisio Pires, Richard Lewis and Jonnathan Brennan. I feel especially indebted to the co-chairs of my committee, and my advisors, Sam and Acrisio. Sam is not only a mentor who is always supportive, but also a friend who never gets tired of offering his warm encouragement and unreserved trust. Tim Chou in fact said what I want to say, “Sam is the kind of advisor and friend that everybody wants.” Sam generously provided his insightful comments on every paper I wrote. His contribution to the dissertation is central and enormous. Many of the ideas were inspired by the meetings with him (and Rick) such that it is impossible to single out his every contribution in the dissertation.

I should thank Acrisio for his valuable comments on every chapter of the dissertation. We have collaborated in several research projects and co-authored several papers. It was a delightful experience to work with him. As a supervisor, Acrisio is very patient and kind. There were times when I approached him with a paper or an abstract that was under a tight deadline, he never complained but generously offered his detailed revisions, which usually made that piece of work ready to submit. I should also thank Acrisio for his care and help on various grant applications and many other administrative issues.

This work would be impossible without the support from the other two members of

my committee, Richard Lewis and Jonathan Brennan. Rick is a great scientist: he is clear, sharp and precise in thinking. Meetings with him have been very stimulating. I greatly benefited from Rick's expertise on the cue-based retrieval theory and on psycholinguistics in general.

I should thank Jon for allowing me to participate in his lab meetings. I received much help from him regarding psycholinguistics and statistics. I enjoyed the time teaching and meeting with him. I can clearly recall that in many meetings with him, a most exciting topic was about his and my little children.

I want to thank other faculty members in the linguistic department at the University of Michigan who spent time discussing with me various topics in this dissertation: Ezra Keshet, Marlyse Baptista, Emily Atkinson and Natasha Abner. Special thanks go to Ezra for his significant contribution to the semantic component of the dissertation. I should also thank Patrice Beddor, San Duanmu, Julie Bolland, Steve Abney, Savithry Namboodiripad, Robin Queen, and Andries Coetzee for their help and support at different points of my studies at Michigan. Thanks also go to staff members Jennifer Nguyen, Sandie Petee and Talisha Reviere-Winston for their assistance.

This dissertation would be not possible without the generous help and support from my fellow students and graduates from the University of Michigan. I would especially like to mention Andrew McInnerney, Jae-Young Shim, Sujeewa Hettiarachchi, Yushi Sugimoto, Marcus Berger, Pyeong Cho, Chia-Wen Lo, Lucy Chiang, Danielle Burgess, Joy Peltier, Aliaksei Akimenka, Fahad Alrashed, Dong Yan, Jian Zhu, Peter Lindes, Marjorie Herbert, Will Nediger, Tim Chou, David Medeiros, and Miki Obata. Thank you for making my time at the University of Michigan such an enjoyable and memorable experience! I am especially grateful to Andrew for his proofreading this dissertation and several of my papers.

I want to thank friends and researchers who are not at the University of Michigan for their discussion and comments on different parts of the dissertation: Lin Yu,

Zahra Mirrazi, Yasutada Sudo, Giorgos Spathas, Quan Wan and Hisatsugu Kitahara. I would also like to extend my gratitude to my church friends for their support, even reading my work, especially Caroline, Ryan, Alex, EJ and Ching-shih.

For my linguistic training in China, I would like to thank my MA advisor Liquan Gao and his collaborator Stephen Crain for their training in child language acquisition. Fuzhen Si's and James Huang's classes triggered my interest in syntax: I was fascinated by the beauty of syntax. I also thank my friends Peng Zhou, Jie Ren, Zheng Shen, Haiwei Zhang and Aijun Huang for their help, support and friendship.

Finally, and most importantly, I would like to thank my wife for her unconditional, ever-lasting and never-failing love. Without her love and support, the Ph.D. life would not have been this smooth and enjoyable. She is certainly always the first person I would like to share my every achievement (if there is any) and every failure with. I also thank my sons, Benjamin and Samuel, for all the joy they bring to our family. I am indebted to my mother-in-law: without her help, I would not have been able to focus on writing and finish this dissertation in time. Many thanks to my parents and other family members in China for their constant love and support. I should also thank members of my church in China and the US for their care, friendship and encouragement.

TABLE OF CONTENTS

DEDICATION	ii
ACKNOWLEDGEMENTS	iii
LIST OF FIGURES	x
LIST OF TABLES	xi
ABSTRACT	xiii
CHAPTER	
I. Introduction	1
1.1 Illusions of grammaticality	2
1.2 Challenges of the cue-based retrieval model	5
1.2.1 Challenge One	5
1.2.2 Challenge Two	10
1.3 Solutions to the challenges	11
1.3.1 For Challenge One	11
1.3.2 For Challenge Two	13
1.4 The framework	16
1.4.1 The language model	16
1.4.2 The <i>v</i> P-internal subject hypothesis	21
1.5 Organization of the thesis	25
II. A Minimal Search-Based Analysis of Agreement	26
2.1 Introduction	26
2.2 Previous analyses of agreement	27
2.2.1 Agreement and Spec-head relations	28
2.2.2 Agree	29
2.2.3 The direction of agree	33
2.3 The proposal	41

2.3.1	Theoretical motivation	41
2.3.2	Defining minimal search	42
2.3.3	An example	46
2.3.4	Breadth-first or depth-first?	47
2.3.5	Minimal search-based agree and labeling	50
2.4	Theoretical implications	56
2.5	Empirical consequences	64
2.5.1	Upward agree	65
2.5.2	Cyclic agree	69
2.5.3	Subject-complementizer agreement	80
2.6	Comparison to the upward agree analysis	87
III.	A Minimal Search-Based Analysis of Reflexive Binding . . .	92
3.1	Introduction	92
3.2	Previous analyses	93
3.2.1	The classical binding theory	93
3.2.2	The movement approach	94
3.2.3	The agree approach	98
3.3	Current analysis	106
3.4	Implications	115
3.4.1	The non-complementary distribution	115
3.4.2	The complementary distribution	117
3.5	A remaining issue	124
IV.	The Interpretability of Φ-Features	132
4.1	Introduction	132
4.1.1	The feature system	134
4.1.2	Semantic ϕ -features as presuppositions	136
4.2	A challenge: the PF Feature Transmission Approach	138
4.3	Challenges to the challenge	141
4.3.1	Evidence	141
4.3.2	Possible responses to the challenges	143
4.4	New evidence for the presupposition account	148
4.4.1	The syntax and semantics of collective nouns	148
4.4.2	Argument One	150
4.4.3	Argument Two	153
4.4.4	Argument Three	159
4.5	Toward an explanation	173
4.5.1	Rooth's theory of focus	173
4.5.2	Previous solutions	176
4.5.3	Current solution	177
4.5.4	Addressing remaining questions	181
4.6	Summary	188

V. Encoding Syntactic Relations as Features	192
5.1 Introduction	192
5.2 Representational c-command	195
5.3 Previous attempts to encode c-command as a feature	197
5.3.1 An “intuitive” approach	197
5.3.2 The spine-based approach	199
5.3.3 The C-COMMAND-AS-CONTENT approach	203
5.4 Proposal: derivational c-commanded-by	205
5.4.1 Derivational c-command	205
5.4.2 Relating derivational c-command to parsing	207
5.4.3 C-commanded-by in parsing	211
5.5 Simplifying the encoding system	217
5.6 Extending to the locality feature	220
VI. Experiments	224
6.1 Introduction	224
6.1.1 Literature review	226
6.1.2 Accounting for the absence of facilitatory effects	236
6.1.3 Asymmetry of interpretability in cue-based retrieval	238
6.1.4 Predictions of the account	243
6.2 An overview of the experiments	247
6.2.1 Materials	247
6.2.2 Procedure	252
6.3 Experiment 1: judgment task with reflexive binding	253
6.3.1 Participants	254
6.3.2 Stimuli	254
6.3.3 Procedure	254
6.3.4 Results	255
6.3.5 Discussion	258
6.4 Experiment 2: speeded judgment task with reflexive binding	260
6.4.1 Participants	260
6.4.2 Stimuli	260
6.4.3 Procedure	260
6.4.4 Results and discussion	261
6.5 Experiment 3: judgment task with subject–verb agreement	265
6.5.1 Participants	266
6.5.2 Stimuli	267
6.5.3 Procedure	267
6.5.4 Results and discussion	267
6.6 Experiment 4: self-paced reading with reflexive binding	274
6.6.1 Participants	274
6.6.2 Stimuli	274

6.6.3	Procedure	275
6.6.4	Results and discussion	276
6.7	Experiment 5: self-paced reading with subject–verb agreement	281
6.7.1	Participants	282
6.7.2	Stimuli	282
6.7.3	Procedure	282
6.7.4	Results and discussion	282
6.8	Experiment 6: a replication study for reflexive binding	286
6.8.1	Participants	286
6.8.2	Stimuli	287
6.8.3	Procedure	287
6.8.4	Results and discussion	287
6.9	General discussion	298
6.10	Future direction	307
VII.	General Conclusions	309
	BIBLIOGRAPHY	316

LIST OF FIGURES

FIGURE

1.1	The Y-shape model	16
1.2	The <i>v</i> P-internal subject hypothesis	21
2.1	Agreement and the Spec-head relations	28
2.2	An illustration of minimal search	47
2.3	Breadth-first search	48
2.4	Depth-first search	48
2.5	Béjar & Rezac’s (2009) analysis of cyclic agree	72
2.6	Béjar & Rezac’s (2009) analysis of cyclic agree	76
2.7	Cycle One Agree	77
2.8	Cycle Two Agree	78
2.9	A-movement	79
2.10	\bar{A} -movement	79
3.1	Reuland’s (2011) agree analysis of SE anaphors	99
6.1	The 7 point Likert scale data from Experiment 1	255
6.2	The 7 point Likert scale data from Experiment 2	262
6.3	Three-way interaction HNoun:DNum:Accu-Rate in Experiment 2 . .	265
6.4	The 7 point Likert scale data from Experiment 3	268
6.5	three-way interaction HNoun:DNum:Accu-Rate in Experiment 3 . .	271
6.6	Reading times by region for Experiment 4	277
6.7	Reading times by region for Experiment 5	283
6.8	Reading times by region for Experiment 6	289
6.9	three-way interaction HNoun:DNum:Accu-Rate in Experiment 3 . .	297

LIST OF TABLES

TABLE

6.1	Mean acceptance rates and reaction times by condition for Experiment 1	256
6.2	LMM results for Experiment 1	258
6.3	Mean acceptance rates and reaction times by condition for Experiment 2	261
6.4	LMM results for Experiment 2	263
6.5	Planned comparisons between conditions in Experiment 2	264
6.6	Mean acceptance rates and reaction times by condition for Experiment 3	269
6.7	LMM results for Experiment 3	270
6.8	Planned comparisons for Experiment 3	270
6.9	Mean reading times of the critical and spill-over positions in Experiment 4	278
6.10	LMM results for Experiment 4 (Region 8)	279
6.11	LMM results for Experiment 4 (Region 9)	280
6.12	Planned comparisons for Region 9 in Experiment 4	281
6.13	Mean reading times of important regions in Experiment 5	284
6.14	LMM results for Experiment 5 (Region 7)	284
6.15	Planned comparisons for Region 7 in Experiment 5	285
6.16	LMM results for Experiment 5 (Region 8)	285
6.17	Planned comparisons for Region 8 in Experiment 5	286
6.18	Mean reading times of important regions in Experiment 6	290
6.19	LMM results for Experiment 6 (Region 10)	290
6.20	Planned comparisons for Region 10 in Experiment 6	291
6.21	LMM results for Experiment 6 (Region 11)	293
6.22	Planned comparisons for Region 11 in Experiment 6	293
6.23	GLMM results for the judgment data in Experiment 6	295
6.24	Planned comparisons for the judgment data in Experiment 6	296
6.25	Planned comparisons for the judgment data from the fast group in Experiment 6	298
6.26	Summary of the main findings of the experiments	299

6.27	Facilitatory effects for the low-rating and high-rating groups in judgment experiments	302
------	--	-----

ABSTRACT

The overall goal of this dissertation is to establish a linking theory between the syntax and semantics and the processing of subject–verb agreement and reflexive binding. This dissertation develops a unified syntactic analysis of agreement, based on a formalization of minimal search. Such an analysis accounts for a variety of agreement patterns observed in various languages, including negative concord in Czech, inflection doubling in Norwegian, Frisian and Swedish, multiple agree in Japanese, cyclic agree in Georgian and Hindi Urdu, and subject-complementizer agreement in Lubukusu. The minimal search-based analysis is also extended to reflexive binding. The minimal search-based analysis of subject–verb agreement and reflexive binding captures the syntactic similarity between these two constructions.

This dissertation then argues that subject–verb agreement and reflexive binding have an important representational difference: the ϕ -features involved in subject–verb agreement and reflexive binding are essentially different in their semantic content. The ϕ -features on bound reflexives (and bound variables generally) have semantic content and are semantically interpretable, whereas those on agreeing verbs/T heads are not semantically interpretable.

The syntactic and semantic analyses of agreement and reflexive binding have crucial consequences for the sentence processing study of subject–verb agreement and reflexive binding. This dissertation proposes that in cue-based retrieval, mismatches of semantically interpretable retrieval cues (in reflexive binding) are less tolerable to the parser than mismatches of phonological cues (in subject–verb agreement). Experi-

mental results are provided to evaluate this Asymmetry of Interpretability Hypothesis. The results reveal that the (un)acceptability of a target sentence by an experimental participant influences the occurrence of illusions of grammaticality/facilitatory effects for that participant: facilitatory effects do not occur for grammatical or acceptable sentences. The results also indicate that there are significant contributing factors to the emergence of illusions of grammaticality/facilitatory effects, involving relative differences between the target and distractor with respect to frequency, phonological length, orthographic length, phonological neighborhood density, and orthographic neighborhood density. These results suggest that the distractors are not completely ignored in the retrieval of the relevant target in subject–agreement and reflexive binding.

CHAPTER I

Introduction

The overall goal of this thesis is to build a linking theory between the linguistic representation and sentence processing of reflexive binding and subject–verb agreement. This is part of the effort to advance our understanding of how humans comprehend language in real-time, by developing and testing a new model of how mental representation of linguistic knowledge interacts with the memory system in sentence processing.

To achieve this overall goal, I investigate an important theory that models the interaction between memory and sentence processing, known as CUE-BASED RETRIEVAL THEORY. Cue-based retrieval theory assumes that a comprehender uses various kinds of information (e.g. syntactic, semantic, discourse, etc.) as retrieval cues to identify meaningful language chunks with matching features in working memory. This thesis emphasizes that the linguistic analysis of the structures being processed must be an important component of the cue-based retrieval theory. Specifically, this thesis develops a unified syntactic analysis of agreement and reflexive binding and defends a semantic approach to the interpretation of features involved in agreement and reflexive binding. This thesis then argues that the syntactic and semantic analyses of the structures in question provide us a critical basis to model sentence processing more accurately. This thesis further discusses two important challenges of the cue-based

retrieval theory, and provides its own solutions based on the syntactic and semantic analyses it develops.

The dissertation makes independent contributions to linguistic theory and sentence processing respectively. The syntax component of the dissertation develops a unified analysis of agreement, based on a formalization of minimal search, for various agreement/concord patterns across languages. The semantic component argues that ϕ -features borne by bound variables (including reflexives and pronouns) have semantic content and are semantically interpretable. By contrast, the ϕ -features on verbs or the T head are phonological features and are not semantically interpretable. The sentence processing component proposes that the parser is less tolerant of mismatches of semantic retrieval cues, compared to those of phonological cues. I argue that this asymmetry may underlie the interesting dissociation in grammatical illusions concerning agreement.

1.1 Illusions of grammaticality

As a window into human sentence processing mechanisms outlined in the cue-based retrieval theory, ILLUSIONS OF GRAMMATICALITY, a type of facilitatory interference, is taken as the empirical focus of this thesis. Illusions of grammaticality occur when competent language users fail to detect ungrammaticality (Clifton et al. 1999; Pearlmutter et al. 1999; Wagers et al. 2009; Phillips et al. 2011). That is, a reader/listener may perceive a sentence as well-formed despite knowing upon conscious reflection that it is an ungrammatical sentence. As a result, in the context of illusions of grammaticality, a sentence sounds acceptable to some comprehenders even though it violates certain grammatical constraints, and the sentence is thus judged grammatical. Such illusions of grammaticality can be demonstrated with sentence (1) (adopted from Dillon et al. 2013).

- (1) The new executive who oversaw the middle managers apparently were dishonest about the company’s profits.

At first glance, many readers fail to see that the linking verb *were* needs to be in the singular form to match the subject *the new executive who...* As a result, these readers do not slow down when they read *were*, whereas such a slow-down is expected if the ungrammaticality is detected.

Previous studies have shown that such illusions of grammaticality are not merely the consequence of random language user errors but rather are systematically associated with intervening material that interferes with the evaluation of grammaticality. This is supported by experimental results showing that when the intervening plural noun phrase *the middle managers* in (1) is changed to its singular form, *the middle manager*, as shown in (2) below, such illusions disappear (Dillon et al. 2013; Pearlmuter et al. 1999; Xiang et al. 2009).

- (2) The new executive who oversaw **the middle manager** apparently were dishonest about the company’s profits.

Therefore, the intervening noun phrase, which we call a distractor, and its plural number feature, must have contributed to the illusions of grammaticality observed in sentence (1).

Vasishth et al. (2008) and Wagers et al. (2009) pointed out that the above illusions of grammaticality can be explained with a cue-based retrieval model such as Lewis & Vasishth (2005). Lewis & Vasishth (2005), henceforth LV05, proposed an influential computational model that implements cue-based retrieval (also see Lewis et al. 2006; Vasishth et al. 2008). Cue-based retrieval assumes that memory retrieval in sentence comprehension is content-addressable, since items are accessed directly by their content, not their locations (McElree 1998, 2000; McElree et al. 2003). The retrieval of linguistic representations (bundles of features) is treated as an assessment of global

familiarity, or of goodness of match between retrieval cues and the feature content of the representations. Linguistic information (e.g., syntactic and semantic constraints) provides the retrieval cues that enable direct access to needed representations. Representations that match the retrieval cues obtain different levels of activation gains according to the strength and goodness of the match. The possibility of a successful retrieval of a specific representation is determined by both its activation gains from cue-match and its original base activation level.

LV05 explains illusions of grammaticality with misretrieval (Jäger et al. 2017; Wagers et al. 2009). Misretrieval occurs when a grammatically inaccessible distractor, i.e., *the middle managers* in (1), matches the [+plural] retrieval cue and is incorrectly retrieved, licensing plural agreement on *were* in (1). As a result, in (1), the comprehender mis-perceives the distractor as determining the number agreement with the verb, which leads to the mis-perception that the sentence is grammatical.¹ As Logačev & Vasishth (2015) point out, mis-retrieval of the distractor will lead to a speed-up in reading time because of a racing situation between the retrieval of the target (the correct chunk supposed to be retrieved) and the distractor (the non-target chunk that interferes with the retrieval process): The chunk which has a higher activation level, no matter whether it is the target or the distractor, will be retrieved. A higher activation level maps to a shorter retrieval time in LV05’s ACT-R model, which in turn results in a shorter reading time, all other factors being equal.

The rest of this chapter is structured as follows. Section 1.2 introduces two challenges or problems faced by the cue-based retrieval theory as is implemented in LV05. One of the challenges concerns the model’s (potentially incorrect) empirical predictions regarding illusions of grammaticality/facilitatory effects in empirical domains other than subject–verb agreement, i.e., in the domain of reflexive binding. Another

¹LV05 is based on a general cognitive architecture, Adaptive Control of Thought-Rational (ACT-R, Anderson & Lebiere 1998; Anderson 2005). ACT-R provides LV05 powerful tools to generate accurate quantitative predictions of reading time patterns in sentence processing.

challenge is a more general problem to all theories that use only features in memory retrieval, including the cue-based retrieval theory: Syntactic relations are hard to encoded as features, yet they are important for a successful retrieval. Section 1.3 provides an overview of the solutions that this thesis intends to offer to address the challenges. Section 1.4 presents a sketch of the framework as well as some of the most important concepts that this dissertation adopts. Finally, the organization of the remainder of the thesis is given at the end of this chapter.

1.2 Challenges of the cue-based retrieval model

As we mentioned in the end of the last section, the cue-based retrieval model such as LV05 provides a neat account for illusions of grammaticality. That is, misretrieval of the distractor causes readers to perceive an illusion that the agreement relation between the verb and the subject has been successfully established. However, the model is not without problems. I will identify two important challenges, to which I will present solutions in Section 1.3.

1.2.1 Challenge One

One problem noticed by psycholinguists is a surprising difference between the processing of subject–verb agreement and that of another similar type of dependency, reflexive pronoun (hereafter REFLEXIVE) binding. The results from reflexive binding have been claimed to be inconsistent with LV05’s prediction. For comparison, I repeat below in (3) the examples of subject–verb agreement. The examples in (4), taken from Dillon et al. (2013) with the conditions relabeled, can be used to illustrate LV05’s prediction concerning illusions of grammaticality in reflexive binding.²

²Notice that (4) is a structural dependency: The reflexive *themselves* needs to refer to a noun phrase that is not right next to it. Under the cue-based retrieval theory, the computational modeling of (4) is possible only when this relation can be encoded as features, another important problem that I will address later.

- (3) a. *Target-mismatch, distractor-match*

The new executive who oversaw **the middle managers** apparently were dishonest about the company's profits.

- b. *Target-mismatch, distractor-mismatch*

The new executive who oversaw **the middle manager** apparently were dishonest about the company's profits.

- (4) a. *Target-mismatch, distractor-match*

The new executive who oversaw **the middle managers** apparently doubted themselves on most major decisions.

- b. *Target-mismatch, distractor-mismatch*

The new executive who oversaw **the middle manager** apparently doubted themselves on most major decisions.

In (4), the retrieval triggered by the reflexive *themselves* has retrieval cues including [+c-command, +plural, +DP], because the antecedent of a reflexive in English must be a c-commanding determiner phrase (DP) matching the number feature on the reflexive. We ignore some other retrieval cues such as [+animate] and [+local], since they are not directly relevant here. In (4a), the distractor *the middle managers* matches the [+plural, +DP] cues, whereas the subject matches the [+c-command, +DP] cues. Since both the subject and the distractor match two retrieval cues, misretrieval of the distractor is predicted to be frequent. By contrast, in (4b), everything remains the same as (4a) except a minimal difference on the number feature of the distractor. Now the distractor in (4b) matches only the [+DP] feature; consequently, the distractor would be less possible to overtake the subject in retrieval. Under the assumption that the dependency in reflexive binding is essentially the same as that in subject-verb agreement, the LV05 model would predict an improvement of grammaticality and a speedup in the reading of the linking verb *themselves* in (4a), compared to (4b), due to the more frequent misretrieval of the syntactically inaccessible distractor,

the middle managers.

Contrary to LV05’s prediction, experimental results from reading tasks reveal that comprehenders do not exhibit facilitatory effects with (4a) (Clifton et al. 1999; Dillon et al. 2013; Jäger et al. 2017; Xiang et al. 2009, etc.) Facilitatory effects/illusions of grammaticality are therefore robustly observed in subject–verb agreement as we have seen in (1), but not with the interpretation of reflexives.³

Based on the differences between the processing of subject–verb agreement (1) and reflexive dependency/binding (4), some researchers (e.g., Phillips et al. 2011; Dillon et al. 2013) claim that reflexive binding may involve a mechanism in addition to cue-based retrieval.⁴ For example, at the site of the reflexive, the parser may initiate a linear search into syntactic objects that c-command the reflexive since syntactically the antecedent must c-command the reflexive (Chomsky 1981, 1986). This search algorithm ensures that only c-commanding DPs will be considered as potential antecedents. Given that in (4), the subject *the new executive who oversaw the middle managers* but not the distractor *the middle managers* c-commands the reflexive, the

³However, it is important to bear in mind that previous experiments in fact produced controversial and mixed results regarding the existence of illusions of grammaticality in reflexive binding. Please see a review of studies on illusions of grammaticality in reflexive binding in Chapter VI.

⁴This type of reflexive dependency is formally defined as reflexive binding. I adopt the following theory of reflexive binding from Chomsky (1981):

- (i) *Binding Condition A for reflexives*
A reflexive must be bound within its governing category (DP and CP).
- (ii) α binds β iff
 - a. α c-commands β , and
 - b. α and β are co-indexed.

And a preliminary definition of c-command would be sufficient for now:

- (iii) *Definition of c-command (to be revised in Chapter V)*
 α c-commands β iff
 - a. neither α nor β dominates the other, and
 - b. the first branching node that dominates α also dominates β .

This theory will be discussed in detail and will be revised in Chapter III. See also Chapter V for a proposal to encode syntactic relations such as the c-command relation as a feature under the cue-based retrieval theory.

syntactically irrelevant distractor will be filtered out by the search algorithm.

I argue that the above SYNTACTIC FILTER APPROACH encounters both theoretical and empirical difficulties. It needs to assume an additional structural search-based algorithm for the retrieval of the antecedent in the case of reflexive binding. Meanwhile, it still needs to resort to cue-based retrieval in other instances including subject–verb agreement. Such an explanation thus introduces extra complications to the theory. In addition, currently there is no easy way to combine the search-based algorithm directly with cue-based retrieval under the ACT-R architecture.⁵ Empirically, the syntactic filter approach predicts that a non-commanding distractor will never be misretrieved or cause intervention effects, a prediction which has been shown to be unjustified (Chen et al. 2012; Parker & Phillips 2017).

Another string of research pursues an alternative explanation, which I call the WEIGHTING APPROACH. The weighting approach suggests that the processing differences between subject–verb agreement and reflexive binding are due to the fact that the c-command relation as a syntactic retrieval cue weights heavier than other cues (Jäger et al. 2017; Parker et al. 2017). Therefore, the c-commanding subject in (4), compared to the distractor, receives a much higher activation. Consequently, the distractor can rarely be misretrieved, explaining why illusions of grammaticality are absent with (4).

This weighting approach is appealing as it does not assume an additional algorithm specifically for the retrieval of the antecedent in reflexive binding. Cue-based retrieval is a unified mechanism applied in both the retrieval of the subject in subject–verb agreement and the retrieval of the antecedent in reflexive binding. However, it is unclear why c-command should be weighted heavier. It is also unclear why the syntactic retrieval cue [+subject] that is needed for the retrieval of the subject in subject–verb agreement does not also have a heavier weight. A fundamental issue

⁵A search-based algorithm might be translated to or be implemented by a version of cue-based retrieval theory that is different from LV05 (cf. Lewandowsky & Murdock 1989; McElree et al. 2003).

concerning the weighting approach is: There is not yet a theory that can help us make accurate predictions on which retrieval cues should have a heavier weight. If the [+subject] cue is also assigned a heavier weight, then only the subject will be retrieved in subject–verb agreement, which predicts that misretrieval of the distractor will rarely happen in subject–verb agreement. Therefore, if both syntactic cues in subject–verb agreement and reflexive binding are assigned a heavier weight, we predict that illusions of grammaticality occur in none of these constructions. This prediction is the opposite to LV05’s, but it is definitely not a better prediction, as it cannot capture the fact that illusions of grammaticality are consistently observed in subject–verb agreement.

Most importantly, both the syntactic filter approach and the weighting approach implicitly assume that the parser in LV05 analyzes the sentence structures of subject–verb agreement and reflexive binding in a similar or even identical way, except that they involve different types of syntactic relations. Without this assumption, no clear expectation could be generated on the processing similarities between these two constructions. In fact, this assumption seems to be held in all previous studies on the comparison of the processing of these two constructions, as far as I am aware. I argue that this assumption mis-categorizes a syntactic **similarity** between these two constructions as a **difference** between them, and meanwhile misses an important semantic difference between these two constructions.

In Section 1.3.1, and with more details in Chapter II, III and IV, I argue that the syntactic analysis of agreement and reflexive binding can be unified with a general search algorithm for minimal search. Therefore, the syntactic relations involved in subject–verb agreement and reflexive binding are in a sense the same in nature. The syntactic filter approach and the weighting approach must assume that the two syntactic dependencies are essentially different. Therefore, given the current syntactic analysis, both the syntactic filter approach and the weighting approach are not

theoretically motivated. In addition, Chapter IV argues that the ϕ -features on the reflexive are intrinsically different from those on the verb. This representational difference has crucial implications for the parsing of these constructions. The semantic analysis in fact predicts these two constructions should be parsed in distinctive ways. In other words, with a formal linguistic analysis of these two constructions, this so-called empirical prediction problem, caused by the unexpected processing of reflexive binding, is no longer a real counterargument to the cue-based retrieval theory in general.

1.2.2 Challenge Two

Another important limitation of the ACT-R cue-based retrieval is that syntactic relations, such as the c-command relation and locality constraint in reflexive binding, are difficult to encoded with features, as required by traditional ACT-R cue-based retrieval models. ACT-R cue-based retrieval models for sentence processing depend crucially on information that is encoded in features. This is because linguistic chunks that are the target of retrieval, including words and phrases stored in working memory, are coded as bundles of features. In addition, only features can be used as retrieval cues. However, many key aspects of linguistic structure (e.g., structure dependence among different chunks) result from RELATIONS between linguistic units. Relations between units are difficult to encode in a feature-based system, as features are better suited to encoding properties of individuals, not relations between individuals. As a result, cue-based retrieval models such as LV05 is unable to properly encode structural dependencies as features. This is true especially for long-distance dependencies, including the c-command relations.

This encoding problem is a serious issue because structural dependencies are what syntax is mainly about. Although some previous studies (Alcocer & Phillips 2012; Kush 2013) have provided interesting solutions for the encoding problems associated

with syntactic relations, I will argue in Chapter V that these solutions are not satisfying as they do not have sufficient theoretical basis. I address the encoding problem by developing a new method based on the derivational approach to syntactic relations (Epstein et al. 1998).

1.3 Solutions to the challenges

1.3.1 For Challenge One

To address the first challenge regarding the (potentially incorrect) predictions of the cue-based retrieval theory in the empirical domains of reflexive binding and subject–verb agreement, this thesis argues that an explicit and formal analysis of these specific constructions needs to be in place before any meaningful predictions can be generated. This is because the model includes a parser that assumes a certain linguistic analysis of the structures under processing. Holding an informal conception of the linguistic analysis of subject–verb agreement and reflexive binding, I argue, is the reason why previous studies ignored important representational differences between them. With a more explicit linguistic analysis in mind, however, differences between the processing of these two constructions are in fact expected.

A syntactic analysis of agreement and binding is presented in Chapter II and III. This analysis suggests that the nature of the dependence relation between the antecedent and the reflexive is an agreement relation similar to that between the subject and the verb (more accurately, the T head). This component of the analysis captures the similarities between these two constructions, as both involve agreement. In subject–verb agreement, the subject and the verb agree in their ϕ -features, including the person and number features (some languages other than English also have grammatical gender agreement). In reflexive binding, the antecedent and the reflexive agree in their referential index feature (Hicks 2009). I assume that the ϕ -features

on the reflexives do not agree with those on their antecedents, as the reflexives bear independent semantically interpretable ϕ -features themselves, a position that I will defend in Chapter IV.

Despite similarities, important distinctions exist between these two constructions. This thesis argues that the ϕ -features on a reflexive are essentially different from those on a verb/T head. For example, the number feature on a reflexive (e.g., plural number on *themselves*) is different from that on a linking verb/the T head (e.g., singular or plural number features on *are* or *likes*) in its interpretability. The number (and person and grammatical gender) feature on the T head is semantically vacuous. For example, *are* or *likes* do not mean more than one thing in the cardinal sense. In this respect, the number feature on T is semantically uninterpretable. On the other hand, the number and person features on a reflexive or a noun phrase have independent meaning, and are semantically interpretable. For instance, *himself* means a single third person, and *we* usually means more than one person, with the speaker included; *cat* means one cat, but *cats* means more than one.

Given such a distinction between semantically interpretable and semantically uninterpretable features, I put forward a hypothesis regarding the asymmetry between these two types of features in parsing:

(5) *Asymmetry of Interpretability in parsing*

Mismatch of semantically interpretable retrieval cues will be more likely detected than that of phonological retrieval cues.

This hypothesis predicts that in the case of reflexive binding in (4), repeated below as (6), since the semantically interpretable number feature on the subject mismatches the semantic retrieval cue [SEM-Num: PL], and since the parser is less tolerant of such mismatches, the ungrammaticality of the sentence can be easily detected. By contrast, in subject–verb agreement in (1), repeated below as (7), only a mismatch of phonological retrieval cues occurs, which is less costly than a mismatch of semantic

retrieval cues.

- (6) The new executive who oversaw the middle managers apparently doubted **themselves** on most major decisions.
- (7) The new executive who oversaw the middle managers apparently **were** dishonest about the company’s profits.

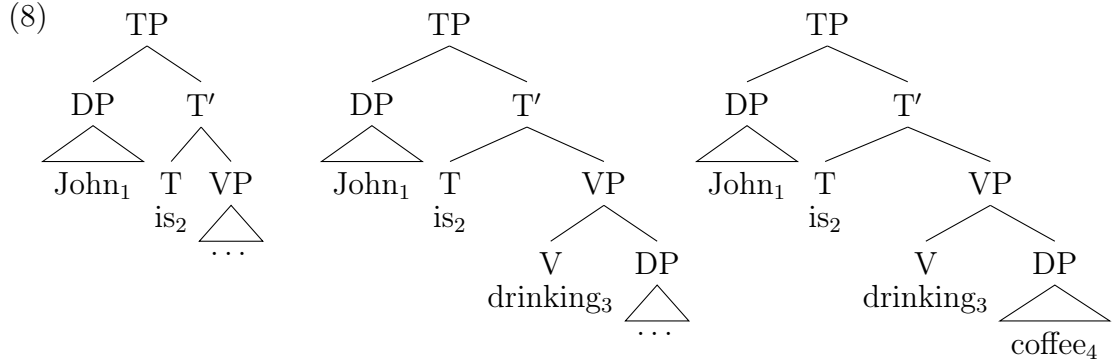
According to the Asymmetry of Interpretability Hypothesis in (5), the parser is less tolerant of the mismatch of semantically interpretable retrieval cues in reflexive binding in (6). Comparatively, the mismatch of semantically uninterpretable retrieval cues in subject–verb agreement has a higher possibility to “escape the parser’s attention.” Illusions of grammaticality are therefore correctly predicted to be less possible for reflexive binding than for subject–verb agreement.

1.3.2 For Challenge Two

Regarding the encoding of syntactic relations as features, this thesis proposes that the problem can be overcome with a revised derivational approach to syntactic relations, built upon Epstein et al.’s (1998) derivational theory. Epstein et al. (1998) formulated an influential theory of syntactic relations, including the central “c-command” relation. However, a direct implementation of their formulation fails to define the relation as a feature. I thus propose the derivational c-commanded-by approach, by changing the c-command relation to the c-commanded-by relation. The revised derivational approach, like Epstein et al. (1998), reduces all syntactic relations to two fundamental local relations: sisterhood and dominance/containment, both of which are results of merge/concatenation. For instance, Y c-commands X iff X or another syntactic object Z that contains X merges with Y. Different from Epstein et al. (1998), the revised derivational approach encodes the c-command features on **c-commandees** rather than on **c-commanders**. That is, the feature is no longer a

[C-command] feature on the c-commanders, but a [C-commanded-by] feature on the commandees: the chunk IDs of the c-commanders are stored in the [C-commanded-by] feature on the c-commandees.

In parsing, since a sentence is parsed cumulatively and the corresponding structure unfolds over time, if the c-command feature is on the c-commanders, the content of the c-command feature, which is a list of indices of c-commandees, needs to be constantly updated and is therefore a very inefficient algorithm. This problem has been discussed extensively by Kush (2013). As an illustration, let us consider the structures built in steps shown in (8), where the indices of chunks are indicated by the subscripts.



In the left structure, *John* and *is* are parsed, and *John* asymmetrically c-commands *is*. If the c-command feature is on the c-commanders, then the corresponding feature on *John* would be [C-command: 2], where “2” is the index of *is*. Then *drinking* is parsed, and the middle tree is constructed. Since both *is* and *John* c-command *drinking*, the parser would need to create a [C-command: 3] for *is*, and also go back to update *John*’s [C-command] feature by adding the index of *drinking*. Similarly, the [C-command] features on *is* and *John* will need to be updated again when the next word, *coffee*, is parsed, given that both c-command *coffee*. Therefore, when a new word is parsed, this algorithm requires all its c-commanders to have their [C-command] feature updated, by adding the index of the new word to the feature content. Imagine a situation where the sentence is very long, the c-commanders of a newly parsed word can be enormous, and the amount of corresponding update operations can dramatically increase. Such

an algorithm cannot be psycholinguistically realistic, as we do not generally observe a dramatic increase in reading times at the end of sentences, even for very long ones.

If we instead store the [C-commanded-by] feature in the c-commandees, the problem goes away. But for us to understand how the problem can be solved, an implementation of the revised derivational approach in parsing should first be in place: If two chunks X and Y join each other directly and therefore enter into a relation, the ID feature of X is passed to the c-commandee Y as the content of Y's [C-commanded-by] feature, and vice versa. This passing of [C-commanded-by] feature is a crucial step for the translation of a syntactic RELATION between X and Y into FEATURES borne individually by X and by Y. The content of the [C-commanded-by] features of X and Y can then be inherited by their daughters which they immediately dominate.

With the above implementation, we can again take the structures in (8) as an example to demonstrate how the problem of the standard derivational approach in parsing can be addressed. When *John* and *is* are parsed (the left tree), T' is projected and *John* joins directly with T'. The indices of John and T' will be passed to each other's [C-commanded-by] features. Then in the middle tree, where *drinking* is parsed, VP is projected and VP inherits the content of the [C-commanded-by] feature from its mother node T' and obtains the index of its sister *is*. The [C-commanded-by] feature on VP is again inherited by *drinking*. As a result, the [C-commanded-by] feature on *drinking* is also [C-commanded-by: 1, 2]. With the same mechanism, it is easy to see that *coffee* in the right tree bears [C-commanded-by: 1, 2, 3]. In sum, for all the steps of the computation of the [C-commanded-by] feature on every parsed word, there is no need to do feature updating. Therefore, switching the c-command feature to c-commandees solves the efficiency problem of the standard derivational approach.

More complicated syntactic dependence relations can also be implemented using similar reformulations that I provide. More details of my proposal to address this

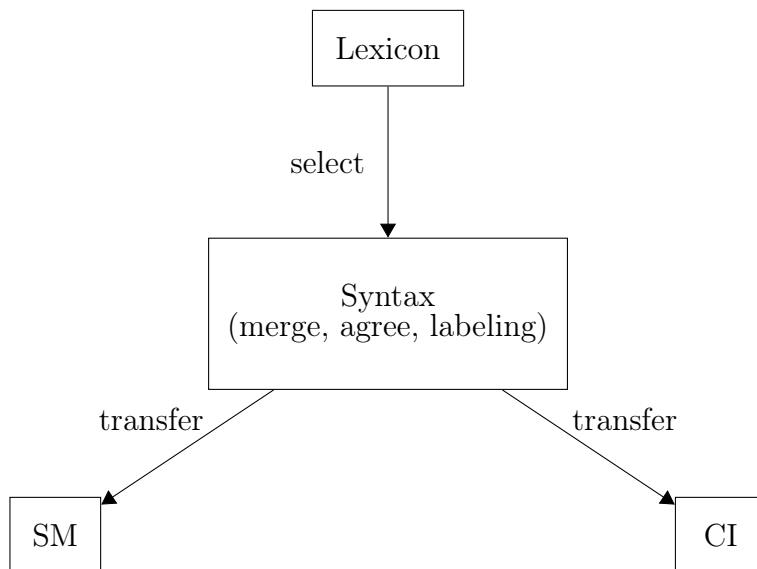
encoding problem can be found in Chapter V.

1.4 The framework

1.4.1 The language model

In this section, I lay out the syntax framework adopted in this dissertation. Minimalist syntax assumes an (upside-down) Y-shape language model as shown in Figure 1.1 (cf. Chomsky 2013, 2015b; Collins & Stabler 2016).

Figure 1.1: The Y-shape model



The lexicon is constituted by lexical items. Each lexical item is a feature bundle, consisting of various types of features including syntactic, semantic and phonological features. Lexical items from the lexicon are selected and inserted into the syntax module, where they are assembled into bigger syntactic objects (sets) through the operation MERGE. Two other operations may further modify the output of merge, namely AGREE and LABELING. The outputs of syntax then are TRANSFERRED to the two interfaces, i.e., SENSORY-MOTOR (SM) and CONCEPTUAL-INTENTIONAL (CI), for phonological and semantic interpretations, respectively.

There is no consensus as to where agree and labeling occur in the Y-shape model. We will leave aside that debate here, since it is not directly relevant to the current discussion. However, as a working hypothesis, I agree with Preminger (2014) that (ϕ -feature) agreement happens in syntax (but see Bobaljik 2008 for a competing proposal). In addition, I adopt Chomsky’s (2015b) assumption that labels of built syntactic objects give the interfaces instructions about what these syntactic objects are, and are thus needed for interpretation purposes at the interfaces. Consequently, labeling needs to occur (right) before or at the time of transfer.

The operation MERGE, which includes external merge and internal merge (= move), is defined below in (9), following Collins & Stabler (2016) (cf. Chomsky 1999).

(9) *Definition of merge*

Given any two distinct syntactic objects α , β , $\text{merge}(\alpha, \beta) = \{\alpha, \beta\}$.

(9) shows that merge takes two syntactic objects, α and β , which could be lexical items or sets of lexical items, and returns a set with α and β as its members.

I take vP and CP as phases, following Chomsky (2001b) and Chomsky (2008). I also assume the unit of transfer is the phase head complement, following the classical transfer theory (Chomsky 2001b and subsequent works). In addition, I assume that the transfer of the complement of a phase head $H1$ applies at the time when the next higher phase head $H2$ enters into the derivation, an assumption adopted from Chomsky (2001b). This is different from Chomsky (2008) and more recent works, where transfer applies to the complement of $H1$ at the time when movement to the edge of that phase occurs, after $H1$ has merged into the derivation. I adopt the theory of transfer of Chomsky (2001b) rather than that of Chomsky (2008) because only the former can account for long-distance agreement patterns, as will become clear in Chapter II (also cf. Ke 2016).

After a set is built by recursive applications of merge, some lexical items in the set

bear unvalued features that cannot be interpreted by the SM or CI interfaces. These unvalued features can be valued during the course of syntactic derivation through the operation AGREE. After these features are valued, they are interpretable at either interface. Otherwise, these unvalued features cause the derivation to crash at the interfaces, and the derived syntactic object is ungrammatical. Notice that here I define uninterpretable features with respect to both the SM and CI interfaces. This is different from the hypothesis held by Chomsky (2001b) and many others, who define interpretability as a semantic issue; that is, a feature is interpretable if and only if the feature is semantically interpretable at the CI interface. In this work, I assume that a feature is interpretable if it can be either phonologically or semantically interpreted; in other words, the feature can be pronounced or has semantic content. In the next chapter, I will provide some motivations for such a revision. To avoid ambiguity, I will use “semantically interpretable” whenever “interpretable” is used in the literature.

For a set to be interpretable to the interfaces, it needs to be labeled; otherwise the interfaces cannot know what these sets are. For instance, the label for {kick, {a ball}} would be “kick_V.” This label tells the interfaces that the set is verbal and must be interpreted as such. Note that the label is not the projection “VP,” as in earlier generative traditions. This is under the framework proposed by Chomsky (2013, 2015b), according to which projections and feature percolation are not legitimate syntactic operations. Projection and feature percolation are not well motivated in the current minimalist program because they are not directly derivable from more fundamental operations such as merge. In addition, projections are excluded from syntax because they violate the Inclusiveness Condition, which is given in (10).

(10) Inclusiveness Condition (Chomsky 1995b:228)

Any structure formed by the computation is constituted of elements already present in the lexical items.

The Inclusiveness Condition states that the computational system (syntax) cannot create new features (Chomsky 2000a). All the features must originate from lexical items. Positing projections requires syntax to create new objects, violating the Inconclusive Condition.

A “VP” is now considered a bare set as a direct result of merge. More generally, all syntax outputs are sets created by recursive applications of merge. Sets do not have features and they are not feature bundles; only heads/lexical items have those properties. Therefore, information about the syntactic nature of sets is not directly accessible. Consequently, traditional symbols such as TP or *v*P are now sets with labels representing their syntactic nature. In this thesis, I continue to use names such as TP or VP, but this is only for the sake of convenience and for expository purposes. Theoretically, I use them to refer to sets with labels such as “T” or “V.”

A critical difference between projections and labels is that projections are syntactic objects that syntax can operate on, whereas labels are not syntactic objects that syntactic computations such as agree and syntactic selection can operate on. Without projection and feature percolation, sets are different from heads as they do not bear features. Syntactic computations such as agree and syntactic selection cannot refer to labels directly. To put it concretely, subject–verb agreement in the example *a boy was...* now cannot be a relation between the subject DP *a boy* and the T head *was*, since the DP is a set and does not have features that are needed for an agreement relation. Instead, the agreement should be between the head of the subject DP, i.e., the D head *a*, and the T head *was*. What’s more, syntactic selection must be a head-to-head relation as well. For instance, a V head such as *kick* does not select a DP like *a ball*. Instead, *kick* selects a D head such as *a*, and then the D head *a* selects an N, *ball*.

Labels are nevertheless useful to the interfaces because they tell what the corresponding sets are. Chomsky (2013) therefore proposes that the syntactic outputs

must be labeled by the labeling algorithm for them to be interpreted at the interfaces. The labeling algorithm, however, is not explicitly defined by Chomsky (2013) nor by any other previous studies, to the best of my knowledge. Fortunately, Chomsky (2013) does provide us three configurations to which the labeling applies:

- (11) *Instances of labeling*
- a. $\{\alpha \text{ H, XP}\}$: $\{\alpha \text{ kick, \{the ball\}}\}$
 - b. $\{\beta \text{ XP, YP}\} = \{\beta \{\delta \text{ X, KP}\}, \{\kappa \text{ Y, ZP}\}\}$:
 $\{\beta \{\delta \text{ a, \{little boy\}}\}, \{\kappa \text{ is, \{kicking the ball\}}\}\}$
 - c. $\{\chi \text{ X, Y}\}$

I provide an example for each configuration in (11a) and (11b). I will not discuss (11c) as there are unnecessary complications that would take us too far afield. According to Chomsky (2013), (11a) is a straightforward case, in which the labeling algorithm looks into the set α and finds the head H. H is thus taken as the label of α . The labeling algorithm makes use of minimal search, a search algorithm which is argued to be a freely available 3rd factor, domain-general search algorithm (Chomsky 2005). The search algorithm is minimal because it terminates as soon as a target is returned. In the corresponding English example, the V head *kick* is returned as the label.

In (11b), the labeling algorithm searches into β and finds two sets simultaneously, i.e., δ (XP) and κ (YP). Then the algorithm continues to search into these two sets, and finds the heads of these two sets, namely X and Y. The set β therefore does not have a unique head as its label. With the corresponding example in (11b), both *a* and *is* are returned by minimal search.⁶ Chomsky (2013) further stipulates that when two heads are simultaneously returned by minimal search, the target set is labeless unless these two heads are somehow identified. For instance, if the two heads are in

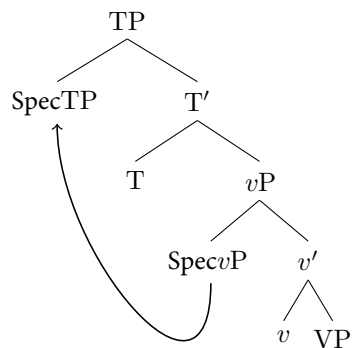
⁶The same situation is observed for (11c), as both the X and Y will be returned by the minimal search into the set γ .

an agreement relation, then the features in agreement are taken as the label of the set. In terms of the example given in (11b), the label for the set χ is the ϕ -features of the two heads that are returned by minimal search, namely *a* and *is*. In other words, the label is a feature pair $\langle\phi, \phi\rangle$. It will become clear that the analysis of (11b) is problematic once minimal search and the labeling algorithm are formally defined.⁷

1.4.2 The *v*P-internal subject hypothesis

I will also assume the so-called *v*P-internal subject hypothesis throughout this dissertation. In this subsection, I will briefly review some important arguments that previous studies offered in favor of the *v*P-internal subject hypothesis. This hypothesis can be illustrated by the following figure, where the subject is originally generated at the *v*P specifier (Spec*v*P) position and finally surfaces at the SpecTP position through movement. That is, the subject of the sentence is not originally generated in the surface subject position but is generated inside the *v*P.

Figure 1.2: The *v*P-internal subject hypothesis



Note that in the review of the literature below, for simplicity, I have changed some terms that the original authors used in their studies and reinterpreted those terms according to the assumptions under the current minimalist program. INFL and IP

⁷A formal definition of minimal search will be provided in Chapter II, which indicates that $\langle\phi, \phi\rangle$ cannot be the label unless we assume a minimal search for (11b) that is different from that for (11a).

have been translated to T and TP respectively, and the higher V and VP in a VP-shell analysis or the split-VP hypothesis (e.g., Larson 1988) have been re-coded as *v* and *vP*.

Theoretically, the *vP*-internal subject hypothesis is a natural consequence of the interaction between case theory and theta theory. It is generally assumed that a V or *vP* assigns DPs θ -roles whereas some other functional heads such as T or *v* assign case. Fukui (1986) proposes that the subject or the external argument of the predicate verb is originally generated in the specifier position of *vP*, where it receives its θ -role. Nevertheless, the external argument cannot get case there. The subject would have to raise to the specifier of the TP to receive case from the T head since T is a case assigner. Therefore, according to Fukui (1986), the *vP*-internal subject hypothesis is a by-product of the interaction between the θ -theory and the case theory.

Kitagawa (1986) further points out that the generation of the subject inside the *vP* has a plausible semantic basis: *vP* denotes a generic event type (without the information of tense and aspect).⁸ For instance, *which* in the following sentences refers to events that include the subject and the predicate but not the tense. If the subject is in fact raised from the Spec*vP* position and the subject and the predicate thus form a constituent, this semantic effect immediately follows.

- (12) a. John will talk to Mary, which usually makes her happy.
 b. Bill was talking to Mary, which usually is the sign of upcoming trouble.

Scope ambiguity provides another piece of evidence for the *vP*-internal subject hypothesis. Examples in (13) are ambiguous in the relative scope between *someone* and *must* or *likely*.

- (13) a. Someone must love her.
 b. Someone is likely to love her.

⁸See Chomsky (1995b:315) and references therein for a similar idea.

As for (13a), the two interpretations are shown in (14), based on whether *someone* or *must* possesses the wider or narrower scope.

- (14) a. *someone* > *must*:

There exists at least one person who the speaker believes loves her.

- b. *must* > *someone*:

The speaker believes that there exists at least one person who loves her.

The scope ambiguity can be easily captured by the LF in (15), with the assumption that *someone* is raised from the specifier of the *vP* and its trace left in that position is inside the scope of *must*, whereas the higher copy at the specifier of TP has scope over *must*.

- (15) [TP someone_i must [_{vP} t_i love her]]

Koopman & Sportiche (1991) use the raising verb *seem* to induce this ambiguity. For example, scope ambiguity between the subject and *seem* is observed in (16) , because as a raising verb, *seem* raises the *vP*-internal subject *a griffin* to the specifier of TP.

- (16) [TP A griffin_i seems to [_{vP} t_i be lurking on the 25th level]]

Sportiche (1988) further shows that scope ambiguity between negation and the floating quantifier *all* also supports the subject raising analysis. In both of the French sentences below (the English translations are mine), *not* can take scope over *all the children*, but only (17a) allows *all the children* to take scope over *not*.

- (17) a. Tous les enfants ne sont pas partis.
all the children did not leave
'All of the children did not leave.'
- b. Les enfants ne sont pas tous partis.
the children did not all leave
'Not all children left.'

This can be readily explained if we assume that *all* optionally pied pipes with the raising subject *the children*. In (17a), *all* moves together with the subject, and thus *all the children* have scope over *not*. In (17b), however, *all* does not pied pipe with the subject but is left in the Spec*v*P position, and therefore is always inside the scope of *not*.

In addition, Sportiche (1988) claims that his floating quantifier analysis also supports the *v*P-internal subject hypothesis.⁹ As exemplified by (18), the floating quantifier analysis assumes that the quantifier *all* attaches to the trace of the raised subject. This presupposes the subject raises from the specifier of *v*P to the surface SpecTP position.

- (18) The children_{*i*} have all t_{*i*} seen this movie.

The floating quantifier analysis provides a neat explanation of why *all* occurs between T and *v*P, and why *all* can quantify over an NP that it does not directly attach to in the surface structure: *All* in fact attaches to the trace of the subject. The success of the floating quantifier analysis in turn supports the *v*P-internal subject hypothesis.

Koopman & Sportiche (1991) then treat all T/INFL categories as raising verbs, paralleling them with traditional raising verbs such as *seem*. Subject raising associated with raising verbs are for case purposes. Take (19) as an example.

- (19) a. It seems that John sleeps all day.
b. John seems to sleep all day.

(19a) suggests that when the embedded clause is a finite clause (TP) and the subject *John* can receive case from T, no subject raising is required. By contrast, (19b) demonstrates that when the embedded clause is infinitive, the subject must raise to obtain case from the matrix T. This analysis thus views *v*P-internal subject hypothesis

⁹See Zyman (2018) for a recent strong argument for the floating quantifier analysis.

from a broader perspective and considers it a special instance of raising for case reasons.

1.5 Organization of the thesis

This thesis is organized as follows. Chapters II and III focus on the syntactic analysis of agreement and binding, followed by an examination of the syntax and semantics of ϕ -features on reflexives and verbs in Chapter IV. Chapter V proposes a novel method of encoding syntactic relations as features and retrieval cues, after a review of the previous attempts to overcome the encoding problem, to pave the way for the cue-based retrieval account (the Asymmetry of Interpretability Hypothesis in (5)) of the processing differences between subject–verb agreement and reflexive binding. In Chapter VI, results from a series of experiments will be presented to test the validity of the Asymmetry of Interpretability Hypothesis. Finally, Chapter VII concludes the thesis.

CHAPTER II

A Minimal Search-Based Analysis of Agreement

2.1 Introduction

This chapter develops a syntactic analysis of agreement. In the next chapter I will extend this analysis to reflexive binding. The goal of this chapter is to provide a syntactic basis for the intuition that subject–verb agreement and reflexive binding are syntactically similar to each other in certain ways. An important difference between these two constructions will be discussed in the Chapter IV.

Any syntactic analysis of (subject–verb) agreement and reflexive binding should capture both similarities and differences between these two constructions. On the one hand, both constructions involve a syntactic dependency. It seems obvious, at least upon initial inspection, that the ϕ -features on the verb/T must match those on its subject, and the ϕ -features on the reflexive must also be the same as those on its antecedent. As illustrated by (1) below, the values of the ϕ -features on the linking verb (the number and person features in English) need to be identical with those on its subject; otherwise the sentence is ungrammatical.

(1) **I** **am**/***are**/***is** writing the dissertation.

(2) **John** would support **himself**/***herself**/***yourself**/***themselves**¹ in college.

(2) shows that the same restriction applies to the reflexive *himself* and its antecedent

in their (stereotypical) gender, number and person features. The syntactic analysis of agreement and binding proposed in this chapter and the next chapter will explain how the relevant dependencies are established in these two constructions.

This chapter is organized as follows. In Section 2.2, previous syntactic analyses of agreement are reviewed, with comments on their pros and cons. Section 2.3 proposes a formal definition of minimal search, a notion that is critical in the current minimalist syntactic theory but is defined formally for the first time in this dissertation. I then argue that the search algorithm in agree and labeling can be unified by minimal search. Theoretical and empirical consequences of the proposal are discussed in Section 2.4 and Section 2.5, respectively. The minimal search-based agree analysis provides a unified analysis of various agreement/concord patterns across languages, including local downward agreement, long-distance downward agreement, upward agree, cyclic agree and complementizer agreement.

2.2 Previous analyses of agreement

With the *vP*-internal subject hypothesis established in Chapter I, we will see a few influential proposals of agreement. The theories of agree that I will review below make a heavy use of the *vP*-internal subject hypothesis.

In this section, I will review two important theories of subject–verb agreement and agreement in general: agreement as a Spec–head relation and the standard probing-based agree. Furthermore, I will mention an important current debate on the direction of agree under the probing-based agree analysis, which I argue can be solved with a new definition of agree.

¹Let us ignore for now the singular, gender underspecified use of *themselves*. I will discuss the effects of the singular *themselves* in the General Discussion of Chapter VI.

2.2.1 Agreement and Spec-head relations

The first influential formal theory of agreement emerges in the early Principles-and-Parameters era. Chomsky & Lasnik (1993) propose that Spec-head relations are responsible for both ϕ -feature agreement and case assignment. The leading idea is that there are independent projections for subject-verb agreement and object-verb agreement, i.e., Agr_sP and Agr_oP , respectively. The structure assumed is shown below in Figure 2.1 (I have relabeled the higher VP as $v\text{P}$ for convenience of discussion).

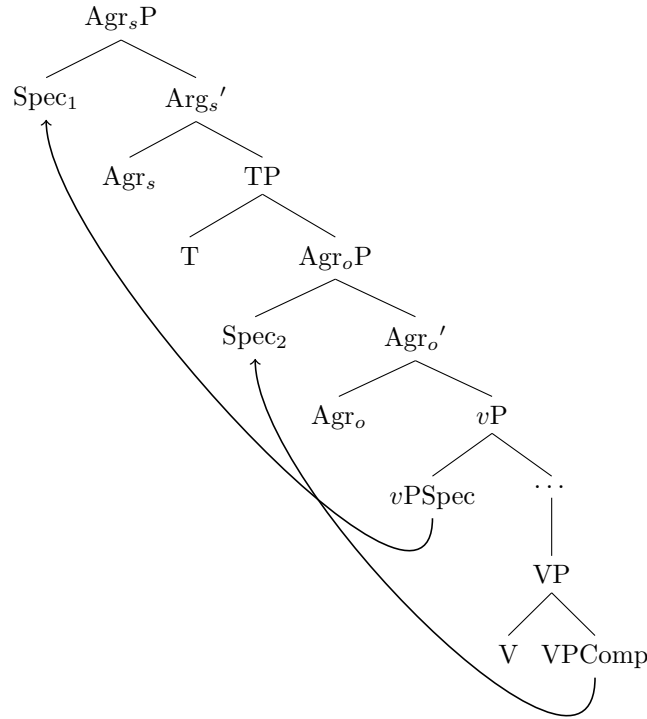


Figure 2.1: Agreement and the Spec-head relations

The Spec of Agr_sP , Spec_1 , is the position that the subject is raised to. The subject is originally generated in $\text{Spec}v\text{P}$, where the subject receives its θ -role. Similarly, the specifier of Agr_oP , Spec_2 , is raised from its θ -role position, i.e., the complement of V. The subject and the object in Spec_1 and Spec_2 agree with their corresponding Agr heads in their ϕ -features. In addition, according to Chomsky & Lasnik (1993), structural case is also assigned through the Spec-head relation. After an agreement

relation is established between a Spec (e.g., the subject or the object) and the corresponding Agr head, case will be assigned to the Spec as a “free rider” of the agreement relation (see also Chomsky 1995b, Ch. 4). Many works have adopted and developed this Spec–head analysis in the literature (e.g., Kayne 1989; Koopman 1992, 2006).

2.2.2 Agree

We have seen that the Spec–head agreement analysis unifies ϕ -feature agreement and case assignment. This is one of its advantages. However, under the Principles-and-Parameters framework, DPs are normally base-generated in θ -positions and then undergo raising/movement to receive case; therefore, this unification entails that agreement should always be accompanied by movement for case-assignment purposes. This unification thus becomes a problem when we examine instances of agreement that are not accompanied by movement.

Agreement in two types of constructions, existential constructions in English and long-distance agreement across languages, provide exactly such instances (where agreement is disassociated from movement and case assignment). Below I will provide a brief demonstration of why these two types of constructions pose a problem for the Spec–head agreement analysis.

Existential constructions are problematic for the Spec–head agreement analysis. The examples in (3) show that, instead of agreeing with the SpecTP *there*, the T heads *is* and *are* agree with *a man* or *three men*, respectively. That is, in these cases, T agrees with a DP in its c-command domain rather than its Spec.

- (3) a. $[_{TP} \text{ There } [_T \text{ is/*are } [\text{a man outside}]]]$
 b. $[_{TP} \text{ There } [_T \text{ *is/are } [\text{three men outside}]]]$

Based on the assumption that *there* is semantically vacuous, a solution to the above problem related to the existential construction is to assume that the expletive

there is inserted for syntactic or phonological reasons (e.g., to satisfy EPP), and the DP moves covertly to replace the expletive at LF (Chomsky 1986). Another solution is that only the formal features (including the ϕ -features) of the DP, rather than the whole DP with its phonological features, move at LF to the SpecTP position.

Both of these solutions encounter difficulties. Since both solutions assume LF movement, we expect this movement would have some semantic effects, such as scope ambiguity, as we have seen in the case of quantifier raising (May 1985). This prediction has been shown to be incorrect (see Lasnik & Saito 1999 for references). For instance, in the following examples (taken from Boeckx 2008), movement of the subject DP to SpecTP induces scope ambiguity (4a). By contrast, when the DP stays in situ in SpecvP, no such scope ambiguity is detected (4b), implying that at LF, the subject DP also does not move across *likely*.

- (4) a. *someone* > *likely* or *likely* > *someone*

Someone from New York is likely to be at the party.

- b. *likely* > *someone*

There is likely to be someone from New York at the party.

In addition, if binding relations can be established at LF, as argued by Chomsky (1995b), then we would expect the suggested covert raising of the DP at LF could feed binding. However, the contrast in (5a, b) shows that binding relations cannot be licensed with the proposed covert raising, whereas overt movement can. As a result, the reflexive *himself* in (5b) is left unbounded.

- (5) a. A man_{*i*} seems to himself_{*i*} to be doing something interesting.

- b. *There seems to himself_{*i*} to be a man_{*i*} doing something interesting.

Furthermore, long-distance agreement other than the existential constructions discussed in (5) provides another piece of evidence against the Spec-head agreement analysis.

Polinsky & Potsdam (2001) strongly argue against the Spec-head agreement analysis based on the following cases of long-distance agreement in Tsez,² where the matrix verb agrees with an embedded nominal in noun classes (boldfaced):

- (6) a. eni-r [už-ā magalu b-āc'-ru-li]
 mother-DAT [boy-ERG bread.**III**.ABS III-eat-PSTPRT-NMLZ]
 b-iy-xo
 III-know-PRES
 'The mother knows the boy ate the bread.'
- b. eni-r [uži ø-āy-ru-li] ø-iy-xo
 mother-DAT [boy.**I**.ABS I-arrive-PSTPRT-NMLZ] **I**-know-PRES
 'The mother knows the boy arrived.'

(6a, b) suggest that the verb ‘know’ agrees with the absolutive nominal inside the sentential (nominalized) argument, i.e., with ‘bread’ in noun class III in (6a), and with ‘boy’ in noun class I in (6b). This is arguably a real case of long-distance agreement if the absolutive nominals do not move to the matrix clause at any stage of the derivation, which implies that these absolutive nominals cannot be in a Spec–head relation with the verb ‘know.’ Polinsky & Potsdam (2001) convincingly argue that in Tsez, these absolutive nominals must stay inside the nominalized phrase and cannot move to the specifier position of the matrix verb ‘know.’

In their analysis, Polinsky & Potsdam (2001) show that all types of movement attested in Tsez cannot cross a clause boundary, including overt A- and A'-movement, covert wh-movement, covert quantifier raising and scrambling. Therefore, the absolutive nominals in (6a, b) are unlikely to be at the specifier position of the matrix verb 'know' through any of these types of movement. Furthermore, Polinsky & Potsdam (2001) illustrate that whenever there is a non-absolutive wh-phrase or a non-absolutive topic present at the left periphery of the embedded clause above the absolutive nominals, the long-distance agreement is blocked, further supporting the

²Long-distance agreement similar to that in Tsez is reported in various other languages such as Passamaquoddy (Bruening 2001) and Innu-aimûn (Branigan & MacKenzie 2002).

idea that the embedded absolutive nominals in long-distance agreement do not move out of the embedded clauses.³ Finally, Polinsky & Potsdam (2001) point out that the absolutive nominals that agree with the matrix verbs need to be at the specifier of the topic position either overtly or covertly, a position which is below the interrogative C in the left periphery. This explains why a wh-phrase and a non-absolutive topic can cause an intervention to the proposed long-distance agreement.

All the above problems to the Spec-head agreement analysis can be solved if we assume that the verb or the T head in fact agrees with a nominal inside its c-command domain. This solution is elegantly implemented by an operation called AGREE as proposed by Chomsky (2000a). The leading idea is that the verb or the T head bears unvalued features, and these unvalued features probe into the verb or T's c-command domain to look for matching valued features on a DP.

A definition of agree is given below in (7), taken from Chomsky (2000a), with minor modifications.⁴ The definition will be revised in Section 2.3.

(7) *Definition of agree*

Agree is a syntactic operation taking place between a probe P and a goal G in the domain of P, D(P), between which a Matching relation holds.

- a. Matching is identity of feature attributes;
- b. D(P) is the sister of P;
- c. Locality reduces to “closest c-command”;

(8) *Definition of closest c-command*

A matching feature G is closest to P if there is no G' in D(P) matching P s.t.

³Such intervention effects caused by a DP above the absolutive target is also taken as a diagnostic of the agree-based agreement, as will be introduced below (Preminger 2009). Similar arguments against the Spec-head agreement analysis that are based on long-distance agreement can be found in Boeckx (2008) and references therein.

⁴Chomsky (2000a) considers matching as feature identity. However, Chomsky (2004) points out correctly that the matching between an uninterpretable feature and a corresponding interpretable feature is not feature identity but feature non-distinctness. I revised this part of definition as “matching is identity of feature attributes,” addressing the problem of the original definition of agree.

G is in D(G').

Let us use (1), repeated below as (9), to demonstrate how agree works.

(9) [TP I_j am [_{vP} t_j writing the dissertation]]

Suppose that the T head *am* agrees with the *vP*-internal subject, i.e., the lower copy of the subject *I*. Before the *vP*-internal subject moves to its surface subject position (SpecTP), and before T agrees with the *vP*-internal subject, the derivation stage is represented below in (10).

(10) [TP be [_{vP} I writing the dissertation]]

The subject–verb agreement relation is now established by a probe–goal relation instead of a Spec–head relation. The T head *be* originally bears unvalued ϕ -features (person and number). The set of unvalued ϕ -features functions as a probe that seeks a goal. The probe looks into T's c-command domain, the *vP*. The ϕ -set on *I* (1ST; SG) matches the probe. As a result, an agreement relation is established by the probe–goal matching relation. The values of the ϕ -features on *I* can then be copied to the probe on T. Consequently, the T head is pronounced as the first-person, singular *am*. It should be noted that the object *the dissertation* cannot be found to match the probe, as the ϕ -features on the *vP*-internal subject intervene and are closer to the probe, based on the definition of closest c-command given in (8).

2.2.3 The direction of agree

Agree as a long-distance feature valuation successfully solves many problems raised by the Spec–head agreement analysis, as we have mentioned. However, further studies reveal that the probing-based agree analysis is not without problems. One of the most prominent problems still lingering in the field concerns the directionality of agree: Agree is not always probing into the c-command domain of the probe-bearer. In

this section, I will briefly review some challenges to the probing-based agree analysis mentioned in the previous section.

The central argument against the probing-based agree approach comes from studies on negative concord (Zeijlstra 2012) and inflection doubling (or parasitic participles) (Wurmbrand 2012a,b). Additional evidence comes from multiple case licensing (or multiple agree) in Japanese. Bjorkman & Zeijlstra (to appear) argue that negative concord, inflection doubling and multiple case licensing are all morphosyntactic phenomena, similar to subject–verb agreement, because they are subject to syntactic locality relations such as clause boundedness. They further assume that feature agreement is the mechanism behind all of these morphosyntactic phenomena. Let us discuss them one-by-one.

Let us start with negative concord. Zeijlstra (2004, 2012) argues that negative concord in Czech should be considered a syntactic agreement between a negative operator and an n-word. (11) indicates that with the negative marker *ne*, the sentence acquires a negative meaning. However, as exemplified in the comparison between (12a) and (11) and that between (12b) and (12a), adding an extra n-word *nikdo* or *nikomu* does not change the negative meaning of the sentence.

(11) *Negative concord in Czech* (Zeijlstra 2004:249)

Milan *nevidi*
Milan NEG.sees

‘Milan doesn’t see.’

(12) *Negative concord in Czech* (Zeijlstra 2012:501)

a. Dnes *nikdo* *(ne)volá
Today n-body NEG.calls
‘Today nobody is calling.’

b. Dnes *nikdo* *(ne)volá *nikomu*
Today n-body NEG.calls n-body
‘Today nobody is calling anybody.’

Logically, if n-words carry a negative meaning, we would expect that the two n-words in (12a) will cancel each other out and the sentence should be interpreted as a positive statement. This is not what we observe for (12a). The same reasoning applies to (12b). Therefore, Zeijlstra (2004) considers n-words as semantically non-negative indefinites that carry a semantically uninterpretable [uNEG] feature. In addition, Zeijlstra suggests that the negative marker *ne* in Czech could be analyzed either as the phonological realization of a [uNEG] feature that needs to be checked at some point during the derivation, or as a negator bearing a semantically interpretable [iNEG] feature. Zeijlstra (2004) further assumes that n-words must be licensed by or agree with an element bearing an [iNEG] feature. This element could be the negative marker *ne* or a null negative operator, *Op*, the latter of which bears an [iNEG] feature and is projected at a position higher than the TP.

Depending on how the negative marker *ne* is analyzed (it may bear either an [iNEG] or a [uNEG] feature), the structure of (12b) can be represented either as in (13a) or as in (13b).

- (13) a. [TP *nikdo*<sub>[uNEG]_j] [_{NegP} *nevolá*_[iNEG] *t_j* *nikomu*_[uNEG]]]
 b. *Op*_[iNEG] [TP *nikdo*_[uNEG] *nevolá*_[uNEG] *nikomu*_[uNEG]]</sub>

(13a) indicates that if we assume that the negative marker *ne* bears an [iNEG] feature, it is possible that *nikdo* can agree with it after it moves to the surface subject position, according to the agree operation defined in (7). However, since agree requires the syntactic object with uninterpretable features to c-command the one with interpretable features, there is no way for the object *nikomu* to agree with *ne*, because *ne* (or *nevolá*), bearing an [iNEG] feature, c-commands the object, which bears a [uNEG] feature. If we adopt the analysis in (13b), on the other hand, all the syntactic objects bearing the [uNEG] are c-commanded by the null negative operator which bears the [iNEG] feature. Again, probing-based agree cannot be applied to check these [uNEG] features.

Zeijlstra (2012) thus claims that in negative concord in Czech, a different type of agree must be adopted, which he calls UPWARD AGREE. The idea is that a syntactic object bearing uninterpretable features agrees upward with a c-commanding syntactic object bearing corresponding interpretable features. Upward agree is defined as in (14).

(14) *Definition of upward agree* (Zeijlstra 2012)

α agrees with β iff:

- a. α carries at least one uninterpretable feature and β carries a matching interpretable feature;
- b. β c-commands α ;
- c. β is the closest goal to α .

I would like to briefly note that according to (14c), in the configuration where there are two c-commanding syntactic objects carrying matching interpretable features, α will agree with the one closer to it. I will return to this point and show that it makes incorrect predictions with regard to the subject-complementizer agreement in Lubukusu.

In addition to negative concord, inflection doubling (parasitic participles) also call for an analysis along the lines of upward agree (Wurmbrand 2012a). Wurmbrand (2012a) encodes morphological selection (e.g., a modal selects an infinitive verb) as feature agreement. She calls this type of feature agreement “reverse agree,” essentially identical to the upward agree in the sense of Zeijlstra (2012).

Parasitic participles can be found in many languages including Norwegian, Frisian and Swedish, examples of which are listed in (15–18). The parasitic participles are in boldface, and the auxiliaries that license the parasitic participles are underlined.

(15) *Norwegian* (Wiklund 2001:201)

Jeg	<u>hadde</u>	villet	[lest	/	lese	boka]
I	had	want.PART	read.PART	/	read.INF	book.DEF

‘I would have liked to read the book.’

(16) *Frisian* (den Dikken & Hoekstra 1997:1058)

hy soe [it **dien** / dwaan] wollen ha
he would it **do.PART** / do.INF want.PART have.INF

‘He would have liked to do it.’

(17) *Swedish* (Wiklund 2001:200)

Jag hade velat [läst / läsa boken].
I had want.PART **read.PART** / read.INF book.DEF

‘I would have liked to read the book.’

(18) *Swedish* (Wiklund 2001:202)

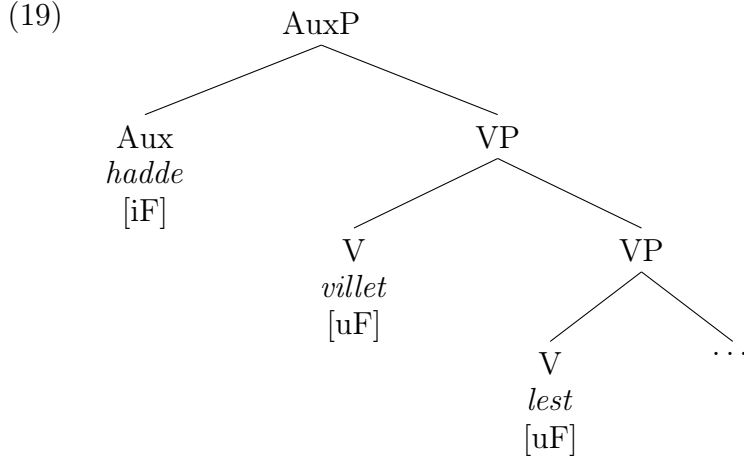
Han hade velat [kunnat [simmat]]
he had want.PART **can.PART** **swim.PART**

‘He would have liked to be able to swim.’

The boldfaced participles in the above examples are called parasitic participles, because they are semantically vacuous. Different from genuine participles (i.e., the participles that are not in boldface in (15-18)), parasitic participles are not interpreted as perfectives; instead, their semantics is identical to that of infinitives. It is possible to have more than one parasitic participle in a single clause, e.g., (18). Parasitic participles are in fact optional alternatives to infinitive verbs, as indicated in the above examples. In addition, they must be licensed by an auxiliary, which are underlined in the examples.

Wurmbrand (2012a) assumes that semantically vacuous parasitic participles carry an uninterpretable feature. Since a c-commanding auxiliary can function as a licenser, Wurmbrand (2012a) suggests that the auxiliary carries a corresponding interpretable feature that agrees with the parasitic participle it licenses. Abstracting away from details of the features involved, the agreement pattern proposed for the example in

(15), for instance, can be represented as follows:⁵



If it is correct that parasitic participles carry an uninterpretable feature and that this feature must be valued against the corresponding interpretable feature on the higher auxiliary, as represented in (19), then this is a case that does call for an upward agree-based analysis.

Let us now consider another case which researchers have suggested to be an instance of upward agree: multiple case licensing in Japanese. (20) illustrates that in Japanese raising-to-subject constructions, multiple DPs can be marked as nominative within an infinitival embedded clause.

(20) *Japanese* (Hiraiwa 2001:76)

John-ga [yosouijouni nihonjin-ga eigo-ga hidoku]
 John.NOM than-expected the-Japanese.NOM English.NOM bad.INF
 kanji-ta.
 think.PAST

‘It seemed to John that the Japanese are worse at speaking English than he had expected.’

⁵It is unclear why the highest perfective participle, e.g., the one on *villet* ‘want’ in (15), while semantically interpretable, should bear an uninterpretable feature. In fact, if *villet* bears an interpretable feature, the upward agree analysis wrongly predicts that the uninterpretable feature on the parasitic participle agrees with the interpretable feature on *villet*. I will show that the analysis proposed in this dissertation is immune to this problem. However, it must be noted that this potential complication does not affect the overall argument that parasitic participles pose a challenge to the standard agree theory.

The structure of (20) is as shown below:

$$(21) \quad [\text{DP1.NOM} [\text{Adv DP2.NOM} \dots \text{DP3.NOM} \dots \text{V.INF}] \text{T}]$$

If we assume that (i) case assignment from the T head to the DPs in (21) is done through agreement, (ii) T or (*v*) bears an interpretable [iCase] feature that is responsible for case assignment, and (iii) all DPs carry uninterpretable [uCase] features that must be valued by agreeing with a T (or *v*), (21) presents a problem to the standard probing-based agree. The problem is that the DPs with the uninterpretable [uCase] feature do not c-command the T head which bears the interpretable [iCase] feature. To address this problem, one must stipulate that there is some uninterpretable feature on T, other than the case feature, that triggers the downward search aiming for DPs bearing corresponding interpretable features, and then case assignment is just a free rider of the agreement relation that has been established. One such choice is to assume that the uninterpretable ϕ -features on T initiate the search for DPs with interpretable ϕ -features. However, this analysis encounters a different problem. If agree is defined as in (7), we would expect that search will be terminated as soon as a DP with interpretable ϕ -features is found. That implies that T in (21) will probe for and agree with the subject DP only, and therefore can assign case to that DP only. In other words, an intervention effect caused by the subject DP is predicted. Contrary to this prediction, no such intervention effects are observed in (21), because all the DPs are assigned case by the T head. As a result, Hiraiwa (2001) proposes the operation MULTIPLE AGREE, as defined in (22), to account for the absence of intervention effects in (21).

$$(22) \quad \textit{Multiple agree} \text{ (Hiraiwa 2001:69)}$$

Multiple agree with a single probe is a single simultaneous syntactic operation; Agree applies to all the matched goals at the same derivational point derivationally simultaneously.

The multiple agree analysis assumes that the T head in multiple nominative case assignment bears a [+multiple] feature, and that this feature enables the probe to look for multiple matching goals that are locally available. Such an analysis therefore requires the T head to be ambiguous between either bearing a [+multiple] feature or not, depending on the construction. To some extent, this is simply a restatement of the facts and does not provide a real explanation for why multiple agree occurs.

It is quite obvious that an upward agree analysis of (20) does not need to assume simultaneous agree or a T that bears the [+multiple] feature, in contrast to multiple agree in (22). This is because each DP carrying an uninterpretable case feature can independently search upward and agree with the interpretable case feature on T. No intervention effects are expected, in accordance with the data. Therefore, an upward agree seems simpler and less stipulative than the multiple agree analysis, as far as the multiple case-assignment in (20) is concerned.

In sum, we see that the classical probing-based agree theory solves significant empirical problems faced by the Spec-head analysis with regard to various types of long-distance agreement. However, there is also evidence that agree, at least in the cases discussed, should be directed upward. More details of the debate on the directionality of agree can be found in Zeijlstra (2012); Preminger (2013); Preminger & Polinsky (2015) and Bjorkman & Zeijlstra (to appear). In the following section, I will re-implement agree with a more fundamental operation, MINIMAL SEARCH, and argue that this re-implementation of agree can provide a solution to the debate on the directionality of agree. Furthermore, the proposed formalization of agree captures the syntactic similarities between subject-verb agreement and reflexive binding.

2.3 The proposal

2.3.1 Theoretical motivation

Before the presentation of a formalization of minimal search, I would like to first give the theoretical motivations, which will be followed by a demonstration of one of the empirical consequences: The formalization of minimal search sheds light on a possible solution to the debate on the directionality of agree. In addition, the proposed theory also leads to a neat unification of the syntactic analysis of subject–verb agreement and reflexive binding. Let us begin with the theoretical motivation, in order to fit the proposal into the overall architecture of the language model.

As previously mentioned in the language model in Chapter I, Section 1.3, the syntactic component includes three operations: merge, agree and labeling. If agree and labeling can be removed from the system and be replaced by the 3rd factor minimal search without a significant empirical loss, then the system would be considerably simplified, as desired.

Chomsky (2013) suggests an appealing approach exactly along these lines. Chomsky (2013) writes, “The simplest assumption is that LA (labeling algorithm) is just minimal search, presumably appropriating a third factor principle, as in Agree and other operations.” (p. 43). In Chomsky (2015a), Chomsky continues this discussion, “Optimally, projection should be reducible to a labeling algorithm LA, a special case of minimal search (like Agree), which in turn falls under MC (minimal computation).” Chomsky therefore hypothesizes that an optimal system would have agree and the labeling algorithm derived from a more general, independently motivated third factor principle, i.e., minimal computation. This conjecture, if successful, could unify the labeling algorithm and agree by resorting to an arguably third factor operation, minimal search, which is considered a type of minimal computation. Consequently, the labeling algorithm and agree might be freely available since minimal search, as a

third factor, is freely and universally available (Chomsky 2005).

A serious implementation of this conjecture requires a formal definition of minimal search. However, to the best of my knowledge, minimal search has not been formally defined. I thus provide a definition of minimal search in the next section. I argue that labeling and agree can only be partially unified by minimal search. On the one hand, labeling and agree are independently needed and they serve distinctive purposes. On the other hand, the search algorithm in labeling and agree can be captured by my definition of minimal search. I will then discuss the implications of this particular definition with respect to agree, labeling and reflexive binding, suggesting that the similarities between agreement and reflexive binding can be captured by the minimal search-based agree.

2.3.2 Defining minimal search

Recall from the last section that Chomsky (2013) does not give a formal definition of minimal search; however, he does illustrate how minimal search works in labeling and agree by examples. Let us consider again some of our previous examples for labeling in Chapter I, repeated below in (23a, b).

- (23) a. $\{\alpha \text{ kick, \{the ball\}}\}$
 b. $\{\beta \{\delta \text{ a, \{little boy\}}\}, \{\kappa \text{ is, \{kicking the ball\}}\}\}$

It seems that for the labeling of the set α in (23a), minimal search needs to look into α , and only the head *kick* is returned. In this case, minimal search seems to look for a head and ignore the set complement *the ball*. This is reasonable because heads can be considered bundles of features, and sets do not carry features (Hisatsugu Kitahara, p.c.), if we do not assume feature projection or percolation, following Chomsky (2013) and Chomsky (2015b). In other words, minimal search in this case could search for any features. As long as it finds a feature, it returns the head bearing that feature,

and then the search is immediately terminated. The term “minimal” therefore refers to the fact that the search ends whenever the first target is returned.

For (23b), minimal search needs to find a label for the set β , and thus looks into it. It cannot find any feature in β ; instead, two sets are found. As a result, minimal search looks into the two sets, δ and κ . The heads it returns for these two sets are *a* and *is*. According to Chomsky (2013), the features in agreement between *a* and *is*, that is, the ϕ -feature pair $\langle 3SG, 3SG \rangle$, will be taken as the label of β .

Notice that based on the labeling operation for structures such as (23b), Chomsky (2013, 2015b) could potentially derive EPP, which requires every (finite) TP to have a subject in its specifier position.⁶ This is because after T agrees with the *v*P-internal subject, an agreement which will be discussed in next paragraph, *v*P-internal subject needs to move to a higher position to receive case; a place that the subject could merge to without a labeling failure is T' ($=[T, vP]$), since only that T is in an agreement relation with the *v*P-internal subject, or more accurately, the D head of this DP.

Now the question is how T can agree with the *v*P-internal subject and how minimal search is involved in this agreement. Let us use (24) as an example. A direct comparison between agree in (24) and labeling in (23) may help us find out what exactly the operation is that labeling and agree might be reduced to. (24a) is the derivation step before the T head, *be*, agrees with the D head of the *v*P-internal subject *this man* in (24b).

- (24) a. $[_{T'} be [_{vP} [_{DP} this\ man]_j [_{v'} writing\ the\ dissertation]]]$
b. $[_{TP} [_{DP} this\ man]_j is [_{vP} t_j [_{v'} writing\ the\ dissertation]]]$

After *be* merges into the derivation, as shown in (24a), a minimal search is initiated, and it looks into the *v*P set for a head bearing valued ϕ -features corresponding

⁶Chomsky (2013, 2015b) works out the derivation of EPP for finite TPs, as he assumes that the T head in English is too weak to be the label of a TP, and therefore requires a SpecTP to be in an agreement relation with T, so that the $\langle \phi, \phi \rangle$ pair can be the label. A non-finite T then does not require a SpecTP because they cannot be in an agreement relation.

to those unvalued ones on *be*. No heads, but two sets, i.e., the DP and v' , are found. So minimal search has to continue to search into these two sets. Finally, the head *the* in the DP is returned, and the minimal search terminates.

Let us now summarize the role minimal search plays in labeling and agree. First, it looks into a set for some specific features (agree) or any kind of features (labeling). Second, it applies iteratively to every available set, and it is terminated whenever a target is returned. To state it more formally, minimal search is a search algorithm that looks into a certain SEARCH DOMAIN (sets) for certain SEARCH TARGET (features). The search is minimal because it is terminated as soon as a target is returned.

Now we are ready to give a definition of minimal search. As shown in (25), the current definition of minimal search includes a search algorithm that applies iteratively to a search domain (SD) to look for a search target (ST).

(25) *A formal definition of minimal search*

MS = <SA, SD, ST>, where MS = minimal search, SA = search algorithm, $SD_{\in set}$ = search domain (the domain that SA operates on), ST = search target (the features that SA looks for).

Search Algorithm (SA):

- a. Given SD and ST, matching against every head member of SD to find ST.
- b. If ST is found, return the heads bearing ST and go to Step (c); Otherwise, get the set members of SD and store them as a list L.
 - i. If L is empty, search fails and go to Step (c); otherwise
 - ii. assign each of the sets in L as a new SD and go to Step (a) for all these new SDs in parallel.
- c. Terminate search.

The description of the search algorithm in (25) can be spelled out more explicitly, e.g., in terms of pseudo-codes:

(26) *Pseudo-code for the SA in minimal search*

Global variable:

new_SD = List()

Functions:

1. List(x, y):

return: $[x, y]$

2. Member($x=\{y, z\}$):

return: List(y, z)

3. Multi_Find($x, y=[a, b, c...]$):

return:

List(Multi_processor(Find(x, a), Find(x, b), Find(x, c), ...))

4. Find(x, y):

found_heads = List()

if y is a head and y includes x :

append y to found_heads;

if y is a set:

join new_SD and Member(y)

if y is a list:

Multi_Find(x, y):

return: (found_heads, new_SD)

Algorithm:

Start: Terminate_search = False, new_SD = List(), given ST, SD

While Terminate_search is equal to False:

(found_heads, SD) = Find(ST, SD)

if found_heads is not empty:

```

        Terminate_search = True
    else if SD is not empty:
        new_SD = List()

```

Note that in the pseudo-code I obtain the results of the parallel minimal search algorithm in (25) by applying the `Multi_Find` function, which aggregates a list of returned heads by executing the `Find` function in multiple processors simultaneously, assuming that multiple processors are available for minimal search. I adopt a parallel search algorithm because, when there is a list of SDs, I do not see a principled reason for the search algorithm to prioritize certain SDs and look into those SDs before looking into others.

In addition, I use lists instead of sets to store the new SDs and the returned heads. Since I do not assume a linear relation between the SDs or the returned heads, sets are most appropriate for this use; however, sets are already used for structure building.

2.3.3 An example

As an illustration of the minimal search as just defined, let us walk through an example. As shown below in Figure 2.2, a minimal search is initiated to search for an $ST = \text{feature } [F]$, in the $SD = \text{set } \alpha$. Three runs of search are conducted before the target is found. In the first run, minimal search looks into α . X , the head member of α , is returned by the first run of search. However, since it does not bear $[F]$, and its co-member is a set, the search is not terminated. The set member of α , that is, the set β , is then assigned as a new SD.⁷ This is the second run. The second run of search finds only set members, i.e., γ and κ , and these two set members are stored as a list L . The two sets in L are then assigned as the new SDs for the third run of search, and

⁷Note that the minimal search defined in (25) does not allow a single minimal search applying to different sets. In order to search into a different set, a new minimal search must be initiated taking that set as a new SD.

two independent minimal searches are initiated in parallel. The head $K_{[F]}$ is found inside the set κ as it bears the ST. $K_{[F]}$ is finally returned by the search algorithm. The minimal search over the set γ fails and is terminated without returning a result.

It is worth noting that the minimality in the breadth-first search is naturally captured by *storing sets as a list L* in the definition (25b), not by, for instance, keeping track of/counting the levels of sets that minimal search looks into. All the sets which are at the same level will be stored in the same list, and new minimal searches are conducted to look into all these sets in parallel, as desired.

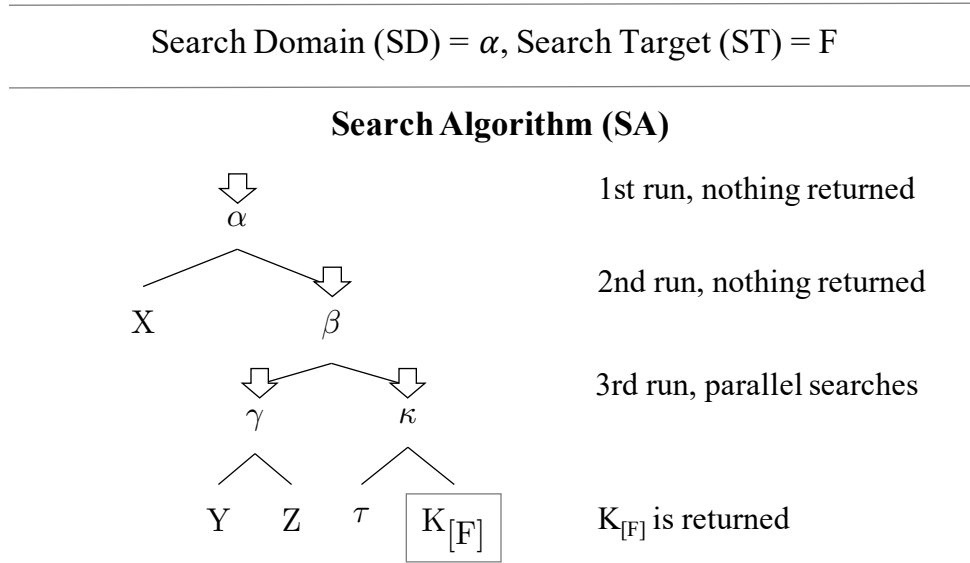


Figure 2.2: An illustration of minimal search

2.3.4 Breadth-first or depth-first?

There are two popular search algorithms in graph or tree search, breadth-first or depth-first search. In the case of traversing through a search tree, a breadth-first algorithm explores all the daughters of a starting node before it searches any deeper. In Figure 2.3, breadth-first search visits the daughters of the starting node N1 before moving on to any of the daughter nodes of N2 or any node deeper. After all the daughters of Node 1 are visited, it then continues to traverse through all the

daughter nodes of Node 2 and Node 3. So on and so forth, until all nodes in the search tree are visited and the algorithm is terminated.

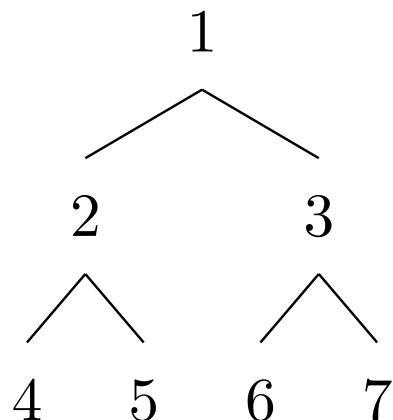


Figure 2.3: Breadth-first search

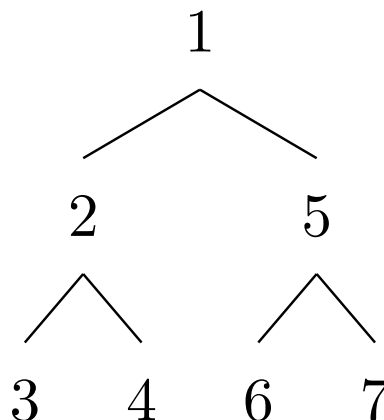


Figure 2.4: Depth-first search

On the contrary, a depth-first algorithm exhaustively searches down along certain nodes before backtracking. Take the tree in Figure 2.4 for an example. Starting from Node 1, depth-first search first visits a certain daughter of Node 1, e.g., Node 2, and then a daughter node of Node 2, e.g., Node 3, before it visits all the daughters of Node 2 and 3. Depth-first search reaches an end node (leaf node) there, so it backtracks to another daughter of Node 2, i.e., Node 4; then another daughter of Node 2, i.e., Node 5; and then a daughter of Node 5; so on and so forth, until it exhausts all the nodes dominated by Node 1.

Given an SD $\{\alpha \beta, \gamma\}$, where Greek letters represent sets, a breadth-first search explores all the members of α first before visiting the members of β and γ . On the other hand, a depth-first search goes to a member of α , for instance, β , and visits the members of β before exploring γ . The search algorithm in the definition of minimal search (25) is breadth-first.

I use a breadth-first algorithm here for two reasons. First, the hierarchical nature of syntactic structures is one of the most important aspect of language and linguistic

analysis, and breadth-first search respects hierarchical structure as it looks into nodes that are at a higher level before visiting nodes at lower levels. By contrast, depth-first search, to some extent, ignores some aspects of hierarchical structure. In many cases, it explores a node in a deeper/lower level before backtracking to a higher node. This implies that depth-first search cannot always capture the asymmetric c-command relations (it primarily captures the containment relations): it may reach a c-commandee before accessing its c-commanders. In principle, a minimal search would need to return as early as possible a node containing the search target if the node is structurally closest to the starting node for search; structural closeness is defined in terms of c-command: If A c-commands both B and C, B is closer to A than C iff B c-commands C. A breadth-first algorithm thus serves exactly this purpose.

Another reason that a breadth-first instead of a depth-first algorithm is employed is that once set theory is adopted for structure building, linear order becomes a peripheral property. This is because the members of a set are standardly assumed to be symmetric with regard to their linear order. Linear order is not specified until syntactic objects are linearized at the SM interface. In addition, the syntax component has no way of determining which member of a set is more important than the other. The head-hood information, which probably gives some sense of “importance” to a certain member of a set, is not available until labels are computed at the time when syntactic objects are sent to CI for interpretation (see below for a discussion of the labeling theory). Although the non-ordering nature of set members is not a problem for the depth-first algorithm technically, it does not favor the depth-first algorithm. Instead, it favors the parallel breadth-first search algorithm, as such an algorithm looks into both set members simultaneously in parallel, without making a distinction between them. Therefore, the breadth-first search algorithm is more consistent with the symmetric relation between set members.

2.3.5 Minimal search-based agree and labeling

We have given a formal definition of minimal search. In this section, we will give a definition for labeling and a revised definition for agree (revised from that proposed by Chomsky 2000a; see (7)). It will become clear that agree and labeling cannot be completely reduced to minimal search.

As shown in (25), minimal search involves three components, the SD, ST and SA. The SA is invariant and identical for all cases of minimal search, no matter whether it is for agree or labeling. What distinguishes agree from labeling is the SD and ST. The SD and ST being independently assigned by different operations is a unique feature of the current definition of minimal search, and I will show in the next section that this feature has far-reaching empirical implications. But first I would like to explain how the SD and ST are determined in agree and labeling.

The SD of minimal search for labeling and agree are determined in accordance with the purpose of these two operations. Agree connects two heads, one carrying unvalued features and the other carrying matching valued features. In order for two heads in an agreement relation to be connected to each other syntactically, they must form a syntactic relation at a certain stage of derivation. That is, they must be contained in the same root set or under a single node in a syntactic tree at a certain stage of derivation. Suppose a head X with an unvalued feature uF , represented as $X_{[uF]}$, needs to be valued by another head Y with a corresponding valued feature vF , represented as $Y_{[vF]}$, there are three situations that are logically possible for X to be connected to Y at a certain point of the derivation, with regard to the c-command relations between X and Y :

- (27) a. $X_{[uF]}$ merges with $Y_{[vF]}$ or a set containing $Y_{[vF]}$: $X_{[uF]}$ c-commands $Y_{[vF]}$;
b. $Y_{[vF]}$ merges with $X_{[uF]}$ or a set containing $X_{[uF]}$: $Y_{[vF]}$ c-commands $X_{[uF]}$;
c. A set containing $Y_{[vF]}$, α , merges with a set containing $X_{[uF]}$, β . Neither

$X_{[uF]}$ nor $Y_{[vF]}$ c-commands the other.

Among the instances of agree we reviewed in Section 2.2, downward long-distance agree is an instance of (27a), upward agree in negative concord, inflection doubling/parasitic participles and multiple case licensing are instances of (27b). None of the agreement patterns we have discussed thus far belongs to (27c). However, I will argue in Chapter III that reflexive binding is an instance of (27c), and the minimal search-based agree proposed in this chapter can capture cases where neither one of the two heads in an agreement relation c-commands the other.

Importantly, with (27a), the minimal search initiated to value $X_{[uF]}$ needs to search into the c-command domain of X . That is, the search domain of the minimal search in this case must be the sister, or co-member, of X . However, a problem immediately arises with regard to (27b). Notice that there is no way for the specific formalization of minimal search in (25) to search “upward.” Minimal search as defined in (25) always searches “downward” into its SD. I will address this problem in the next section. I will discuss the details concerning the minimal search for (27c) in Chapter III.

The ST for agree is simply the unvalued features on the head that initiates the minimal search. For instance, when the purpose of a minimal search is to find a head with valued features matching the unvalued feature attributes [Person: , Num:] on a T head, then [Person: , Num:] are taken as the ST.⁸

On the other hand, since the labeling algorithm is meant to provide a label for a given set, the SD will be that set. For instance, if the set {kick, {the, ball}} is the set that needs to be labeled, this set is the SD of the minimal search for the labeling purpose.

⁸I leave open for now the question whether all the unvalued features should be combined as a single ST and initiate a single minimal search, or each of them should initiate an independent minimal search. These two options do not make a distinction for the agreement patterns we have discussed thus far. For cases where we may need to split the unvalued features, please see e.g., Béjar & Rezac (2003) and Preminger (2009). The current definition of minimal search is in principle compatible with the split analysis. A mixed analysis with a combined ST in some cases but with split STs in some other cases may also be possible.

The ST of minimal search for labeling is more complicated. Take (28) as an example. In the case of (28a), the ST would be any feature or any head, not a particular head or feature.

- (28) a. $\{\alpha \text{ H, XP}\}$: $\{\alpha \text{ kick, \{the ball\}}\}$
 b. $\{\beta \text{ XP, YP}\} = \{\beta \{\delta \text{ X, KP}\}, \{\kappa \text{ Y, ZP}\}\}$:
 $\{\beta \{\delta \text{ a, \{little boy\}}\}, \{\kappa \text{ is, \{kicking the ball\}}\}\}$

However, for (28b), if we want to obtain a $\langle\phi, \phi\rangle$ pair, as suggested by Chomsky (2013), then the ST must be ϕ -features as well, a set of specific features. This assumes that the labeling algorithms for (28a) and (28b) should be different. A unified labeling algorithm, by contrast, has to (i) set the ST to any feature or any head (not a particular one), and therefore will return only heads for both (28a) and (28b); or (ii) set the ST to some specific features such as ϕ -features. A problem for (ii) is that when features other than ϕ -features are in agreement, e.g., in an interrogative CP, the features in agreement, i.e., the $\langle\text{Q}, \text{Q}\rangle$ pair, cannot be returned as a label, unless the ST of the minimal search for an interrogative CP is set to Q. As a result, if we want a unified labeling algorithm for all the cases we discussed (we surely do!), the label for (28b) must be $\langle\text{a, is}\rangle$ rather than a $\langle\phi, \phi\rangle$ pair.

To summarize, I have shown that the **search algorithm** in agree and labeling can be unified by the **search algorithm** of minimal search. However, the **values** of the SD and ST of minimal search are independently determined by agree and labeling, and thus the SD and ST values for agree are different from those for labeling. Minimal search for agree and labeling thus serve distinctive purposes.

We are now ready to give definitions of minimal search-based agree and minimal search-based labeling, which I will henceforth abbreviate as MS-agree and MS-labeling:

- (29) Definition: minimal search-based agree (MS-agree)

Agree = minimal search + valuation

i. **Minimal search**

a. Input to minimal search:

- SD = co-member of $H_{[uF]}$ ⁹
- ST = feature attributes of uFs (e.g., [Person: , Num:])

b. Output of MS: heads bearing the ST

ii. **Valuation**

- Trigger: the head bearing uFs
- Valuation: copy [vFs] values in the minimal search output to the corresponding [uFs] in the trigger

In the definition of agree in (29), agree consists of two steps: minimal search and valuation.¹⁰ The separation between minimal search and valuation is a natural consequence of the definition of minimal search, as minimal search does not in any way tell us how to copy the valued features on the returned head to the unvalued features that have initiated the search. The only thing minimal search does is to search for certain features in some domain. When a head bearing that feature is found, the head is returned as the output. The output will then feed other operations such as valuation.

Below, I give an example to illustrate the implementation of MS-agree. *Be* in (30) can agree with *a man*. This is a case of long-distance agreement. The steps of MS-agree are shown in (31).

(30) Example: *(there) be likely to be a man outside*

⁹This captures only the (27a) case of agree. (27b) will be discussed later in this chapter, and (27c) will be discussed in Chapter III.

¹⁰Proposals separating agree into two independent steps can be found in Arregi & Nevins (2012); Bhatt & Walkow (2013); Smith (2017) and Bjorkman & Zeijlstra (to appear). The separation can potentially solve a debate on the nature of agree: The minimal search part likely happens in syntax, but the valuation part can be done at the PF interface.

- (31)
- a. $SD = \{\text{likely}, \{\text{to be a man outside}\}\}$
 - b. $ST = \phi$ -feature attributes of the unvalued ϕ -features on *be*
 - c. Search into the SD, the target is not found
 - d. Set $\{\text{to}, \{\text{be a man outside}\}\}$ as the SD
 - e. Search into the SD, the target is not found
 - f. Set $\{\text{be}, \{\text{a man outside}\}\}$ as the SD
 - g. Search into the SD, the target is not found
 - h. Set $\{\{\text{a man}\}, \text{outside}\}$ as the SD
 - i. Search into the SD, the target is not found
 - j. Set $\{\text{a}, \{\text{man}\}\}$ as the SD
 - k. Search into the SD, the target is found on *a*, return *a*
 - l. Valuation: copy feature values to *be* from the matching features on *a*

Furthermore, a formal definition for MS-labeling is presented below:

- (32) Definition: minimal search-based labeling (MS-labeling)

The output of minimal search is taken as the label of the set to be labeled.

- i. Input to minimal search:
 - $SD = \text{the set to be labeled}$
 - $ST = \text{any feature/head}^{11}$
- ii. Output of minimal search:
 - Heads

Since the labeling algorithm is straightforward, and since we have already discussed the critical instances of labeling in (23), I will not give an additional example here.

¹¹This is equal to the statement that ST is not a set and ST is not empty.

Nevertheless, an important question regarding the current formalization of labeling is whether labels can be the input to the labeling algorithm. This question should be taken seriously because, based on our assumptions, labels are heads returned by minimal search. If they are heads, why can't they behave like regular heads in labeling? For instance, suppose the goal is to provide a label for the set $\alpha = \{\alpha X, \{\beta Y, \{\kappa Z, \dots\}\}\}$. If the label of β has already been computed, that is, Y has been taken as its label, and if we further assume that labels feed minimal search as its input for the labeling of α , α is then equivalent to $\{X, Y\}$. We now have a labeling problem, which does not exist if labels are not used as legitimate objects for labeling, as in that case the label of α should be simply X . The reason that including labels as input to MS-labeling causes a problem is because this obscures the distinction between heads and sets. The head-set distinction is an important assumption that makes labeling work in the first place. Otherwise even for simple cases such as $\{H, XP\}$, H cannot be simply returned as the label.

A possible way to avoid taking labels as input to MS-labeling is that we can simply stipulate that minimal search does not take labels as input. However, then we owe an explanation as to why labels are, on the one hand, exactly heads but, on the other hand, not usable by minimal search. A better solution could be that labels are computed top-down rather than bottom-up. That is, the labeling algorithm does not apply throughout the derivation, but only right before transfer, taking the syntactic object undergoing transfer (Chomsky 2015b) as its search domain. Take $\alpha = \{\alpha X, \{\beta Y, \{\kappa Z, \dots\}\}\}$ again as an example. The label of α is computed before the label of β . Consequently, the computation of the label for the superset α will not be affected by the label of its set member β . This is a desirable result. Such a solution is compatible with the assumption that labels are crucial for interpretation purposes at the interfaces, not for syntax. In other words, labels are not computed in the course of derivation, for example, right after each merge operation. Another

important consequence of this solution is that labels cannot be used for MS-agree as well, given the assumption that MS-agree may be initiated right after the merge of a head with unvalued features.

2.4 Theoretical implications

The minimal search-based approach has both theoretical and empirical implications. This section focuses on its theoretical implications, and the next section on its empirical consequences.

Recall that the theoretical motivation for giving a formal definition of minimal search is to evaluate the appealing conjecture proposed by Chomsky (2013, 2015a): Agree and labeling can be reduced to minimal search.

With the formal definition in (25), it is quite clear that agree and labeling cannot be completely reduced to minimal search. Our definition of MS-agree in (29) shows that agree uses minimal search to find the target. However, besides minimal search, agree must resort to the second step, valuation, in order to copy the valued features from the target to the corresponding unvalued features on the trigger. Our definition of MS-labeling in (32), on the other hand, is essentially minimal search. But still labeling is an independent linguistic operation that takes the output of minimal search as the label. In other words, labeling specifies a way to use the output of minimal search. Minimal search itself does not give instructions or put any restrictions on how to use its output, as the output could feed either agree or labeling.

In addition, there is another distinction between the definition of agree (29) and the definition of labeling (32); namely, the specific values of SDs and STs for minimal search are determined by agree and labeling independently, although an SD and ST are required/presupposed by minimal search. The assignment of the values of SDs and STs thus serves specific purposes in agree or labeling. Therefore, even if we take the valuation step in agree as a separate operation that occurs at PF and not as part

of the syntax of agree, so that both labeling and the syntactic part of agree are just minimal search, we still cannot reduce agree and labeling fully to minimal search, given that the SDs and STs are assigned by agree and labeling separately.

It is worth noting that with minimal search, we do unify a core component of agree and labeling. The search algorithms for agree and labeling are essentially minimal search. This is partially in accordance with Chomsky’s intuition regarding a possible reduction of agree and labeling to minimal search.

Therefore, I conclude that although agree and labeling both involve minimal search, they are linguistic mechanisms that are independently needed, arguably for interface interpretation purposes. If we take agree and labeling away from syntax, minimal search as a general search algorithm will not know for what and into which domain it should search.

Furthermore, as we have mentioned, if we have a fully unified labeling algorithm, we cannot on the one hand obtain a head as the label for instances such as {kick, {the, ball}}, and on the other hand return a ϕ -feature pair for instances such as {{a, {little boy}}, {is, {kicking the ball}}} (see (28)). This is because in order to return a ϕ -feature pair in the second case, the ST must be set to ϕ -features. If for both cases we have the same ST, that is, ST = any feature (not a particular one) or any head, then what is returned in the second case must be a head pair. That is, <a, is> will be returned as the label for {{a, {little boy}}, {is, {kicking the ball}}}.

There may be a question as to what the label <a, is> actually is. The same question is equally applicable to Chomsky’s (2013) < ϕ , ϕ > pair. Details aside, one possible way to interpret <a, is> at the CI interface is that this label tells us its syntactic category, as the heads contain categorical features. Therefore, <a, is> includes the information <D, T>. This is a desirable result. As argued by Shim (2018), taking categorical features instead of the ϕ -features as the label solves an independent problem caused by the < ϕ , ϕ > label. Narita (2012) points out that if

the label for a TP is $\langle \phi, \phi \rangle$, then it requires the ϕ -features on T to be visible to the CI interface, which is in contradiction with the hypothesis that unvalued ϕ -features on T are not semantically interpretable and thus must be not visible to the CI interface. MS-labelling returns a head pair instead of a ϕ -feature pair for a finite TP, naturally circumventing this problem.

A related problem is how we can force the movement of the v P-internal subject to SpecTP? Recall that Chomsky (2013) achieves this by resorting to the idea that ϕ -features in agreement could be taken as a legitimate label for TP. If the subject and T head are not in an agreement relation, the TP set is labelless. The same reasoning applies to wh-movement. In the following example, if the wh-phrase *who* moves to SpecCP, then the CP can be labeled as $\langle Q, Q \rangle$. Movement of any other syntactic objects to SpecCP renders a labelless set.

(33) $[_{CP} \text{ who}_{[+Q]} [_{C'} \text{ does}_{[-Q]} \text{ John like } \langle \text{who} \rangle]]$

However, under MS-labeling, nothing prevents a DP which is not in an agreement relation with the T or C head to stay at SpecTP or SpecCP, as the label has nothing to do with the features in agreement.

I suggest that the labeling problem related to TP and CP is in fact not a CI interface issue, but an SM interface issue, assuming that labels are input to both interfaces. That is, the problem is not that TP and CP are labelless if their two members are not in an agreement relation. Instead, the problem is whether a $\langle D, T \rangle$ pair or a $\langle D, C \rangle$ pair can be considered a legitimate SM relation. If a $\langle D, T \rangle$ pair is $\langle a, be \rangle$, where *a* and *be* are not in an agreement relation, then the pair $\langle a, be \rangle$ cannot be licensed as an SM relation. This is natural if we consider an agreement relation (especially the valuation component of agree) to also be an SM relation. The same reasoning applies to the labeling of the CP in (33); namely, the SM interface would need to know that *who* and C_{does} are in an agreement relation to license $\langle \text{who}, C_{does} \rangle$ as a legitimate SM relation, so that the SM interface can assign

a proper prosodic interpretation to *who* and C_{does} . The SM restriction on labels is summarized as below:

(34) *SM Condition on Labels* (to be revised)

$\langle X, Y \rangle$ is a legitimate label at SM only when X and Y are in an agreement relation.

(34) covers all empirical cases that Chomskyan labelling theory covers with feature pairs as labels. In addition, (34) seems to also solve the invisibility problem regarding lower copies of movement. Recall that according to Chomsky (2013), the vP -internal subject in (35) needs to move further to solve the labeling problem with vP , as vP will be labelless if the subject stays in vP , since it does not agree with v .

(35) $\{_{TP} \text{ a little boy}, \{_{T_{be}}, \{_{vP} \langle \text{a little boy} \rangle, \{_{v_{be}}, \{\text{kicking the ball}\}\}\}\}\}$

Under MS-labeling, vP will not be labelless. Instead, it has a label $\langle a, v_{be} \rangle$, which may encode a θ -relation at CI. However, $\langle a, v_{be} \rangle$ cannot be licensed at SM due to the SM Condition on Labels (34). Therefore the vP -internal subject needs to move to SpecTP. However, this is not the end of the story, because the lower copy of the subject is at Spec vP , and it is exactly the same with its higher copy at SpecTP. So Chomsky (2013) has to stipulate the following condition on lower copies:

(36) *Invisibility of Lower Copies*

Lower copies are invisible to labeling.

If the vP -internal subject is invisible to labeling, then I can hardly imagine how it becomes visible again at CI. Nevertheless, the vP -internal subject *a little boy* must be visible to CI because it encodes the original θ -role position (Samuel Epstein, p.c.).

MS-labeling offers a reasonable solution to the problem caused by the Invisibility of Lower Copies Hypothesis in (36). In fact, we do not have to assume (36). I assume, following Nunes (1995) and Bošković & Nunes (2007), that all the copies except one

must be deleted at SM as a result of the following requirement on linearization: linear order is determined by asymmetric c-command relations (Kayne 1994). Multiple copies give rise to symmetric c-command relations with other syntactic objects. For example, in the sentence (35) above, after moving from Spec*v*P to SpecTP, *a little boy* occurs in two syntactic positions. There is not an asymmetric c-command relation between *a little boy* and e.g., T_{be}, because *a little boy* at SpecTP c-commands T_{be}, but T_{be} c-commands the lower copy of *a little boy* at Spec*v*P. This causes a linearization problem. A way to avoid this linearization problem is to delete all but one of the copies at SM.

Turning back to the example in (35), if the lower copy at Spec*v*P is deleted at SM due to linearization issues, resulting in all its phonological features being deleted, leaving syntactic features such as categorial features unaffected, then the label for *v*P after the movement of the *v*P-internal subject is $\langle D, v_{be} \rangle$. I suggest that $\langle D, v_{be} \rangle$ composes an exception to the SM Condition on Labels, as D does not have any phonological content, and should not be subject to such an SM condition. We can now revise the SM Condition on Labels to make it explicit that cases such as $\langle D, v_{be} \rangle$ are not subject to the SM condition.¹²

(37) *SM Condition on Labels* (to be further revised)

Given two heads X, Y, if X, Y have phonological content then $\langle X, Y \rangle$ is a legitimate label at SM only when X and Y are in an agreement relation.

However, the revised version of the SM Condition on Labels still encounters a problem with the following sentences:

- (38) a. [_{*v*P} David/Him/*He being the chairman] has meant more work for all of us.
b. [_{*v*P} Those people/Them/*They being the committee members] has meant more work for all of us.

¹²See Takita (to appear) for a similar way to derive the Invisibility of Lower Copies Hypothesis.

It is important to note that the v Ps marked in (38) are problematic for Chomsky's (2013) labeling theory. This is because the v P-internal subjects do not move to SpecTP but remain in the Spec v P positions; nevertheless, they do not agree with the v heads. Chomsky's (2013) labeling theory predicts that these cases must be filtered out by labeling failures, as no $\langle \phi, \phi \rangle$ in an agreement relation can be taken as the label of the v Ps.

It seems reasonable to assume that either the ϕ -features on the v heads in (38) are underspecified or that the ϕ -features on the DPs at Spec v P are not accessible for agreement because they bear a default case, as indicated by the pronouns that must be marked with the default accusative case in (38). Both of these assumptions have the following consequence: there is no possibility for the v P-internal subjects and the v heads to form an agreement relation in the first place. In this case, a $\langle D, v \rangle$ label also does not cause a problem at SM.

Existential constructions cause a similar problem:

(39) [TP There is [SC [DP a child] [PP in the room]]].

The DP *a child* stays in the specifier position of the small clause SC without a need to move to a higher position. This is a problem to Chomsky's labeling theory, as there is no (overt) agreement between D *a* and P *in*. However, (39) is not a problem to my analysis, because P does not bear ϕ -features in English and it cannot be involved in an agreement relation with a D head, exempting the label $\langle D, P \rangle$ from the SM Condition on Labels.

In addition, why can *there* in (39) stay at SpecTP? Is there an agreement relation between *there* and *is*? It may be argued that *there* is inserted after syntax. This will solve the problem for Chomsky's labeling theory. (39) is not a problem for MS-labeling either, if we assume that *there* does not bear ϕ -features and thus cannot be involved in an agreement relation with *is*.

To capture the above observations, the SM Condition on Labels is revised to its final version in (40).

(40) *SM Condition on Labels* (final)

Given two heads X, Y, if X, Y are able to be in an agreement relation then $\langle X, Y \rangle$ is a legitimate label at SM only when X and Y are in an agreement relation.

With the SM Condition on Labels as formalized in (40), can we exclude (41)¹³?

(41) * $[_{TP1}$ There $[_{T'1}$ seems $[_{TP2}$ $[_{DP}$ a man] $[_{T'2}$ to be outside]]].

EKS (Epstein et al. 2014) suggest that (41) can be filtered out by labeling failure, because *a* and *to* in TP2 are not in an agreement relation and thus TP2 cannot be properly labeled. However, the SM Condition on Labels (40) predicts that $\langle a, to \rangle$ is a legitimate label at SM, because *to* does not bear ϕ -features and therefore it cannot be in an agreement relation with *a*. Consequently, instead of resorting to labeling failure, I draw on case theory to filter out (41), since *a man* does not have case.¹⁴

Summarizing what we have just discussed, the following cases are exemptions to the SM Condition on Labels as the two heads involved are not able to be in an agreement relation in the first place: (i) either X or Y is phonologically null, (ii) the relevant agreeing features on either X or Y are underspecified or for some reason not accessible to agreement.

Another example for Case (ii) where (40) is not applicable comes from languages that have quirky subjects (e.g., subjects that are dative case marked). (42a) indicates that the verb/T agrees with the subject when it bears a nominative case, but this

¹³I should thank Samuel Epstein for bringing up this example.

¹⁴I then have to assume that in (i) below, the null head of the small clause (represented as X) can (optionally) assign case to *a man* (see Chapter III for a formal analysis of small clauses built upon Kayne (1984)).

(i) There seems to be $[_{SC}$ a man $[X$ outside]].

A similar analysis applies to the small clause in (39).

agreement does not occur when the subject is a quirky subject; that is, when the subject bears a dative case (42b).

(42) *Icelandic* (Sigurðsson 1996:1)

- a. Strákunum leiddist/*leiddust.
 boy.the.PL.DAT were.bored.3SG/*3PL
 ‘The boys were bored.’
- b. Strákarnir leiddust/*leiddist.
 boy.the.PL.NOM walked.hand.in.hand.3PL/*3SG
 ‘The boys walked hand in hand.’

The purpose here is not to provide an explanation as to why the quirky subject cannot agree with T. Instead, the focus is on the problem that (42b) imposes for Chomsky’s (2013) labeling theory. The quirky subject is in the SpecTP position, but the TP cannot be labeled as $\langle \phi, \phi \rangle$, as no agreement relation holds between SpecTP (more accurately, its D head) and the T head. However, (42b) is not as problematic for the current MS-labeling, because the TP will be labeled as $\langle X, T \rangle$, where X represents the highest head in the dative case-marked subject, whatever it is. This label does not cause a problem at SM because the ϕ -features on X are normally assumed to be not accessible for agreement, exempting (42b) from the SM Condition on Labels as formalized in (40).¹⁵

Finally, besides the points we mentioned above, we have also pointed out that in order to keep the head-set distinction, labels cannot feed minimal search, and therefore labels cannot be used for MS-agree. As a result, a system that resorts to the idea that the label of a set functions as a head for agree purposes is problematic theoretically. For instance, to account for cyclic agree, Béjar & Rezac (2009) suggest that the label of v' , i.e., the v head, can act as a head with unvalued features that

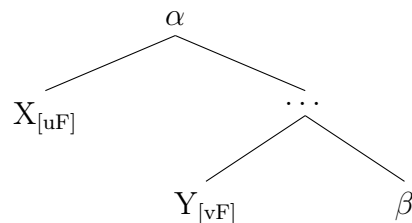
¹⁵Different proposals have been made to explain why no agreement/default agreement is triggered with quirky subjects in Iceland (see Preminger 2014: Ch. 8 for a review). A review of these approaches is beyond the scope of this dissertation, but it is worth noting that a common property of all these proposals is that the ϕ -features on quirky subjects are inaccessible for agreement.

probes into its sister, SpecvP. Such analysis loses the set–head distinction and cannot be formalized under MS-labeling. The empirical basis of the cyclic agree analysis can be reanalyzed with MS-agree, as I will argue in the next section.

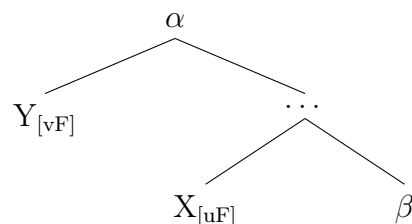
2.5 Empirical consequences

We have mentioned that there are three logically possible configurations in which a head X with an unvalued feature can be related to and agree with another head Y with a corresponding valued feature, as listed in (43), repeated from (27) (I add tree diagrams where dots indicate optional additional structure):

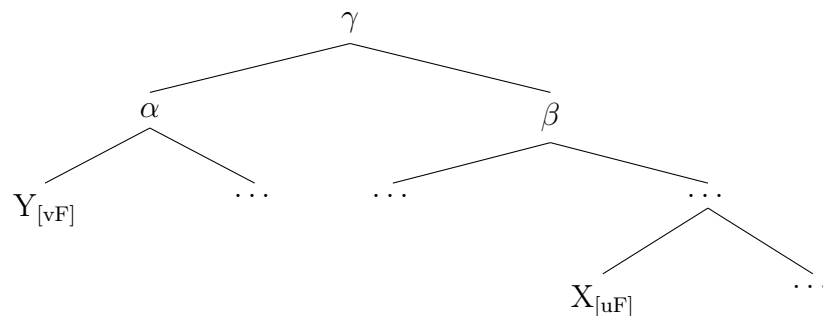
- (43) a. $X_{[uF]}$ merges with $Y_{[vF]}$ or a set containing $Y_{[vF]}$: $X_{[uF]}$ c-commands $Y_{[vF]}$;



- b. $Y_{[vF]}$ merges with $X_{[uF]}$ or a set containing $X_{[uF]}$: $Y_{[vF]}$ c-commands $X_{[uF]}$;



- c. A set containing $Y_{[vF]}$, α , merges with a set containing $X_{[uF]}$, β .



MS-agree is an instance of downward search. The downward agreement configuration

as in (43a) therefore does not cause any problem to MS-agree. However, the upward agree configuration in (43b) does raise a problem. What's more, (43c) becomes a serious problem if we do not assume feature percolation and projection, as neither $X_{[uF]}$ c-commands $Y_{[vF]}$ nor $Y_{[vF]}$ c-commands $X_{[uF]}$, whereas the upward agree analysis requires $Y_{[vF]}$ to c-command $X_{[uF]}$, and the standard downward agree requires $X_{[uF]}$ to c-command $Y_{[vF]}$. I will address the problems raised by the configurations in (43b) in the rest of this section, and leave (43c) to the next chapter, as such a configuration is observed in reflexive binding.

2.5.1 Upward agree

The main goal of this subsection is to address the debate on the direction of agree reviewed in Section 2.2.3. I argue, based on a unique feature of our definition of minimal search, that upward agree can be re-implemented as an instance of MS-agree.

In our review of the direction of agree in Section 2.2.3, negative concord, inflection doubling/parasitic participle and multiple agree provide strong evidence for the existence of upward agree. I repeat one example from each of these constructions below, coupled with corresponding syntactic analyses:¹⁶

(44) *Czech* (Zeijlstra 2012:501)

- a. Dnes nikdo *(ne)volá nikomu
 Today n-body NEG.calls n-body
 'Today nobody is calling anybody.'
- b. $[_{CP} [_{TP} Op \neg_{[vNEG]} [_{TP} nikdo_{[uNEG]j} nevolá_{[uNEG]} [_{vP} t_j nikomu_{[uNEG]}]]]]]$

(45) *Norwegian* (Wiklund 2001:201)

- a. Jeg hadde villet [lest boka]
 I had want.PART **read.PART** book.DEF

¹⁶I have replaced interpretable features in the original representations, e.g., $[iNEG]$, with corresponding valued features, e.g., $[vNEG]$, to make the syntactic representations consistent with the framework this dissertation adopts (see Chapter IV for a further discussion on the feature system this dissertation adopts). In addition, we have abstracted away from the actual valued and unvalued features involved in (45) and (46), to simplify the discussion.

‘I would have liked to read the book.’

- b. $[_{CP} \dots [_{AuxP} \text{had}_{[vF]} \text{want.PART}_{[uF]} [_{VP} \text{read.PART}_{[uF]}]]]$

(46) *Japanese* (Hiraiwa 2001:76)

- a. John-ga [yosuijouni nihonjin-ga eigo-ga hidoku]
 John.NOM than-expected the-Japanese.NOM English.NOM bad.INF
 kanji-ta.
 think.PAST

‘It seemed to John that the Japanese are worse at speaking English than he had expected.’

- b. $[_{CP} \dots \text{DP1.NOM}_{[uF]} [_{\dots \text{DP2.NOM}_{[uF]} \dots \text{DP3.NOM}_{[uF]} \dots \text{V.INF}}] \text{T}_{[vF]}]$

For all these examples except the negative concord in Czech, as indicated by the syntactic representations in the (b) part of each example, these attested cases of upward agree are all phase-constrained. That is, the same minimal phase (the CPs in the examples) dominates both the valuee carrying the unvalued features and the valuer bearing the corresponding valued features. This gives us a hint on how to model these cases of upward agree with MS-agree. If minimal search looks into the CP phase and searches downward, then we obtain the desired results.

For the negative concord in Czech (44a), the syntactic analysis (44b) adopted from Zeijlstra (2012) indicates that the lowest n-word *nikomu* is in the *vP* whereas the negative operator is located in the CP phase. The valuation of the object and the negative operator crosses a *vP* phase. A slight revision, that is, positioning the negative operator in *SpecvP* instead, will give us a configuration similar to other examples. As far as I know, there is no reason to reject this alternative analysis.

- (47) $[_{CP} [_{TP} \text{nikdo}_{[uNEG]j} \text{nevolá}_{[uNEG]} [_{vP} \text{Op}\neg_{[vNEG]} [_{vP} \text{t}_j \text{nikomu}_{[uNEG]}]]]]]$

Given (47), the $[uNEG]$ on the verb *nevolá* can be valued by the operator through regular downward search. But the $[uNEG]$ features on the object *nikomu* and the subject *nikdo* can be valued by the operator if minimal search looks down from the

*v*P phase.

Recall that the reason why our definition of minimal search can partially unify agree and labeling is that the values of SD and ST are independently assigned by relevant operations. This unique feature of the definition can probably help us connect the valuee (the head with unvalued features) with the valuer (the head with corresponding valued features) in (44) to (46).

More specifically, I assume that in the above cases of upward agree, the SD is not the sister or co-member of the valuee. By contrast, if we still want to find the valuer by downward search, the SD in cases of upward agree must be sets that include the *v*P (in 44), the AuxP (in 45), and the TP (in 46). Therefore, as long as these examples are concerned, it could be either the *v*P or CP that is assigned as the SD of minimal search for these cases.¹⁷ Minimal search then searches downward until the head with the target is found.

An immediate question is whether there is a reason why *v*P and CP could be assigned as the SD of minimal search here. A reasonable conjecture is that generally, the phase status of *v*P and CP may be relevant. This is because, given the phase theory that I adopt in this dissertation, when a phase is built, the lower phase needs to be transferred. If a valuee remains unvalued in the phase that is subject to transfer, as in the upward agree cases, it will cause the derivation to crash. The [uNEG], the unvalued feature associated with the participle morpheme, and the unvalued feature connected to the nominative case morpheme are all uninterpretable to the interfaces. Therefore, minimal search taking the next higher phase dominating the valuee is the last chance for the unvalued features to find a valuer if the valuee is inside the lower phase head complement, otherwise the valuee will be transferred. Although if the valuee is in the edge of the lower phase, it will not be transferred until the next two higher phases being built, as a general mechanism, minimal search must be triggered

¹⁷TP is less relevant as (46) shows that infinitive cannot be the domain that constrains upward agree.

when the next higher phase above the valuee is built. In addition, compared to any other set that is contained in the next phase, the next built phase (containing the valuee) is the set that contains the most head members, and therefore it is the SD within which the minimal search is most likely to find a valuer.

If the phase status of vP and CP is relevant here, given that the heads with unvalued features are in the c-command domain of the valuers, we can formalize the locality constraint in upward agree cases by proposing that the next built phase above the valuee is taken as the SD for the relevant minimal search:

- (48) In upward agree examples, minimal search takes as its SD the next phase built above the valuee.

With the simple assumption in (48), the cases of upward agree can be perfectly captured by MS-agree. For the negative concord in Czech in (44) with an analysis as in (47), *nikdo* and *nikomu* as well as the verb *nevolá* trigger separate minimal searches with the same SDs and STs. The SDs are the vP , and the STs are the [uNEG] feature. Minimal search looks into the vP and finds the Op with the [vNEG] feature that matches the STs. Consequently, the [uNEG] features on all these triggers can be valued by the [vNEG] feature on the Op.

Likewise, in (45), the unvalued feature on the verb bearing the parasitic participle can be valued by the auxiliary *hadde* ‘had’ by downward searching from the CP. In (46), the unvalued case feature on the DPs can also be valued by the T head through multiple independent minimal searches with SD = CP.

Therefore, MS-agree provides a neat solution to cases where upward agree is claimed to be necessary, with the assumption in (48). This gives rise to two questions. The first question is: since these cases of upward agree are not standard ϕ -feature (e.g., negative concord), should they be treated as non-syntactic concord and thus be treated differently from regular ϕ -feature agreement? If this is the case, then we should not try to account for upward agree and downward agree with a single mech-

anism, and it is even misleading to do so (Preminger & Polinsky 2015). Bjorkman & Zeijlstra (to appear) address this question by pointing out that treating concord phenomena such as negative concord and inflection doubling as non-syntactic would fail to explain their locality constraints. If they are syntactic in nature, then there is no reason in principle why these phenomena of concord should be treated separately from ϕ -feature agreement.

Even if we leave the locality constraints aside, it seems at least inflection doubling should be taken as a syntactic phenomenon and be better treated as a case of agree-ment. This is because the parasitic participles are interpreted as infinitives rather than perfective participles. Nevertheless, they are pronounced as perfective participles. Syntactic selection may determine the semantic interpretation of the parasitic participles (as it determines the participles' syntactic category); however, it cannot explain why the parasitic participles can be pronounced as such. With agree, this possibility is readily available because the participle marker on parasitic participles can be a pronounced unvalued phonological feature.

To provide an even stronger argument, however, I will discuss in the next two subsections cases where upward ϕ -feature agreement is attested, and show that even in those cases, MS-agree with the assumption in (48) still offers a simple analysis. These upward ϕ -feature agreement, however, cannot be fully captured by the upward agree analysis.

Another question that I will address in the next section is when and why minimal search searches into the next built phase containing the valuee.

2.5.2 Cyclic agree

In this subsection, I use cyclic agree to illustrate that MS-agree provides a neat analysis to upward ϕ -feature agreement. Cyclic agree also gives us a clue about when and why minimal search takes as its SD the next built phase containing the valuee

in cases of upward agreement.

Béjar & Rezac (2003); Rezac (2004) and Béjar & Rezac (2009) investigate an interesting agreement pattern in Georgian: agreement occurs in two cycles. In the first cycle, a head with unvalued features (the trigger) probes downward to find a head matching the unvalued features. If the search in the first cycle is not successful, the trigger may instead search upward. Let us see some examples.

(49) Georgian (Rezac 2004:72)

- a. m-xedav-t
1-see-PL
'You(.PL) see(.PL) me.'
- b. g-xedav-t
2-see-PL
'I see(.PL) you(.PL).'

Béjar (2003) and Béjar & Rezac (2003) assume that person and number features are represented separately in Georgian: *v* bears a [uPerson] feature and T bears a [uNum] feature. (49a) shows that, when the subject is plural, T_[uNum] agrees with the subject and *v*_[uPerson] agrees with the object. By contrast, (49b) exemplifies that, when the subject is singular, *v*_[uPerson] still agrees with the object, but T_[uNum] now agrees with the object instead of the subject. Béjar and Rezac suggest that the singular number feature on the subject is underspecified or not present in syntactic representations in Georgian, so downward searching from T_[uNum] skips the subject and expands the search domain down to the object. The intervention effects caused by the subject to the number agreement in (49a), compared to (49b), strongly support the idea that this is a case of downward search.

Let us ignore the issue regarding what it exactly means for a head to be underspecified for a certain feature (see Preminger 2011 for an alternative analysis). The examples in (49) illustrate the first cycle agree, which is an instance of downward

agree. The following examples show the second cycle agree, which is more important for our purposes:

(50) *Georgian* (Rezac 2004:73)

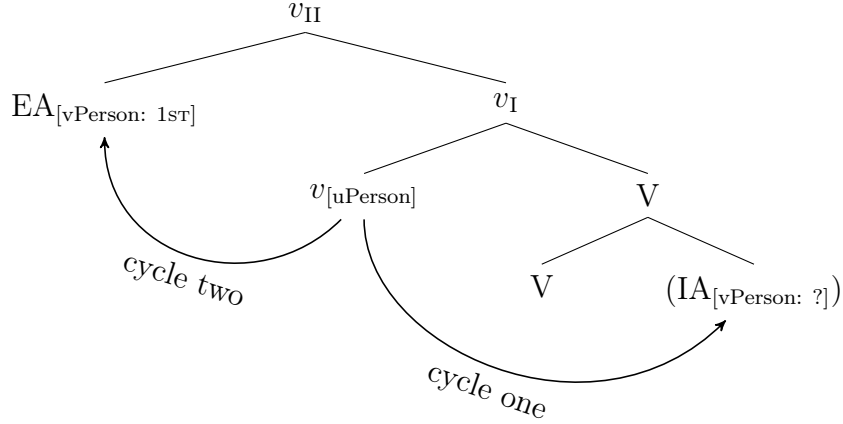
- a. v-k'vd-eb-i
1-die-x-ind.PRS
'I am dying.' (originally from Hewitt 1995:118)
- b. mi-v-c'er-e
X-1-write-AOR
'I wrote it to her.' (originally from Hewitt 1995:190)

(50a) indicates that when the object is absent, $v_{[uPerson]}$ agrees with the subject instead of the object, differently from what we have seen in (49). Similarly, in (50b), where the object is an underspecified third person pronoun, $v_{[uPerson]}$ also agrees with the subject. Béjar & Rezac (2009) argue that the reason $v_{[uPerson]}$ can agree with the subject when the object is not available for agreement is that agree can change its direction when necessary. That is, in the second cycle, given that nothing can be returned in the first cycle downward search, upward search is initiated and the subject is returned. As shown below in Figure 2.5, in cycle one, $v_{[uPerson]}$ looks for the object/internal argument (IA). If IA's person feature is underspecified (indicated by the question mark) or if the IA is absent (indicated by parentheses), the second cycle agree is initiated. $v_{[uPerson]}$ searches upward to agree with the subject/external argument (EA).

Crucially, Béjar & Rezac (2009) propose that the second cycle agree could be regular downward search if we assume that the label of v_I and v_{II} are both the v head. That is, $v_{[uPerson]}$ can probe down from the v_{II} projection as v_{II} is just $v_{[uPerson]}$.

This analysis builds upon Chomsky's (1995b; 2000a) labeling theory, where merge of two syntactic objects α and β produces a label, and the label can be either α or β , depending on which one syntactically selects the other (setting adjunction/pair merge aside). Béjar & Rezac (2009) move a step further: the label can initiate a search for

Figure 2.5: Béjar & Rezac’s (2009) analysis of cyclic agree



agree, like a regular head. This analysis thus identifies regular heads with labels in the computation of agreement.

This is a clever use of the labeling theory proposed by Chomsky (1995b, 2000a). However, as previously mentioned, such an approach does not work under the current formalization of agree and labeling, following Chomsky (2013, 2015b). This is because labels cannot be input to minimal search and thus cannot feed agree; otherwise the head–set distinction is lost. Instead, labels are computed only right before syntactic objects are sent to the interfaces.

In addition, Béjar & Rezac’s (2009) analysis faces the difficulty that the whole vP (v_{II}) cannot be labeled as v until the EA moves to a higher position, SpecTP for example, as vP is an instance of $\{XP, YP\}$. This implies that it is only after the TP is built that we can come back to the v_{II} to probe for the lower copy of the EA. Note, however, that after movement, the lower copy of the EA is standardly assumed to be invisible for agree.

A similar cyclic agree analysis has also been argued for Hindi-Urdu. Keine & Dash (2017) (see also Keine & Dash 2018) call attention to data such as (51):

- (51) *Hindi-Urdu* (Keine & Dash 2017:3)

saare shikṣak [raam-ko kitaab paṛhne] **dete** hāĩ
all teachers.M Ram-DAT book.F read.INF **let.M.PL** AUX

‘All the teachers let Ram read a book.’

We see that in (51), the verb *dete* ‘let.M.PL’ (or the matrix T) regularly agrees with the matrix subject, i.e., *saare shikṣak* ‘all teachers.M.’ The embedded object *kitaab* ‘book.F,’ being located in a lower position, cannot agree with the matrix T since the subject intervenes. However, when the matrix and the embedded subject are both case-marked and thus inactive for ϕ -agreement (Chomsky 2001b), T can probe further and agree with the embedded object, as is shown in (52).

(52) *Hindi-Urdu* (Keine & Dash 2017:3)

shikṣakō-ne [raam-ko **kitaab** paṛhne] **dii**
 teachers.M-ERG Ram-DAT **book.F** read.INF **let.F.SG**

‘The teachers let Ram read a book.’

This is the first cycle agree, where T searches downward for a target and an overtly case-marked intervener can be skipped; otherwise the intervener causes an intervention effect.

Keine & Dash (2017) also detect the second cycle agree in Hindi-Urdu when the first cycle downward search returns nothing. (53) shows that the object undergoes A-movement by scrambling to the matrix SpecTP position, and thus binds the pronoun (as indicated by the indices).

(53) *Hindi-Urdu* (Keine & Dash 2017:4)

har kitaab₂ [unke₂ lekhakō-ne] t₁ **dii** [raam-ko t₂ paṛhne]₁
every book.F its authors-ERG **let.F.SG** Ram-DAT read.INF

‘For every book *x*, *x*’s authors let Ram read *x*.’

In this case, the matrix T cannot find a matching target by the first cycle downward search. Therefore, the second cycle agree is initiated and the A-scrambled SpecTP agrees with T.

In comparison, an \bar{A} -scrambled DP cannot feed the second cycle agree against T:

- (54) **har kitaab₂** [unke_{3/*2} lekhakō-ne] t₁ kahaa/***kahii** [raam-se t₂
every book.F its authors-ERG say.DFLT/***say.F.SG** Ram-INSTR
 parhne-ko]₁
 read.INF-DAT
 ‘Its₃ authors told Ram to read every book₂.’

(54) is taken to be a case of \bar{A} -scrambling because the scrambled DP does not feed the binding of the pronoun. If the pronoun is forced to be co-indexed with the scrambled DP, a crossover effect is observed. Therefore, only default marking is possible.¹⁸ Keine & Dash (2017) assume that the \bar{A} -scrambled DP is ultimately positioned at the matrix SpecCP position. This is supported by the following contrast:

- (55) *\bar{A} -scrambling in Hindi-Urdu* (based on (26) in Keine to appear:17)

[_{CP} **kitaab-ko₁** mǎĩ caahtaa hũũ [_{TP} ***kitaab-ko** kahnaa [_{CP} ki mǎĩ-ne
book-ACC I want AUX ***book-ACC** say.INF that I-ERG
 t₁ parhaa hai]]]
 read AUX

‘I want to say that I read the book.’

- (56) *A-scrambling in Hindi-Urdu* (Keine to appear:18)

Sita-ne caahaa thaa [_{TP} **har larḱii-ko₁** [uskii₁ shaadii ke-dauraan] t₁
 Sita-ERG wanted AUX **every girl-ACC** her wedding during
 dekhnaa]
 see.INF

‘Sita wanted to see every girl *x* at *x*’s wedding.’

(55) is an instance of \bar{A} -scrambling, not A-scrambling, as the scrambling in this case is from a finite clause (whereas A-scrambling in Hindi-Urdu is finite clause bounded), and it cannot target SpecTP of the embedded infinitive TP, but can end up at the matrix SpecCP. On the contrary, since the scrambling in (56) is from an infinitive TP and the scrambled DP can bind the pronoun, it is a case of A-scrambling. (55) and

(56) thus suggest that A-scrambling can target the SpecTP position of a finite TP, whereas \bar{A} -scrambling cannot (see Keine to appear for more details of the argument).

Based on the contrast between (53) and (54), Keine & Dash (2017) argue that in the second cycle agree, the search direction is reversed to be upward, and thus T can agree with the A-scrambled subject. However, it is unclear why upward search cannot reach the \bar{A} -scrambled DP at SpecCP in (54).¹⁹

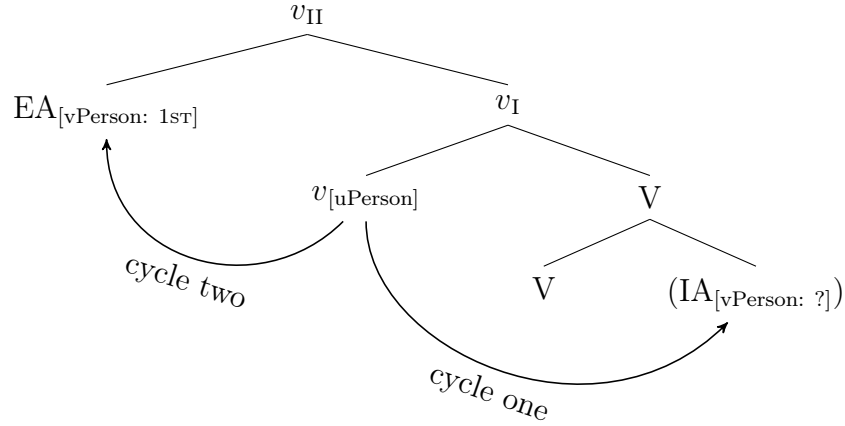
Keine & Dash (2018) suggest that a proposal adopted from Béjar & Rezac (2009) can probably explain why T can agree with its own specifier but cannot agree with SpecCP. As we have mentioned, in Georgian (and some other languages), *v* normally agrees with the IA (the object) (the so-called first cycle agree). However, when the IA is absent or its person feature is underspecified, *v* agrees upward with a local EA (the *v*P-internal subject) instead of the IA (see (49) and (50)). To account for this type of agreement (the so-called second cycle agree), Béjar & Rezac (2009) claim that the direction of agree is not in fact reversed from downward to upward. Instead, in the second cycle, it is the label of *v*, which dominates the EA, searches down, as has been illustrated in Figure 2.5, repeated below in Figure 2.6.

Building upon Béjar & Rezac (2009), Keine & Dash (2018) suggest that $X_{[uF]}$ can agree with $Y_{[vF]}$ if and only if X (either as a head or as a label) c-commands Y. Back to the Georgian examples in (50), the schema of the structures relevant to person agreement is the one shown in Figure 2.6. For Keine & Dash (2018), *v* can value the person feature on the IA if the IA is available; however, when the IA is absent or is

¹⁸Keine & Dash (2017) consider default marking a result of agreement failure (see also Preminger 2011, 2014).

¹⁹This analysis hinges on assumptions that A and \bar{A} scrambling either directly targets the SpecTP or SpecCP position without going through intermediate positions such as the matrix specifier of *v*P, or an alternative assumption that even if there are intermediate copies of the movers, T somehow does not agree with the intermediate DPs that will undergo further scrambling. The former assumption is what is presented and defended in Keine & Dash (2018). Although they do not exclude the latter analysis, Keine & Dash (2018) do convincingly argue against the possibility that T only agrees with an intermediate DP that will undergo further A-movement but does not agree with one that will undergo further \bar{A} -movement, thus discrediting a potential explanation to the A and \bar{A} difference between (53) and (54) that the latter assumption is likely to provide.

Figure 2.6: Béjar & Rezac’s (2009) analysis of cyclic agree



case-marked, the label v_I can agree with the EA, as v_I c-commands the IA.

It is important to note here that an advantage of Béjar & Rezac (2009) and Keine & Dash (2018), inherited from Chomsky (1995b, 2000a), is that there is no need to stipulate feature percolation. Instead, labels naturally provide the head that can agree with its specifier. In addition, an attractive feature of Béjar & Rezac and Keine & Dash’s system is that the locality of upward agreement is well-captured. For Béjar & Rezac, the controller or target of the agreement must be dominated by the maximal projection of the agreeing head. For Keine & Dash (2018), on the other hand, the controller must be either c-commanded or dominated by the agreeing head. As a result, Keine & Dash (2018) can explain why in (54) the \bar{A} -moved DP, being at SpecCP, cannot agree with the T head, given that TP neither c-commands nor dominates the \bar{A} -moved DP.

However, both Béjar & Rezac (2009) and Keine & Dash (2018) identify labels with regular heads and thus labels can function as probes. This, by implication, identifies heads with sets. We have pointed out the loss of set–head distinction brings in several undesirable theoretical consequences to the labeling theory. In fact, Béjar & Rezac and Keine & Dash’s assumption simply renders the labeling algorithm useless (see

Rizzi 2016). Furthermore, I would like to mention here that empirically, the identification of regular heads and sets/phrases is at odds with all studies that suggest a distinction between heads and sets, including the distinction between head movement and phrase movement (e.g., Chomsky 2001b:37–38, and the citations therein). Therefore, the solution provided by Béjar & Rezac (2009) and Keine & Dash (2018) should be reformulated under the current framework.

Now we are ready to offer our re-implementation of cyclic agree with MS-agree as defined in (29). I adopt the idea that agree occurs in two cycles. The single, crucial difference between these two cycles is the search domain (and correspondingly, the timing) of minimal search. In the first cycle, the search domain is the sister/co-member of the trigger (the head bearing unvalued features that initiate the search), as shown in (57), with a tree diagram in Figure 2.7. The box indicates the SD.

(57) *Cycle One Agree*

SD = $X_{[uF]}$'s sister = β

Timing: when $X_{[uF]}$ enters in the derivation

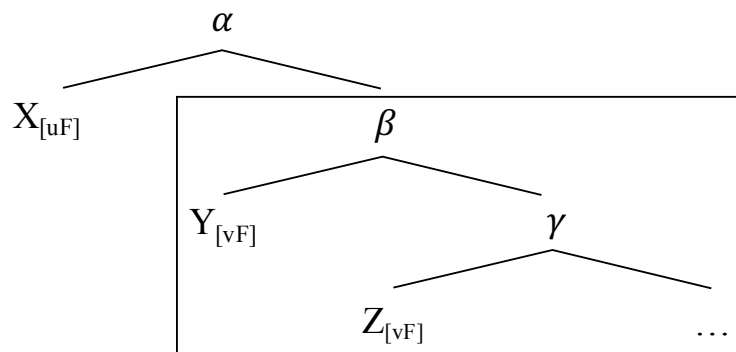


Figure 2.7: Cycle One Agree

This corresponds to the first cycle agree in Béjar & Rezac (2009) and Keine & Dash (2018), where normally $Y_{[vF]}$, which c-commands $Z_{[vF]}$, causes an intervention to Z. That is, only Y will agree with X. But in the cases where the relevant feature

on Y is underspecified, minimal search can skip Y and find Z instead.

On the other hand, in the second cycle, the search domain should be a set containing both the trigger and the search target. We have mentioned in Section 2.5.1 that all the attested cases of upward agree can be accounted for if we assume the search domain is the next phase that to be built or completed. This solution can be exactly carried over to cyclic agree discussed in this section, as shown below in (58), with a tree diagram in Figure 2.8.

(58) *Cycle Two Agree*

SD = the next phase ρ

Timing: when the lexical array for ρ is exhausted

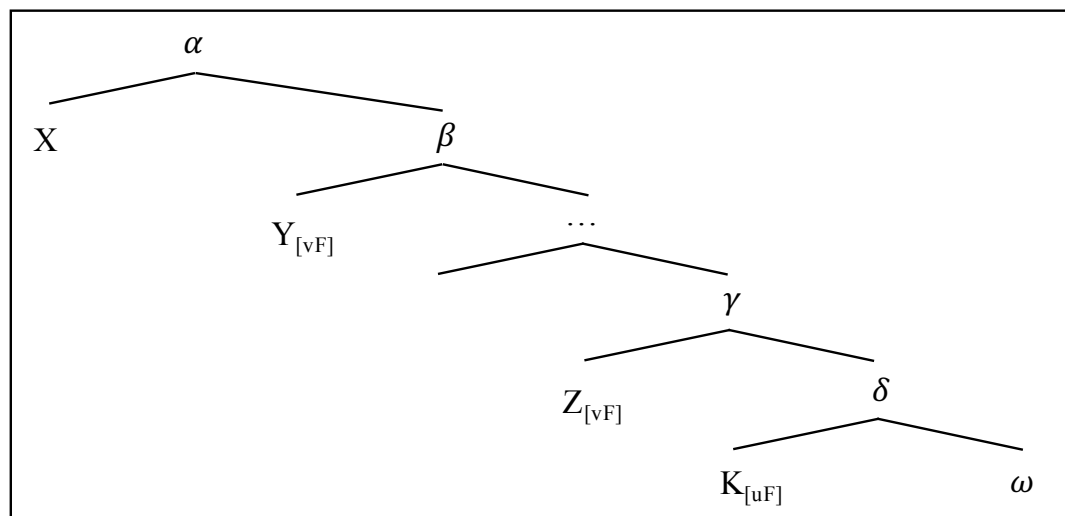


Figure 2.8: Cycle Two Agree

Here we further specify the timing of the second cycle agree, namely, at the time when the lexical array for the next phase is exhausted. This captures the contrast between A-movement and \bar{A} -movement: The former but not the latter feeds the second cycle agree. In the example of A-movement (53), the structure of which is shown below in Figure 2.9, when the DP moves to SpecTP, the C head has not yet entered into the derivation. The lexical array for the matrix CP is exhausted after

the C head merges in. This is why in Figure 2.9, where the SD is the matrix CP, the scrambled object in SpecTP can be found by minimal search and it thus agrees with T.

Note that with the definition of Cycle Two Agree in (58), the locality constraint of upward agree in Georgian can also be captured. Since the *v* head as a trigger for the second cycle agree must initiate a minimal search after the local *v*P is completed, Spec*v*P will always be returned as long as such a Spec*v*P exists. We will see a case where the trigger is a phase head and it does not have a specifier; in that case, non-local agreement is predicted, a prediction that will be confirmed by the complementizer agreement in Lubukusu.

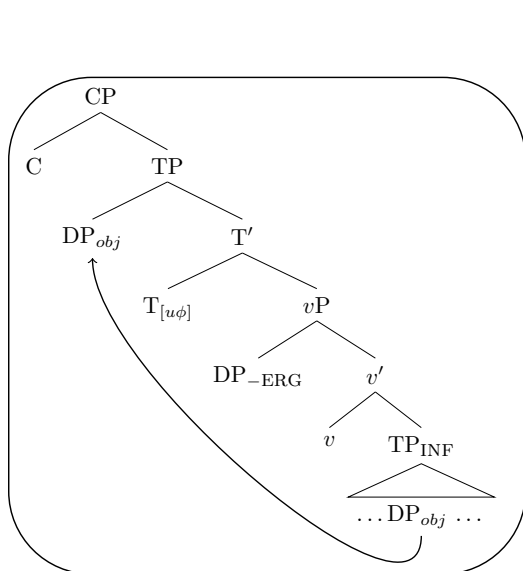


Figure 2.9: A-movement

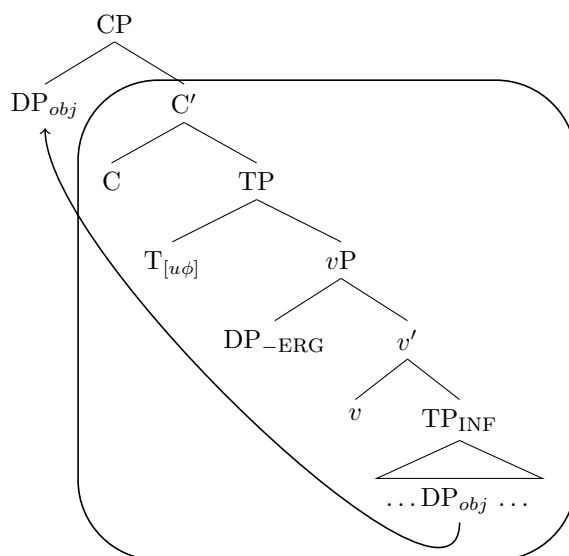


Figure 2.10: \bar{A} -movement

Figure 2.10 shows the structure of the \bar{A} -movement example. It indicates that the lexical array for the matrix CP phase is exhausted when C merges in, and therefore C' is taken as the search domain for the second cycle agree, with the \bar{A} -scrambled DP excluded. This explains why the \bar{A} -scrambled DP cannot agree with T.²⁰

²⁰It seems I should assume that both the first cycle and second cycle agree can be initiated only once, and in Figure 2.10, T obtains default agreement after the second cycle agree fails.

2.5.3 Subject–complementizer agreement

The last type of phenomenon I want to discuss is subject–complementizer (S–C) ϕ -feature agreement in Lubukusu, a Bantu language. The reason S–C agreement in Lubukusu is pertinent is that it is an instance of long-distance upward agree. As Preminger & Polinsky (2015) point out, the most convincing evidence for upward (ϕ -feature) agree is long-distance upward agree²¹ that crosses a finite TP (or a CP) clause boundary; namely, a head with unvalued features triggers agree with a target in a higher TP where the target receives its case or θ -role. This is because, due to derivational opacity, we can neither justify nor falsify upward agree if at least in some circumstances there is covert movement of the trigger to a higher position c-commanding the target; it is also quite unlikely that we can completely exclude the possibility that the target is in fact raised from a lower position c-commanded by the trigger in an earlier derivation stage. In both of these cases, the trigger c-commands the target at certain point of the derivation, an agreement pattern that can be handled with downward agree. However, with long-distance upward agree, covert movement of the trigger that crosses a finite TP, as an instance of head movement, is widely considered unlikely. Raising of the target to a position in a higher TP where it receives its case or θ -role, as an instance of A-movement, cannot be easily justified either. Therefore, long-distance upward agree cannot easily be formalized as instances of downward agree, and thus strongly argues in favor of upward agree.

More importantly, as we have mentioned in the discussion of cyclic agree, local upward agree can usually be captured by the Spec–head relation, which can in turn be formalized as downward agree if Chomsky’s (1995b; 2000a) labeling theory or a theory of projection is assumed, although we have argued that these assumptions are in contradiction with the labeling theory and encounter empirical problems.

²¹I use “long-distance agree” for long-distance downward agree, like one can usually find in the literature. Now I am using “long-distance upward agree” to distinguish it from long-distance downward agree.

Unfortunately, examples of long-distance upward agree are extremely rare in the literature. Preminger & Polinsky (2015) thus take it as an unattested agreement pattern in human languages.

However, at least at the surface form, S–C agreement in Lubukusu is an instance of long-distance upward agree, as the embedded C head agrees with the immediately superordinate subject. Preminger & Polinsky (2015) are aware of this type of upward agreement, but following some previous analyses of S–C agreement in Lubukusu, e.g., Diercks (2013), they assume that S–C agreement in Lubukusu can be successfully explained by an analysis that does not resort to upward agree. I argue below that such a non-agree analysis encounters many difficulties and is not tenable; on the other hand, MS-agree gives us a much simpler analysis that is free of these problems.

Diercks (2013) argues convincingly that in Lubukusu an embedded C agrees only with the subject, not the object, of the immediately superordinate clause in full ϕ -agreement, in terms of noun class agreement. Exemplified in (59), the embedded C head *-li* agrees with the matrix subject (*ba-ba-*)*ndu*, rather than with the object *Alfredi*, although the latter is structurally closer to the embedded C.

- (59) Ba-ba-ndu ba-bol-el-a Alfredi ba-li a-kha-khil-e.
 2-2-people 2SA²²-said-APL-FV 1Alfred 2-that 1SA-FUT-conquer
 ‘The people told Alfred that he will win.’

Further evidence shows that (59) is not a surface reflection of an agreement between the logophoric center of the embedded CP (i.e., the “source,” see Sells 1987) and the lower C. (60a) indicates that a non-logophor subject can agree with the lower C head; on the other hand, (60b) suggests that the non-subject logophor, *Sammy*, cannot trigger agreement with the lower C. Taken together, it is clear that the subject of the higher clause rather than the logophor of the embedded CP agrees with the C head in the lower CP.

²²The gloss SA that follows a verbal noun class agreement indicates agreement with a ‘subject.’

- (60) a. Sammy ka-bol-el-wa a-li ba-keni b-ola.
 1Sammy 1SA-say-APL-PASS 1-that 2-guests 2SA-arrived
 ‘Sammy was told that the guests arrived.’
- b. Ba-sasi ba-bol-el-wa nende Sammy mbo/*a-li ba-keni
 2-parents 2SA-say-APL-PASS by 1Sammy that/*1-that 2-guests
 ba-a-rekukha.
 2SA-PST-leave
 ‘The parents were told by Sammy that the guests left.’

It should also be noted that S–C agreement is subject to locality constraints. A C head must agree with a subject that is right in the immediately superordinate clause, not with subjects in any higher clause, as illustrated by (61).

- (61) Alfredi ka-a-loma a-li ba-ba-andu ba-mwekesia bali/***ali**
 1Alfred 1SA-PST-say 1-that 2-2-people 2SA-revealed 2-that/***1-that**
 o-mu-keni k-ola.
 1-1-guest 1SA-arrived
 ‘Alfred said people revealed that the guest arrived.’

Diercks (2013) claims that such subject-orientation in agree in examples such as (59) can be explained if we assume a null operator in SpecCP which agrees with C *-li* through a Spec–head relation. This null operator is a subject-oriented anaphor, akin to the reflexive clitic *se* in French (Safir 2004). This reflexive raises to adjoin to a higher T where it is bound by the matrix subject (cf. Chomsky 1986), as shown below:

- (62) $[_{CP} C [_{TP} \mathbf{ba-ba-ndu}_{[v\phi]j} \text{ REFL}_j\text{-T} [_{vP} t_j v [_{VP} V [_{DP} \text{Alfredi}] [_{CP} \text{REFL}_j \mathbf{ba-}$
 $\mathbf{li}_{[u\phi]} \dots]]]]]$

This analysis correctly captures the lack of intervention effects with regard to the object *Alfredi*. However, a potentially serious drawback of Diercks’ analysis is that a null reflexive is positioned in every embedded CP, yet they make no semantic (or phonological) contribution to their local CP. As acknowledged by Diercks (2010), these reflexive operators are not logophors or elements that are related to any thematic

roles, and they receive no identifiable thematic interpretation themselves. To make their status even more dubious, the null reflexives are required to undergo covert movement to the T head in the superordinate clause, which is a crucial aspect of the analysis to account for the subject-orientation effects in S–C agreement, but this assumption is essentially unfalsifiable, as it is a null element going through a covert movement.

There are still other drawbacks of Diercks’s (2013) analysis. Diercks has to assume two distinct agreement relations, a subject–reflexive and a null reflexive–C dependency. The subject–reflexive binding dependency has been argued *not* to be a primitive operation, but one that is reducible to agree (e.g., Hicks 2009 and Reuland 2011; see various approaches to reduce reflexive binding to agree in Chapter III). An alternative analysis that can directly analyze S–C agreement as a case of agree is thus simpler (and preferable). In addition, the null reflexive–C agreement requires the reflexive to be involved in ϕ -feature agreement, an agreement that has been shown to be impossible or unattested (so called “anaphor (anti-)agreement effects,” cf. Preminger 2018; Rizzi 1990; Woolford 1999). More problems of this indirect agree analysis will be discussed later.

The above problems can be avoided if we assume a direct agree between C and the superordinate subject. With a direct agree analysis, we do not need to assume a null reflexive, a hypothesis that does not have empirical basis. The analysis is therefore significantly simplified. In fact, Diercks (2010) considered such a possibility. Nevertheless, Diercks (2010) ends up not adopting a direct agree analysis, mainly due to the lack of intervention effects when there is an indirect object intervening between the agreeing subject and C. (59), repeated below in (63), shows that the indirect object ‘Alfred’ cannot agree with the embedded C.

- (63) Ba-ba-ndu ba-bol-el-a Alfredi ba-li a-kha-khil-e.
 2-2-people 2SA-said-APL-FV 1Alfred 2-that 1SA-FUT-conquer
 ‘The people told Alfred that he will win.’

This is a problem if we assume upward agree (e.g., Baker 2008; Zeijlstra 2012), which Diercks (2010) does. An upward agree analysis will incorrectly predict that in (63), the embedded C must agree with the object, which is structurally closer than the subject to the embedded C.

However, if we adopt MS-agree instead of the upward agree approach, this major obstacle for the direct agree analysis is naturally removed. With the MS-agree analysis, we no longer predict an intervention effect caused by the indirect object, as minimal search starts its search from full phases, the *v*P set in (63); that is, the search is top-down in the sense of tree representations.

In fact, I argue below that MS-agree provides a straightforward analysis of S-C agreement in Lubukusu, without running into the problems encountered by the indirect agree analysis. Before diving into the proposal, it is helpful to remind the reader again that our definition of minimal search can partially unify agree and labeling because the values of SD and ST are independently assigned by relevant operations. In my analysis of cyclic agree, I drew on this unique feature and proposed that in the first cycle, the SD is the co-member or sister of the trigger. In the second cycle, however, the SD is the next built phase (when the lexical array for that phase is exhausted). We will see that the same analysis can be directly extended to S-C agreement in Lubukusu.

“Subject-orientation” in Lubukusu can in fact be derived by MS-agree defined previously in (29) with an independently assigned SD, as in the second cycle of cyclic agree. Instead of being mediated by a null reflexive, I assume that the embedded C can directly agree with the matrix subject if the next built phase is taken as the SD. To make this concrete, consider the structure under discussion:

$$(64) \quad [_{vP} \text{ba-ba-ndu}_{[v\phi]} \ v \ [_{VP} \ V \ [_{DP} \text{Alfredi}] \ [_{CP} \text{ba-li}_{[u\phi]} \ \dots]]]$$

As a reminder, I assume that minimal search observes Chomsky’s (2001b) Phase Impenetrability Condition; namely, a phase head complement is transferred and is

not accessible to minimal search after a higher phase head enters the derivation. Therefore, when the matrix *v* head merges in, the complement of the embedded C will be transferred, indicated by the dots in (64). The lexical array for the matrix *v*P phase includes the *v*P-internal subject. Now the crucial step is: A minimal search is conducted with ST = [*uφ*] and SD = the phase that is built so far, i.e., the matrix *v*P. This minimal search first finds the *v*P-internal subject *ba-ba-ndu* (more accurately its D head), which causes intervention to the object *Alfredi*. The subject-orientation phenomenon in Lubukusu is therefore explained.

Is the current analysis compatible with the cyclic agree analysis proposed for Georgian and Hindi-Urdu? The answer is probably yes if an additional assumption proposed by Carstens (2016) is adopted. According to Carstens (2016), there are three CP layers in the left periphery in Lubukusu, namely, ForceP, Int(errogative)P and Fin(ite)P, following classical work by Rizzi (1997; 2001). The FinP is argued to be a host for the non-agreeing C *mbo*, which allows A-movement out of the CP headed by *mbo* as exemplified in (65).

(65) *Lubukusu* (Carstens 2016:19)

- a. E-fula e-lolekhana (mbo/*e-li) e-kha-kw-e muchiri.
9-rain 9SA-appear that/*9-that 9SA-FUT-fall-FUT tomorrow
‘It seems like it will rain tomorrow.’
(Lit. ‘Rain seems [that] will fall tomorrow.’)
- b. O-mu-keni ka-suubil-wa mbo/(a-li) k-ola.
1-1-guest 1SA-believe-PASS that/1-that 1SA-arrive
‘The guest was believed to have arrived.’

Since *mbo* is transparent for A-movement, Carstens (2016) considers it not a phase head (see also Carstens & Diercks 2013). On the contrary, A-movement cannot cross an agreeing C, i.e., *-li* in (65a, b), which is assumed to be a phase at the Force position. Force then selects an Int, overt versions of which can be found in some languages (Rizzi 2001).

An important empirical motivation for Carstens (2016) to assume two C phase heads is the differences between the complementizer agreement in West Germanic and Lubukusu: Differently from Lubukusu, unvalued ϕ -features on a West Germanic complementizer are valued by the subject of its complement TP. This difference is derived if the complementizer in West Germanic is at a low C head, the Fin head, as the embedded subject is still available when Fin merges in.

If we follow Carsten’s (2016) assumption about the CP periphery in Lubukusu, then, as pointed out by Carsten, S–C agreement can be implemented as a case of the second cycle of cyclic agree. (66) presents the corresponding structure.

$$(66) \quad [_{vP} \textbf{ba-ba-ndu}_{[v\phi]} \ v \ [_{VP} \ V \ [_{DP} \text{Alfredi}] \ [_{ForceP} \textbf{ba-li}_{[u\phi]} \ [_{IntP} \text{Int} \ [_{FinP} \text{Fin} \ [_{TP} \text{DP}_{\text{subject}} \dots]]]]]]]]$$

When the $C_{Force} -li$, bearing $[u\phi]$, merges with the IntP, the complement of Int, i.e., FinP, is transferred. Then the Cycle One minimal search is conducted, taking this IntP as its SD and the $[u\phi]$ as its ST. This minimal search fails, as the lower subject has been transferred. Only when the first cycle minimal search fails to value the $[u\phi]$ will a second cycle minimal search be executed. The derivation continues until the lexical array of the next phase is exhausted, that is, when the vP -internal subject merges in. In the second cycle, minimal search takes the vP as the SD and the $[u\phi]$ as the ST, and the vP -internal subject (its D head) is returned for agreement.

However, a potential critique against Carstens (2016) regarding her split-CP hypothesis in (66) concerns the IntP. Why is an Int posited in a declarative sentence, given that no evidence is given for the existence of an Int head or SpecIntP in Lubukusu? In addition, why should both the FinP and IntP be phases?

If one does not adopt Carstens’s (2016) assumption regarding the CP system in Lubukusu, an alternative is that the values of SD and ST for minimal search are determined as a language parameter or on an individual agreement basis, and must be learned for specific languages and specific linguistic constructions. What MS-

agree does is to offer a unified search algorithm. The specific values for SD and ST for a particular agreement construction in a particular language will be learned independently. The same assumption can be extended to cyclic agree. If minimal search does not return anything in the first cycle agree, another SD will be assigned for the second cycle agree. The relative order between these two types of minimal searches must be learned for specific languages and constructions. This is in a sense similar to the parametric approach proposed by Baker (2008).

To summarize, the current MS-agree analysis of S–C agreement in Lubukusu does not assume (i) the existence of a null reflexive, nor (ii) covert movement of the null reflexive, nor (iii) agreement between the reflexive and the embedded C head. In short, it does not run into any of the problems that we identified above for Diercks’s (2013) indirect agree analysis. In addition, compared to Diercks’s (2013) analysis, the current MS-agree based analysis is simpler and more intuitive, because now S–C agreement is established directly by agree, whereas Diercks needs to assume two distinct agreement relations, subject-reflexive and reflexive-C agreement.²³

2.6 Comparison to the upward agree analysis

In this subsection, I will make a comparison to the upward agree analysis proposed by Zeijlstra (2012) and Bjorkman & Zeijlstra (to appear). Zeijlstra (2012) argues that agree is always upward; that is, a head bearing uninterpretable features always probes upward to find the corresponding valued features. Such an upward search approach can account for the various cases of concord we have reviewed in Section 2.5.1, e.g., negative concord in Czech, case assignment in Japanese (multiple agree) and inflection

²³MS-agree can capture another constraint on Lubukusu S–C agreement: the logophoric requirement. Specifically, the subject that agrees with the subordinate C must also be a logophoric center of the superordinate clause (not that of the subordinate clause). Due to the complications beyond the scope of this chapter, I leave the account of the logophoric requirement to future research. Please see Ke (in preparation) for details of an MS-agree analysis that captures the logophoric requirement in addition to subject-orientation.

doubling in Norwegian, Frisian and Swedish. Upward agree may be also involved in the second cycle of cyclic agree in Georgian.

Bjorkman & Zeijlstra (to appear) revise their upward agree analysis by separating agree into two steps: checking and valuation. Feature checking is defined below:

(67) *Feature checking* (Bjorkman & Zeijlstra to appear)

α checks an uninterpretable feature on β iff:

- a. α carries a matching interpretable feature;
- b. α c-commands β ;
- c. α is the closest goal to β

This is exactly the upward agree that was defined previously by Zeijlstra (2012), as we discussed in (14).

Feature valuation is simply a copy operation that copies the values of the interpretable features on α to the uninterpretable features on β . The crucial revision Bjorkman & Zeijlstra (to appear) make is that feature checking establishes a relation of accessibility between syntactic items in agree. The Accessibility Condition is defined as below:

(68) *Accessibility Condition*

α and β are accessible to each other iff an uninterpretable feature on β has been checked (via feature checking) by a corresponding interpretable feature on α .

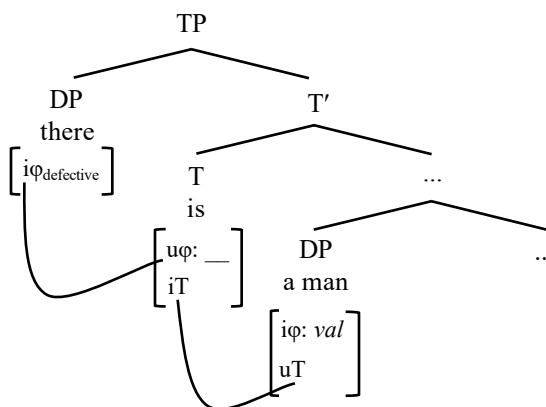
The Accessibility Condition in (68) suggests that as long as α and β are in a checking relation for any feature, they are accessible to each other. This implies that all the features on α are accessible to all the features on β . However, the Accessibility Condition needs to work together with another hypothesis:

(69) Only checked features can be valued.

(69) is in fact a restriction only on uninterpretable features.²⁴ Together with the Accessibility Condition, it might be the case that some uninterpretable features UniF on a syntactic object α are accessible to features on another object β if α and β have a certain feature in a checking relation, but as long as UniF has not been checked, it cannot be valued by any features on β .

To be concrete, let us use their analysis of *there*-constructions as an illustration:

(70) There is a man outside.



In (70), the [uT] feature on *a man* is checked upward against the [iT] feature on T. According to the Accessibility Condition in (68), *a man* and T are accessible to each other. Why then can't the ϕ -features on T be valued by the corresponding features on the DP *a man*? This is where the restriction in (69) comes into play: Only checked features can be valued. That is, in the case of (70), the [u ϕ] on T cannot be valued even though T is accessible to the DP *a man*, because the [u ϕ] on T have not been checked.

The crucial step is that the [u ϕ] feature on T also probes upward and finds *there*, which Bjorkman & Zeijlstra (to appear) assume to bear defective ϕ -features. As a result, the [u ϕ] feature on T, though checked by the defective ϕ -features on *there*,

²⁴Bjorkman & Zeijlstra (to appear) assume that (69) in particular does not apply to unvalued interpretable features, as shown by their analysis of the long-distance agreement in Tsez (pp. 40–41).

cannot be successfully valued by *there*. This leaves a chance for the $[u\phi]$ feature on T to be valued by the DP *a man*. Note that now this valuation is possible because the $[u\phi]$ features on T have been checked by those on *there*, and T is accessible to *a man*. In other words, feature checking and feature valuation can be dissociated: An uninterpretable feature being checked by a corresponding interpretable feature does not necessarily mean that it must be valued by that interpretable feature.

Such an upward agree analysis is interesting and worthy of further exploration. However, theoretically, this analysis has some surprising features. These features are not necessarily problematic, but probably call for an explanation.

In order to be valued, a feature needs to be checked. This is a natural assumption if feature checking is associated with feature valuation. That is, feature checking is intended to find matches between interpretable and uninterpretable features and to establish a relation between these two types of features. Then feature valuation makes use of this relation by copying the feature value from an interpretable feature to the corresponding uninterpretable one. However, feature checking under Bjorkman & Zeijlstra’s system is not intended to establish such a relation.

In addition, in such an upward search approach, the movement of an external argument from the *v*P-internal subject position to SpecTP is motivated by the need to value the uninterpretable ϕ -features on T. On the other hand, the subject DP has its case feature ($[uT]$) valued by upward agree against the T head when it is at Spec*v*P. In other words, the external argument receives its case when it is in its original θ -position. This is at odds with the standard assumption that SpecTP is a case position. The theoretical and empirical consequences of this new assumption have not been discussed and await further investigation.

Empirically, as we have pointed out, upward agree cannot be the mechanism behind the long-distance “upward” S–C agreement in Lubukusu. S–C agreement provides a crucial test for the upward agree approach, but the lack of intervention

effects in constructions with an object intervening between the superordinate subject and the embedded C head strongly discredits upward agree as a general analysis for S–C agreement in Lubukusu.

Bjorkman & Zeijlstra (to appear) argue that their revised upward agree analysis can, like the standard downward agree analysis, account for long-distance downward agree such as the long-distance agreement (T/*v* agreeing with an immediately subordinate topic) observed in Tsez, as reviewed in Section 2.2.2. Compared to the standard downward probing analysis, their analysis is more like an indirect agree analysis. For instance, they need to assume that *v* agrees with the subordinate C head,²⁵ and the C head in turn agrees with the subordinate topic. The ϕ -features are copied from the topic DP to the subordinate C head, and then the subordinate C head transmits these ϕ -features to the matrix *v* head. The subordinate C head bears interpretable ϕ -features as DPs, but meanwhile these ϕ -features on C are assumed to be unvalued interpretable features so that they can be valued by the valued interpretable features on the topic DP. Ignoring all the details and assuming that the upward agree provides an analysis for the long-distance agreement as the standard downward agree analysis, the standard downward agree analysis is still preferable, as it establishes a simpler, direct relation between the T/*v* head with the subordinate topic.

In the next chapter, I will extend MS-agree to reflexive binding, another domain in which an upward agree analysis has been proposed. I argue that reflexive binding is better analyzed as an instance of MS-agree rather than upward agree.

²⁵If Bjorkman & Zeijlstra (to appear) assume that it is the T head that agrees with the embedded C, then they should further establish a relation between T and *v* such that the ϕ -features on *v* can be transmitted to T.

CHAPTER III

A Minimal Search-Based Analysis of Reflexive Binding

3.1 Introduction

As briefly pointed out by Wurmbrand (2012b), reflexive binding can potentially be modeled by upward agree. That is, it may be hypothesized that a reflexive bears certain unvalued features, and these features search upward to find an antecedent that matches these features (see also Wurmbrand 2017). Wurmbrand (2012b) does not mention the details of this upward agree analysis. However, a version of such an upward agree analysis for reflexive binding is developed by Hicks (2009).

In this chapter, I will develop a minimal search-based agree (MS-agree) analysis of reflexive binding, building on Hicks’s (2009) agree-based analysis. I will discuss a common problem faced by both Hicks’s (2009) agree-based analysis and my MS-based agree analysis (e.g., the double object and dative constructions). A possible solution along the lines of the MS-agree analysis is offered.

The discussion will then reveal some empirical gains of the MS-agree analysis: It correctly predicts the non-complementary distribution between reflexives and pronouns in some cases and complementary distribution in some other cases, with an additional assumption that PPs are optionally phases. Finally, I compare the rela-

tions established by MS-agree to the c-command relation.

3.2 Previous analyses

The goal of this subsection is not to give an extensive review of the previous binding theories, nor even to exhaustively cover all the most influential ones. Instead, the goal is to review some previous theories of binding that are directly relevant to the topic of this chapter: the MS-agree analysis of reflexive binding. In addition to the classical binding theory, I will focus on analyses that intend to reduce binding to more primitive operations in syntax. Given the minimalist framework I assume throughout this dissertation, there are only three operations available in syntax: merge (including external merge and internal merge), agree and labeling. Consequently, binding, or co-reference, has no independent status under the current assumption. In addition, binding as a component of syntactic theory does not explain how exactly two syntactic objects can be involved in a long-distance relation. Recent efforts have focused on various reductionist approaches. Specifically, researchers have tried to reduce binding to movement (Hornstein 2001; Kayne 2002; Zwart 2002; Hornstein 2007; Drummond et al. 2011) or to agree (Reuland 2001; Hasegawa 2005; Heinat 2006; Reuland 2006; Hicks 2009; Reuland 2011; Wurmbrand 2017).

3.2.1 The classical binding theory

Let us begin with the classical binding theory as proposed by Chomsky (1981, 1986), given below in (1)–(5). I will not discuss Condition C but will restrict the domain of investigation primarily to Condition A, occasionally extending the discussion to Condition B.

(1) *Definition of c-command*

A c-commands B iff

- a. neither A nor B dominates the other, and
 - b. the first branching node that dominates A also dominates B.
- (2) *Binding Condition A*
- A reflexive must be bound within its governing category.
- (3) A binds B iff
- a. A c-commands B, and
 - b. A and B are co-indexed.
- (4) *Binding Condition B*
- A pronoun is free (= not bound) in its governing category.
- (5) *Definition of governing category*
- X is a governing category for Y iff X is the minimal category (minimal TP or DP) containing Y, a governor of Y and a SUBJECT accessible to Y.

For instance, (6a) satisfies Condition A because the reflexive is properly bound by *John*. *John* binds *himself* in this sentence, because it c-commands and co-indexes with *himself*. (6b, c) are ungrammatical because *John* does not bind *himself* as it does not c-command *himself*.

- (6)
- a. John_{*i*} likes himself_{*i*}.
 - b. *John_{*i*}'s mother likes himself_{*i*}.
 - c. *Himself_{*i*} likes John_{*i*}.

3.2.2 The movement approach

To derive Condition A through movement, Hornstein (2001, 2007) proposes that the antecedent/binder is base-generated with the reflexive it binds. Anaphors are residues of overt A-movement. That is, for sentences such as (6a), one of the early derivation steps is shown in (7a), after which *John* moves to the SpecvP position,

which is the position for the external argument and thus is a θ -position, as illustrated by (7b).

- (7) a. [_{VP} like John-self]
 b. [_{vP} John_j v like t_j-self]
 c. [_{TP} John_j [_{T'} likes_k [_{vP} t_j t_k him_j-self]]]

After the derivation is complete, the trace of the antecedent is pronounced as a pronoun (*him* in (7c)) at PF, an operation which is called “pronoun-support,” reminiscent of *do*-support in English. The ϕ -features of the antecedent are correctly reflected in the corresponding pronoun at spell-out. The reason *-self* needs to be supported by a pronoun is that it cannot stand alone at PF given that it is a bound morpheme (Hornstein 2001:160).

A few questions arise immediately. How are case and θ -roles assigned? Taking (6a) again as an example for the sake of concreteness: Regarding case assignment, *-self* receives accusative case, and *John* is assigned nominative case when it occupies SpecTP. Regarding θ -role assignment, the puzzle is why *John*, after acquiring the θ -role of a theme, moves into a θ -position, SpecvP, to receive a second θ -role. A DP bearing two θ -roles violates the theta filter, and movement into θ -positions violates the θ -theoretic principle (Chomsky 2000a:102).

Kayne (2002) and Zwart (2002) intend to provide a solution to these problems. An important contribution of Kayne (2002) is the elimination of the configuration in which a single DP receives two θ -roles. This is done by assuming that the reflexive and its antecedent merge together directly. For (6a), the object will be a complex DP, [John-he]’s self]. Then *John* moves as under Hornstein’s (2001) analysis. It is now the complex DP that receives a θ -role, and then *John* moves upward to SpecvP to get its own θ -role.

Zwart’s (2002) solution is quite similar. Zwart assumes that the complex DP is the antecedent + a pronominal, e.g., [John + he]. The merge of these two lead to a

co-reference relation between them in syntax. The pronominal is then pronounced as a reflexive. In other cases such as (8), the relation between *John* and *him* must be an accidental co-reference relation, not a co-reference relation given by syntax through merge. However, it is unclear why merge should give rise to a co-reference relation.

- (8) John_{*i*}'s mother loves him_{*i*}.

Hornstein (2001, 2007) (as well as Kayne (2002); Zwart (2002)) also tries to derive Condition B through the movement analysis. The leading idea is that pronouns are more costly than reflexives; whenever reflexives are possible (e.g., in the local domain), pronouns must not be used. The reason why inserting a reflexive is preferred to inserting a pronoun in binding cases such as (7) is that the former is a result of move, whereas the latter is not. Hornstein (2001, 2007) assumes that move is more economic than merge. Therefore, inserting a reflexive is more economic than inserting a pronoun.

It is worth noting that this analysis does have many advantages. First, no co-indexation or an additional concept of binding has to be stipulated, consistent with the minimalist assumptions. Second, Binding Condition A is reduced to the independently motivated move and copy theory of movement; the long-distance relation between reflexives and their antecedents receives an explanation. Third, the c-command requirement for binding is derived by upward movement. Fourth, the locality requirement in reflexive binding is derived through the locality constraints on movement.

However, the proposal cannot work due to both theoretical and empirical difficulties. As previously mentioned, the movement approach assumes movement into θ -positions, which is at odds with the θ -theoretic principle.

In addition, no compelling motivation is given for the assumption that inserting a reflexive is more economic than inserting a pronoun. As pointed out by Hicks (2009), if the pronunciation of trace-*self* as a reflexive requires pronoun insertion, why that is less costly than inserting a pronoun directly? In fact, Hornstein has to resort to

movement in addition to inserting a pronoun to deal with the trace-*self* constituent, which makes the derivation of (9b) more economical than (9a); this predicts that (9b) is grammatical whereas (9a) is not, contrary to fact.

- (9) a. John_i likes himself_i.
 b. *John_i likes him_i.

Empirically, given that the analysis assumes movement of the antecedent, it is important to test this movement with syntactic islands. Available evidence strongly indicates that island effects are not attested with reflexive binding. Whereas the movement approach assumes that *the children* should be moved from the position of *each other* in (10a), such a movement from the possessor position is expected to be blocked by the Left Branch Condition proposed by Ross (1967). For instance, this violation of the Left Branch Condition leads to ungrammaticality in the case of wh-movement in (10b).

- (10) (Antonenko 2012:36)
 a. The children_i like each other_i's toys.
 b. *Whose do the children like [t_i toys]?

Similarly, the suggested movement of the antecedent does not give rise to island effects for other syntactic islands, as exemplified in (11b), (12b) and (13b), whereas such a movement from a *picture*-NP, a complex DP or a conjunct phrase is impossible for wh-phrases, as illustrated by (11a), (12a) and (13a), respectively:

- (11) a. *Who_i does [the picture of t_i] terrify John?
 b. Does [the picture of himself] terrify John?
 (12) a. ?*Who_i did they write [an article about t_i]
 b. They_i wrote [an article about themselves_i]
 (13) a. *Who_i does Mike trust [Sue and t_i]?

- b. Mike_i trusts [Mary and himself_i].

It may be argued that the reflexives inside a *picture*-NP, a PP or a conjunct phrase are not plain reflexives but exempt anaphors or logophors, which behave like a pronoun (Reinhart & Reuland 1993). I will turn back to this issue after the discussion of Reuland’s analysis of reflexive binding (e.g., Reuland 2011). The overall counter-argument is that the reflexive inside *picture*-NPs, PPs and conjunct phrases are not necessarily logophors and cannot be always exempt from Binding Condition A; therefore, the lack of island effects in (11)–(13) does pose a serious challenge to the movement approach.

3.2.3 The agree approach

Reuland (2001, 2006, 2011) proposes an agree analysis of reflexive binding with the SE type reflexives (= simplex reflexive) such as *zich* in Dutch. I will not deal with SE anaphors but will still include Reuland’s analysis here for the sake of completeness, as his agree analysis is quite influential. The leading idea is that an agreement chain is formed through a series of agreement relations. An example from Dutch is shown in (14a), with its syntactic representation in (14b).

- (14) a. *Dutch* (Reuland 2011:157, simplified, and with slight modifications to the gloss)

[Oscar_i [T [t_i [voelde [zich T [wegglijden]]]]]]
 Oscar felt.FIN SE slide-away.INF
 ‘Oscar felt him slide away.’

- b. [Oscar_i [Tns [t_i [voelde [zich T [wegglijden]]]]]]

A chain formed by linking five independent agreement relations between the following heads:

SE anaphors are assumed to be underspecified in number and gender features. As demonstrated in Figure 3.1, the ϕ -features are copied from the antecedent to the SE

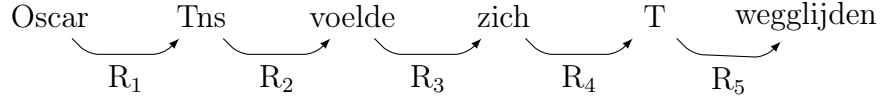


Figure 3.1: Reuland's (2011) agree analysis of SE anaphors

anaphor, passing through the chain.

English reflexives such as *himself* are considered as a SELF type anaphor. The binding of SELF can be illustrated as below:

- (15) a. Mike likes himself
 b. $[_{TP} \text{ Mike T } [_{SELF_i} \text{ adore}] [_{DP} \text{ him-} \langle \text{SELF}_i \rangle]]$
 c. $\lambda x. \text{ SELF-adore}(x, x)(\text{Mike})^1$

The syntactic representation of (15a) in (15b) assumes that the reflexive *himself* is composed with the pronoun *him* and *self*. *Self* covertly raises to adjoin to the closest predicate (the verb *adore* in (15a)). This is to reflexive-mark the verb, so that the predicate must apply to arguments with an identical reference, which imposes a requirement on the interpretation of the pronoun *him* (as the remnant of the movement of *self* out from *himself*): It must be (near-)identical to the other argument of the predicate in reference, as indicated by the semantic presentation of the sentence in

¹Reuland (2011) suggests that an additional function should apply to the object reflexive; this function is designed to capture various proxy readings of the reflexive. Since this is irrelevant to the current discussion, I leave out the function in the semantic representation.

(15c). This is built upon Reinhart & Reuland’s (1993) predicate-based analysis.²

The most interesting prediction of the predicate-based analysis is the non-complementary distribution of reflexives and pronouns in non-argument positions (where the arguments are enclosed in square brackets):

(16) (Reinhart & Reuland 1993:661)

- a. Lucie saw [a picture of herself/her].
- b. Max likes [jokes about himself/him].
- c. Max saw [a gun near himself/him].

(17) Max said that the queen invited [both Lucie and himself/him] for tea.

(the judgment is Reinhart & Reuland’s)

According to Reinhart & Reuland (1993), the fact that the reflexives and pronouns optionally occur in (16) indicates that in *picture*-NPs, adjunct PPs and conjuncts, the reflexives are not plain reflexives, but logophors. (18a) suggests that the reflexive in the *picture*-NP and the conjunct DP is very likely a logophor, as long-distance binding is allowed (only for a subset of speakers), especially when the reflexive is stressed.

(18) a. John_i thought Mary had a picture of ??/?*himself_i/him_i.

b. Mike_i thinks that Mary trusts [DP Sue and ??/?*himself_i/him_i].

²The matching between syntax and semantics of reflexive binding is not as straightforward as one may expect. ECM cases are problematic to such a simple mapping:

- (i) (Reuland 2011:253)
John heard himself criticize Lucie.

Reuland (2011) suggests that SELF moves to reflexive-mark the matrix predicate *heard*. However, after being reflexive-marked, the predicate should be interpreted as follows:

- (ii) $\lambda x.$ SELF-hear(x , criticize(x , Lucie))(John)

We see that a reflexive-marked predicate does not impose an interpretation requirement on its arguments, different from what has been claimed in general. Therefore, the predicate-based analysis is forced to innovate the term “syntactic argument” to include reflexives which are case-marked by the predicate but not really one of its arguments (see Reinhart & Reuland 1993:678).

Nevertheless, for both (17), (18a), and a different version of (17) with a reflexive in a conjunct phrase as in (18b), my consultants definitely prefer to use pronouns instead of reflexives, implying that the anaphoric reading, rather than the logophoric reading, is a prominent interpretation; that is, if possible, the reflexive is interpreted as a plain anaphor instead of a logophor.

Furthermore, long-distance binding for reflexives inside adjunct PPs seems unlikely as (19) indicates:

- (19) a. Mary_i said Max likes jokes about *herself_i/her_i.
 b. Max_i said Mary saw a gun near *himself_i/him_i.

Hestvik & Philip (2001) and Bruening (2014) report similar judgment regarding the inaccessibility of long-distance binding in (19). This is not predicted if the reflexives are logophors, which are assumed to be pronominal and normally allow long-distance binding (Reinhart & Reuland 1993; Charnavel 2018a).

The logophoric status of the reflexives inside *picture*-NP constructions, adjunct PPs and conjunct phrases can be further tested with inanimate reflexives. As pointed out by Charnavel & Sportiche (2016), inanimate reflexive requires an inanimate antecedent, which does not have a mind or perspective to report. Therefore an inanimate reflexive is incompatible with the logophoric reading, which requires a logophoric center. A logophoric center is an attitude holder or empathy locus, according to Charnavel (2018a), or is the source, self, and point of view, according to Sells (1987). The following sentences show that the reflexives in *picture*-NPs, PPs, or conjunct phrases are not necessarily exempt anaphors or logophors:

- (20) a. The cat_i stood on [a picture of itself_i].
 b. The mouse_i dragged [a piece of cheese with itself_i].
 c. The student's report indicated that the table_i broke [the chair and itself_i].

If reflexives in these constructions must be logophors, then we predict that inanimate reflexives cannot occur in these environments, contrary to fact. (20) thus strongly support Charnavel’s (2018) argument that reflexives in these contexts are NOT NECESSARILY logophors. Instead, the reflexives can be locally bound reflexives, although logophoric interpretation is sometimes (weakly) accessible.

Ke & Pires (accepted) find that body-part nouns in Mandarin Chinese bear an implicit reflexive argument and this argument must be locally bound. The implicit reflexive argument of body-part nouns does not have a logophoric reading. The fact that body-part nouns can be bound inside *picture*-NPs, adjunct PPs, and conjunct phrases, as illustrated by the examples below, confirms that these constructions do not force a logophoric reading of the (implicit) reflexives.

- (21) a. Zhangsan_i zhao-le [yi-zhang REFL_i lianbu de texie].
 Zhangsan take-ASP one-CL face DE closeup
 ‘Zhangsan_i took a closeup photo of his_i face.’
- b. Zhangsan_i [zai REFL_i shou-shang] xiezi.
 Zhangsan at hand-LOC write-character
 ‘Zhangsan_i wrote characters on his_i hand.’
- c. Zhangsan_i pao-huai-le [REFL_i jiao he wushu-shuang xiezi].
 Zhangsan run-bad-ASP foot and countless-CL shoe
 ‘Zhangsan_i injured his_i feet and wore out countless shoes.’

Literally: Zhangsan’s feet and countless shoes were bad because of running (too much).

In all the above examples, the implicit reflexive arguments of the body-part nouns can be replaced with the locally bound reflexive *ta-ziji* ‘himself,’ again suggesting that these reflexives are not logophors.

The empirical observation that poses a challenge to Reinhart & Reuland (1993) and Reuland (2011) (as well as his earlier work) is summarized as below:

- (22) Reflexives in non-argument positions are not necessarily (although could be) logophoric. Instead, a plain anaphoric interpretation of the reflexive is preferred.

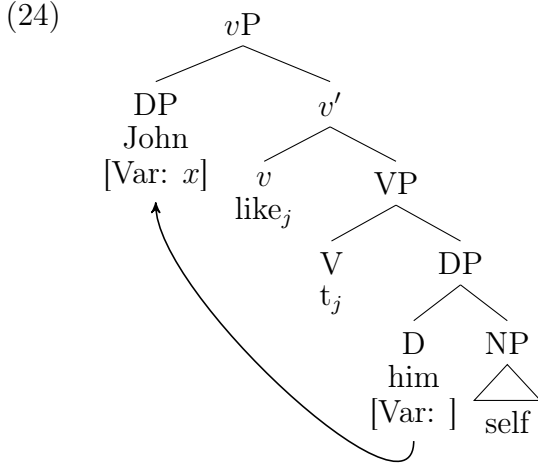
Hicks (2009) does not follow the predicate-based analysis but turns back to the classical analysis of Binding Condition A (see a similar approach in Charnavel 2018a), which does not treat reflexives in non-argument positions differently from those in argument positions. Therefore, Hicks (2009) does not run into the same problems as Reuland’s (2011) analysis does. Such an analysis is consistent with the empirical observation in (22).

Hicks’s analysis reduces Condition A to agree. Hicks’s (2009) agree-based analysis is actually familiar to us, because he formalizes binding as cyclic agree. As a standard assumption, Hicks considers anaphors a combination of a pronoun D head and a NP complement *self*. For instance, the structure of *himself* is shown as below:

$$(23) \text{ himself} = D_{him} + NP_{self}$$

Hicks (2009) further assumes that the D head in a reflexive enters the derivation with an unvalued [Var:] feature, which is responsible for its reference. In order to acquire a reference for the reflexive, its [Var:] feature must be valued by agree. Building upon the cyclic agree analysis, which we reviewed in Section 2.5.2, Chapter II, Hicks proposes that D_{him} first probes into its NP sister *self*. This is the first cycle agree. Given that *self* does not have a specified reference, the [Var:] value on D cannot be valued. Then in the second cycle agree, the D head with [Var:] probes upward. As we know regarding upward agree, the [Var:] feature can only be valued by a matching valued feature of a c-commanding antecedent hierarchically closest to D. According to Hicks (2009), (LF) phases (CP, *v*P and *n*P with a specifier) are local binding domains for Condition A, which require an antecedent to be found in the minimal binding domain where the reflexive is located.

The example in (24) demonstrates how Hicks’s (2009) analysis works.



The D head of the reflexive DP bears a $[\text{Var: }]$ feature, and this feature must be valued by upward agree, finding the DP in *SpecvP* with a $[\text{Var: } x]$ feature. The feature on the reflexive is valued against that on the DP. It is worth noting that the upward agree analysis derives the c-command requirement on the antecedent of reflexives, as the algorithm always searches upward into the sisters of the nodes dominating the reflexive.

To account for the constructions that allow either a reflexive or a pronoun, as the *picture*-NP, adjunct PPs and conjunct phrases mentioned in (16) and (17), Hicks (2009) assumes that the binding domains for Condition B (for pronouns) are different from those for Condition A (for reflexives). The binding domains for Condition B are called PF phases, including PPs, DPs/*n*Ps (either with or without a subject), whereas the binding domains for Condition A are LF phases, including CP, *v*P and *n*P with a subject.³ If the *picture*-NP with a determiner, conjunct phrases, as well as adjunct PPs, as types of DPs, are binding domains for pronouns, pronouns are predicted to be allowed in these phrases as no antecedent is available to bind them in these binding domains, consistent with the data in (16) and (17).

³A similar suggestion regarding the adoption of different binding domains for Condition A and B can be found in Charnavel (2018a).

Hicks (2009) has a broad empirical coverage. But instead of introducing all aspects of his analysis, I would like to identify below a serious problem of this analysis. The problem is caused by optional binding from the subject or the first object in double object or dative constructions:

- (25) a. John_i gave Jim_j a picture of himself_{i/j}.
- b. John_i gave to Jim_j a picture of himself_{i/j}.
- (26) a. The teachers_i sent the children_j each other_{i/j}'s crafts.
- b. The teachers_i sent the children_j to each other_{i/j}'s house.
- (27) (Revised from Barss & Lasnik 1986:347)
- a. Bill_i showed John_j himself_{i/j} (in the mirror).
- b. Bill_i showed John_j (in the mirror) to himself_{i/j}.

These examples are problematic for Hicks's (2009) agree-based analysis because as agree searches upward, it will return the first syntactic object it finds. If we assume that the subject c-commands the first object (let us refer to the PPs in these constructions as objects of the predicates for the sake of convenience),⁴ and the first object c-commands the second object which contains the reflexive, then the first object will always be returned by upward search. In other words, there is no way to also return the subject as the antecedent of the reflexive. On Hicks's approach, I do not see an easy solution to this problem.

We have seen that reflexive binding can be implemented as an instance of agree. There are other notable agree-based analyses, such as Heinat (2006), similar in spirit to Reuland's (2011) agree-based analysis to SE anaphors, Hasegawa (2005), who assumes that T agrees with the antecedent and the reflexive through multiple agree, and Antonenko (2012), who hypothesizes that either the most local V, *v*, or T head

⁴No matter the first object is the direct or indirect object, it has been argued to be in a higher position than the second one (Barss & Lasnik 1986; Larson 1988; Pesetsky 1995).

above a reflexive can bear a specific unvalued feature that can be valued by the reflexive by regular downward probing. Given that Hicks’s (2009) analysis is highly plausible and theoretically reasonable, and given that my analysis will be built on Hicks (2009), I will not dig into the details of these alternative analyses. In the next section, I will provide an MS-agree analysis of reflexive binding and address the mentioned problem caused by double object or dative constructions.

3.3 Current analysis

I argue that reflexive binding is another instance of MS-agree. To be more specific, reflexive binding is a case of minimal search that takes the next built phase as the search domain (SD) and the index feature [Ind:] (similar to the [Var:] feature in Hicks’s (2009) system) as the search target (ST). The purpose of this minimal search is therefore to value the index feature on reflexives.

Why is the ST not the ϕ -features? There are two reasons. First, I assume that the ϕ -features on reflexives are brought from the lexicon, as the pronoun part of a (complex) reflexive specifies the ϕ -features of the reflexive. Second, I also assume that these ϕ -features are semantically interpretable (as I will argue extensively in the next chapter). Following Chomsky (2000a, 2001b), I assume only unvalued, semantically uninterpretable features can be the ST; therefore, the semantically interpretable, valued *phi*-features on reflexives cannot be.

I assume all D heads (e.g., regular determiners and pronouns) bear an unvalued [Ind:] feature from the lexicon.⁵ An exception must be set for reflexives: Reflexives are not assigned such a variable as the *self* component of reflexives signals that the [Ind:] feature must remain empty.

⁵In fact, in order to solve the repetition–copy distinction problem, I have to assume all lexical items bear an [Ind:] feature and a variable is assigned to this feature when a given lexical item is brought to the lexical array or workspace. I must also assume that lexical items for a phase must be assigned different variables that have different consequences on their semantic interpretation.

Index features can be independently motivated (cf. Collins & Stabler 2016). They can be used to distinguish different syntactic objects with the same formal features (called repetitions) from copies of the same syntactic object. For instance, in the following structure representation (28), assuming the copy theory of movement, there are two *likes* as independent syntactic objects, namely *like_y* and *like_k*, but both *likes* have two copies.

$$(28) \quad [_{TP} \text{ John}_x \text{ T } [_{vP} \text{ John}_x \text{ like}_y [_{VP} \langle \text{like}_y \rangle [_{CP} \text{ what}_z [_{TP} \text{ you}_h \text{ T } [_{vP} \text{ you}_h \text{ like}_k [_{VP} \langle \text{like}_k \rangle \text{ what}_z]]]]]].$$

The two *likes* and their copies cannot be confused, otherwise the SM interface cannot determine which copies must be deleted in the course of linearization (Nunes 1995; Bošković & Nunes 2007), and the CI interface cannot know if two DPs with identical formal features must be assigned the same reference.

If no index feature is assigned to the heads, we would need a much more complicated algorithm to achieve the same goal, e.g., by “remembering” the derivational history in which copies are produced (Chomsky et al. 2017). Alternatively, the syntax has to identify copies and distinguish them from repetitions with a rather complicated algorithm (to the best of my knowledge, there is not yet a simple solution to this problem) (Collins & Groat 2018). Since the distinction between repetitions and copies has an influence on their interpretation at both the CI and SM interfaces, the best place to “mark” the distinction is in the syntax (Andrew McInnerney, p.c.). This is especially the case if we resort to θ -theory, case theory, and/or derivational history to make such a distinction, because the only place that all the information is available is in syntax. If the repetition–copy distinction is established in syntax, whatever algorithm this may require, in order to not forget whether two syntactic objects are repetitions or copies when the syntactic objects are sent to interfaces for interpretation, some information must be added to individual syntactic objects. This information plays a role at best identical to that of index features. In this case, the Inclusiveness Condition

(Chomsky 1995b:225) is obviously violated.⁶

This is the principal reason that Chomsky does not want to use indices in syntax (see e.g., Chomsky 2013): assigning indices in syntax is a violation of the Inclusiveness Condition. The Inclusiveness Condition requires that the output of a system does not contain anything beyond its input. This condition is primarily for the computational system of human language, namely, the syntax system. To circumvent this problem, I assume that lexical items bear an unvalued [Ind:] feature in the lexicon, and this feature is normally assigned a variable when the relevant heads are selected into either the lexical array/numeration (Chomsky 1995b) or the workspace (Chomsky 2017; Collins & Stabler 2016), before entering the derivation. Since the values for the index feature is assigned in the process of selection rather than in the course of derivation in syntax, the Inclusiveness Condition is essentially not violated. This removes the main reason not to use indices (Chomsky 1995b; Reuland 2011).

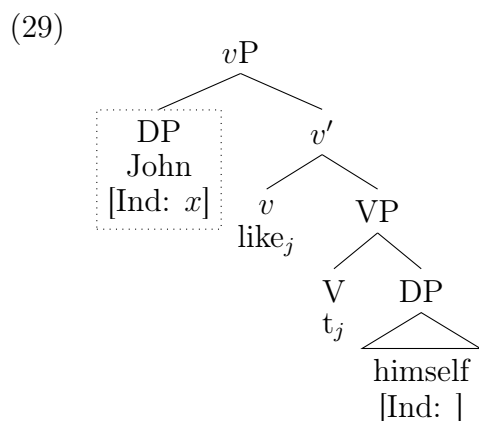
As a result, with index features that are brought from the lexicon, and with variables assigned to the index features in selection, the distinction between repetitions and copies is straightforward. Now we do not need to resource to a memory in syntax or a storage of chains of copies or any other algorithms specifically designed to derive the copy–repetition distinction. Instead, copies are simply lexical items that are of the same index feature, and repetitions are lexical items that have identical formal features but different index feature values.

Given an unvalued index feature on a reflexive, an agreement relation should be established between the reflexive and its antecedent. This can be accomplished by MS-agree. If minimal search with the unvalued [Ind:] feature (attribute) as the ST can find a D head that can value the [Ind:] feature on the reflexive, the index feature value (as a variable) on D is copied to the reflexive as in a regular valuation process. Note that it is a D head, rather than a DP set, that values the [Ind:] feature

⁶See (10) in Chapter I for an introduction of the Inclusiveness Condition.

on the reflexive, because indices are assigned only to lexical items (not sets) when they are selected. This can be perfectly fit into the MS-agree system, as it is exactly the index feature on a D head that values the reflexive's [Ind:] feature. The next chapter will explain how this agreement between index features can feed the semantic interpretation.

Let us illustrate the idea with a simple example. Recall that minimal search will always search downward, and in the following example, minimal search will take the built phase vP as its SD and the [Ind:] feature (or a combination of the [Cat: D] and the [Ind:]) as its ST.



The D head of *John* will be found (in the dotted rectangle), giving rise to the binding relation between *John* and the reflexive.

Can we filter out the ungrammatical example in (30)⁷?

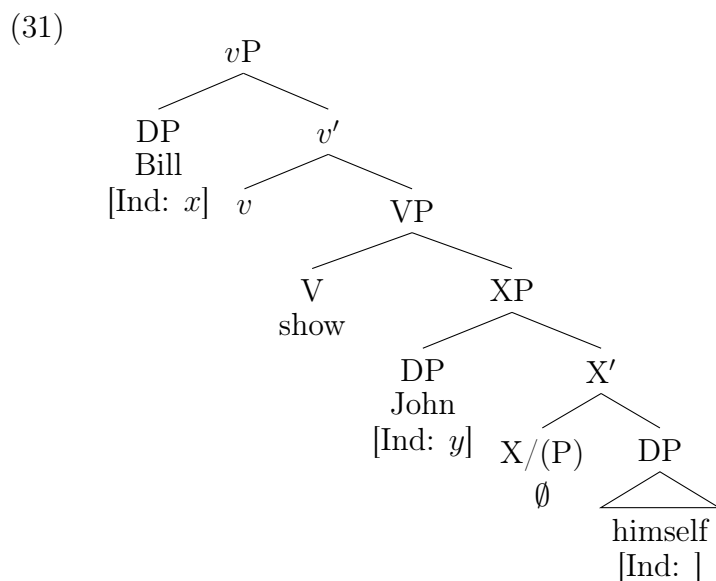
(30) *_[CP] Himself_[Ind:] likes Bill_[Ind: x].

The answer is positive. The unvalued index feature on *himself* will trigger a minimal search with the matrix CP as the SD. Given that the ST is the feature attribute [Ind], *himself* will be returned. However, the unvalued feature on *himself* cannot value the unvalued feature on itself. Therefore, (30) is out because the unvalued index feature on *himself* is left unvalued.

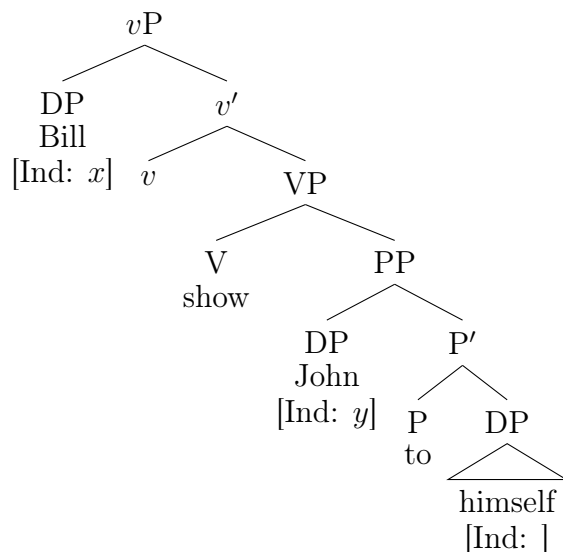
⁷I should thank Richard Lewis for bringing sentences such as (30) to my attention.

Now how can we solve the problem caused by double object or dative constructions as shown in (25–27)? Recall that the problem is that either the subject or the first object can be the antecedent of the reflexive in the second object position. This may also cause a problem to the MS-agree analysis proposed above, because if only vP is a phase, then minimal search with the vP as the SD will return the subject only; the first object will never be returned. So the MS-agree analysis predicts the opposite to the upward agree analysis: The upward agree analysis (e.g., Hicks 2009) predicts only the first object can be the antecedent, whereas MS-agree predicts only the subject can be the antecedent. The data suggest that both predictions are incorrect.

The answer should be based on the syntactic analysis of the double object and dative constructions. Following Kayne’s (1984) small clause analysis, incorporating insights from Larson (1988) and Pesetsky (1995) about the structures of VP/ vP and dative constructions, I assume the following structures for the double object (31) and dative constructions (32). Only the vP structures are relevant to our discussion here.



(32)



To be more specific, the structures are essentially adopted from the structures for double object and dative constructions that Pesetsky (1995) develops based on Kayne's (1984) analysis. For the double object constructions, demonstrated in (31), the XP resembles the small clause in Kayne's (1984) analysis. The null X is a predicate according to Kayne, carrying the meaning of HAVE. This works for most verbs that can take double objects, including act of giving (e.g., *give*, *sell*, *pay*, *trade*, *lend*, *loan*), instantaneous causation of ballistic motion (e.g., *throw*, *toss*, *kick*, *shoot*), verbs of sending (e.g., *send*, *mail*, *ship*), verbs of creation (e.g., *bake*, *make*, *build*, *cook*, *sew*), etc. (Pesetsky 1995:141–142). The example in (33a) denotes a small clause analysis of the double object construction. The meaning of (33a) might be represented as in (33b) (cf. Beck & Johnson 2004).

- (33) a. Max gave [_{HAVEP} Bill HAVE a gift].
 b. [Max's giving a gift] CAUSE [BECOME [Bill HAVE a gift]]

However, it is less obvious that some other verbs such as verbs conveying communicated message, e.g., *tell*, *show*, *ask*, *teach*, *write*, *read*, have anything to do with the possession meaning. Based on this and several other reasons, Pesetsky (1995) identifies X in (31) as a P head, which is a null case assigner, instead of a predicate.⁸

I adopt Pesetsky's view of the null element X as a P head here.

Let us now return to the binding issue with regard to the double object and dative constructions. The problem can be solved if we adopt the following hypothesis:

- (34) PP is optionally a phase.

Whether PP is a phase is currently under debate. Lee-Schoenfeld (2008) argues that in German, PPs with a P head that assigns its own θ -role(s) to its argument(s) (which we normally call adjunct PPs) can be considered phases, whereas PPs with a P head that does not assign its own θ -role(s) are not phases. In addition, Hicks (2009) hypothesizes that PP is not an LF phase but is a PF phase, and that an LF phase is relevant to Condition A, whereas a PF phase is relevant to Condition B. Abels (2003), based on the fact that P-stranding is not possible in Serbo-Croatian but possible in English, argues that PP in Serbo-Croatian is a phase whereas in English it is not. This is based on (i) the assumption that movement must proceed through the phase edge, i.e., the specifier of the phase, and (ii) the anti-locality constraint, which states that movement from the complement of a phrase to the phrase's specifier is blocked as the movement is too local. If PP is a phase, P-stranding would be blocked since the DP complement of P needs to move through the specifier of PP before moving any higher, and this step of movement violates the anti-locality constraint. By contrast, Bošković (2014) suggests that PP is a phase even in English, and P-stranding in English is allowed because it has a richer internal structure between P and DP: DP then is not the complement of PP; therefore movement through SpecPP does not violate the anti-locality constraint.

The debate on the phase status of PP becomes even more difficult to resolve if we adopt a cartographic approach to PP structure. With the assumption that PPs have a more complex internal structure across languages, P-stranding cannot

⁸Then we have to assume that the meaning of HAVE is brought in by other factors such as the verb that selects X.

be a general diagnostic of the phase status of PPs. Based on various cross-linguistic studies reported in Cinque & Rizzi (2010), Cinque (2010) concludes that stative and directional PPs have the following structures, represented respectively in (35a) and (35b), where PPdir = directional PP, PPstat = stative PP, AXPartP = axial part phrase (see e.g., Svenonius 2006):

- (35) a. *under the table*: $[_{PPstat} (AT) [_{DPplace} [_{AXPartP} \text{under} [X^0 [_{PP} P [_{NPplace} \text{the table} [PLACE]]]]]]]$
- b. *from under the table*: $[_{PPdir} \text{from} [_{PPstat} (AT) [_{DPplace} [_{AXPartP} \text{under} X^0 [_{PP} P [_{NPplace} \text{the table} [PLACE]]]]]]]$

If we follow Bošković (2014) and assume that when multiple PPs are present, only the highest PP can be a phase, then movement of NPplace to the Spec of PPStat or PPdir does not violate the anti-locality constraint. That is, the ungrammaticality of P-stranding in Serbo-Croatian cannot be explained with the anti-locality constraint once we assume (35) for PPs in Serbo-Croatian.

Citko (2014) provides various tests (including P-stranding) for the diagnosis of PPs as phases. She finds that the diagnostics give mixed results. Like standard phase heads such as C and *v*, P is the source of uninterpretable features (assigning case to its complement); PP is a domain for feature valuation (agreement between P and its complement); movement out of PP has to proceed through the edge of PP (P-stranding); and PP can be a binding domain. However, Citko (2014) also finds that quantifier raising cannot target SpecPP and the complement of P cannot be deleted in languages that lack argument ellipsis. These results indicate that PPs are different from standard phases in the sense that PPs do not share all standard properties of phases.

Following Rizzi's (1982: 107) parametric approach regarding PPs as a binding domain, I hypothesize that the phase status of PPs is subject to cross-linguistic variation as a parameter. There are logically three possible variations: PP is optionally

a phase, PP is not a phase, PP is a phase. These possibilities are worth a detailed discussion, which is beyond the purpose of this thesis. To address the problem in this chapter, I assume that in English PP is optionally a phase. On the other hand, in Polish, PP is not a phase, because PP in Polish cannot be a binding domain (Citko 2014).

With the hypothesis in (34), the double object and dative examples we mentioned, repeated below, can now be analyzed as a consequence of optional minimal search due to the optional phase status of the PPs. As previously shown, both the double object and dative constructions involve a PP below *vP*. If a given PP is optionally a phase, and minimal search is thus optionally triggered for that PP, then the fact that the reflexive inside the second object can be bound by either the subject or the first object is correctly derived.

- (36) a. John_i gave Jim_j a picture of himself_{i/j}.
- b. John_i gave to Jim_j a picture of himself_{i/j}.
- (37) a. The teachers_i sent the children_j each other_{i/j}'s crafts.
- b. The teachers_i sent the children_j to each other_{i/j}'s house.
- (38) (Revised from Barss & Lasnik 1986:347)
- a. Bill_i showed John_j himself_{i/j} (in the mirror).
- b. Bill_i showed John_j (in the mirror) to himself_{i/j}.

Taking (38a) as an example: If minimal search is conducted when the derivation reaches the PP [_{PP} John [_P himself]], minimal search with that PP as the SD and [Ind:] as the ST will return *John*. However, if minimal search is not conducted at the PP, then *John* will not be returned. Instead, a minimal search will be triggered at the *vP* phase, and thus the *vP*-internal subject *Bill* will be returned instead.

It must be noted that the problem raised by double object and dative constructions cannot be easily solved by the upward agree analysis proposed by Hicks's (2009). Even

with the hypothesis that PP is optionally a phase, or that PP is a PF phase, not an LF phase, upward agree would always find the first object as the antecedent. One way to avoid the problem is to assume that the reflexive or the *self* part moves to the phase edge of PP (Reuland 2011; Lee-Schoenfeld 2008), which Hicks (2009) does not assume. Such a move-based analysis is usually proposed for simplex reflexives (or SE anaphors) (e.g., Pica 1987; Cole et al. 1990; Huang & Liu 2001; Safir 2004). There is no strong motivation for complex reflexives (or parts of them) to undergo movement. Future research is necessary to evaluate which analysis is simpler and more natural theoretically, as well as more plausible empirically. The current analysis is at least theoretically desirable as it resorts to the theoretically well-motivated MS-agree.

3.4 Implications

By formalizing reflexive binding as an instance of MS-agree, we have reduced the locality requirement in reflexive binding to phase theory. Such a phase-based account of the locality constraint on reflexive binding was discussed and supported in many previous studies, including Lee-Schoenfeld (2004, 2008); Quicoli (2008); Despić (2015) and Charnavel & Sportiche (2016), among others. We have also reduced reflexive binding to a general, minimal search-based agree analysis. No additional mechanisms or operations are needed besides merge, agree and labeling. In addition to these theoretical advantages, I discuss some of the empirical consequences below.

3.4.1 The non-complementary distribution

A bonus of the hypothesis regarding the optional phase status of PPs (34) is that the surprising non-complementary distribution of reflexives and pronouns in PPs can be readily derived. A sentence we mentioned above in Section 3.2.3 is repeated in (39a), and more examples are drawn from Hicks (2009:178, 179, 180), with their structures reanalyzed according to the current proposal.

- (39) a. Max_i saw [PP a gun near himself_i/him_i].
 b. Max_i rolled [PP the carpet over himself_i/him_i]
 c. John_i turned [PP the anglepoise lamp towards himself_i/him_i]
 d. John_i found [PP a snake near himself_i/him_i]

The reason that either the reflexive or the pronoun is allowed when they are in adjunct PPs is because when the minimal search for a reflexive is initiated for the PPs, the inanimate SpecPPs will be returned. However, the inanimate SpecPPs are not possible antecedents for the animate reflexives. Consequently, the pronoun referring to the *v*P-internal subject can occur there instead. On the other hand, if a minimal search is initiated with the *v*Ps as the SD, then Spec*v*Ps will be returned. This gives rise to the interpretation where the reflexives inside the PPs are bound by the *v*P-internal subjects.

Interestingly, the non-complementary distribution between reflexives and pronouns are also detected if they are embedded inside regular small clauses:

- (40) a. Bill_i makes Mary angry at himself_i/him_i.
 b. Bill_i makes Mary tired of himself_i/him_i.
 c. [John and Tom]_i want Mike in each other_i's houses.
 d. John_i found Mary in himself_i/him_i. (Meaning: John found Mary became part of him.)
 e. John_i found Mary part of himself_i/him_i.

Such a distribution can be explained if PPs are optionally phases, and predicate adjectives are not phases.

Therefore we have derived the non-complementary distribution inside PPs without any additional complications besides the hypothesis regarding the phase status of PPs mentioned in (34).⁹

⁹In principle, the non-complementary distribution of reflexives and pronouns inside *picture*-NP

3.4.2 The complementary distribution

Interestingly, the current analysis also derives the complementary distribution between reflexives and pronouns in adjunct PPs when the reflexives and pronouns refer to the first object. As demonstrated by the examples in (41) (the syntactic analyses are mine),

- (41) a. (Reinhart & Reuland 1993:689)
 Max rolled [PP the carpet_j over *it_j/itself_j]
 b. (Hicks 2009:179)
 John turned [PP the anglepoise lamp_j towards *it_j/itself_j]
 c. (Bruening 2014:380)
 Max put [PP the books_j next to each other_j/*them_j].

The explanation of the complementary distribution of reflexives and pronouns in (41) is straightforward. Since the reflexives mismatch the *v*P-internal subject either in animacy (41a, b) or in number (41c), the *v*P-internal subjects are not legitimate antecedents even if they can be returned by minimal search. This leaves SpecPP as the only SD that may establish a binding relation for the reflexives. Since SpecPPs and conjunct phrases, examples of which are repeated below, can potentially be explained with a similar account. I would have to assume that subjectless DPs are optionally phases.

- (i) a. Lucie_i saw [a picture of herself_i/her_i].
 b. Max_i said that the queen invited [both Lucie and himself_i/him_i] for tea.

However, the peculiarity of *picture*-NPs and the inconsistency and individual variation in the judgment of sentences such as (i) prevent me from investigating these constructions any further. The peculiarity of *picture*-NPs can be illustrated with the fact that the reflexive inside it can be bound by an antecedent that is embedded inside a possessive phrase (sub-command contexts) (iia), whereas this is not possible for bare reflexives (iib):

- (ii) (Hestvik & Philip 2001:121)
 a. Clinton's car carried a picture of himself on the roof.
 b. *Clinton's car backfired/collapsed/explored behind himself.

The contrast in (ii) shows that the reflexives in *picture*-NPs are easier to be interpreted as a logophor, although I suspect that the reflexive in *picture*-NPs are always logophors, a possibility I previously argued against.

in (41) are legitimate antecedents, the use of reflexives is justified. As a result, the pronouns are predicted to be not possible.

Assuming the proposed MS-agree analysis, some properties of the distribution of reflexives and pronouns appears to be quite puzzling. For instance, at first glance, the examples in (42) present a challenge to the MS-agree analysis. We have to explain why *him* cannot be bound by *Max* in (42a), whereas such a binding relation is possible in (42b).

(42) (Reinhart & Reuland 1993:688)

- a. *Max_{*i*} stepped on him_{*i*}.
- b. Max_{*i*} looked around him_{*i*}.

With a closer examination of the differences between these two sentences, an answer suggests itself. The (intended) meanings of these two sentences might be spelled out as in (43).

- (43)
- a. Max_{*i*}'s stepping caused the fact that Max_{*i*} (was) on him_{*i*}.
 - b. Max_{*i*} looked (at) things around him.

Accordingly, the sentences seem to have the following structures, where the same PP analysis (that resembles a small clause analysis) is applied, and SpecPP is filled either with a PRO controlled by the subject (44a), or with a *pro* that refers to things in the context.

- (44)
- a. *Max_{*i*} stepped [_{PP} PRO_{*i*} on him_{*i*}].
 - b. Max_{*i*} looked [_{PP} *pro*_{*j*} around him_{*i*}].

Now it is clear why (44a) is out. Suppose that the pronoun is in fact a reflexive, (i) if a minimal search is initiated for the PP, the PRO that is controlled by *Max* will be returned, and therefore the reflexive must be bound by the PRO. A pronoun therefore is not allowed in the same context.¹⁰ (ii) if no minimal search is initiated until the

vP phase, then still *SpecvP Max* is returned as the antecedent of the reflexive. A pronoun is thus not allowed in this context as well. In other words, no matter whether the PP or the vP is the search target, the reference of the antecedent is always *Max*, which correctly entails that a reflexive rather than a pronoun must occur in (44a).

(45) Max_i stepped [$_{PP}$ PRO_i on himself $_i$].

The pronoun, by contrast, can occur in (44b) because *SpecPP* is not controlled by the *SpecvP*, and thus the PP can be the domain where a reflexive bound by *Max* is impossible. The presence of the pronoun is thus correctly licensed. In the semantic interpretation, the PP in (44b) is identical to a DP.

The above analysis of (44) receives support from constructions that have a PP as the subject:

- (46) a. [$_{PP}$ Behind the shed] needs mowing.
 b. [$_{PP}$ Before ten] is a bad time to call me.
 c. [$_{PP}$ Under the refrigerator] is disgusting!

Why can these PPs be in the subject position? These interesting cases can be explained if these PPs have a *pro* in their specifier position. For instance, (46a) means “the place behind the shed needs mowing,” and the subject [$_{PP}$ *pro behind the shed*] is interpreted equivalent to a DP as the *pro* refers to a place given by the context.

¹⁰Given that the main focus of this dissertation is on reflexive binding, I ignore details about how to derive Condition B. However, I assume an economy-based/competition-based approach for Condition A and B, an idea that is behind Chomsky’s (1981; 1986) formalization of his binding theory. Such an idea is well-explored in many influential works on binding theories (e.g., Hornstein 2001; Reuland 2001; Safir 2004). The central hypothesis is that pronouns are only allowed in contexts that reflexives are not allowed. Under the framework this dissertation adopts, binding relations can and only can be established by MS-agree. This implies that only reflexives can establish such a binding relation, given that only reflexives bear an unvalued [Ind:] feature that can initiate a minimal search. On the other hand, pronouns can contribute to two types of semantic relations: When a pronoun shares the same index feature with a D head that is outside the minimal phase containing the pronoun, an accidental binding relation can be formed (cf. Hicks 2009); when a pronoun’s index accidentally assigned the same reference as that of another D head, an accidental co-reference relation can be formed.

Reinhart & Reuland (1993) discuss another example that may challenge the current analysis.

(47) (Reinhart & Reuland 1993:688)

Max praised/examined the carpet_{*j*} underneath it_{*j*}.

This sentence should be put under comparison to another sentence Reinhart & Reuland (1993) discuss:

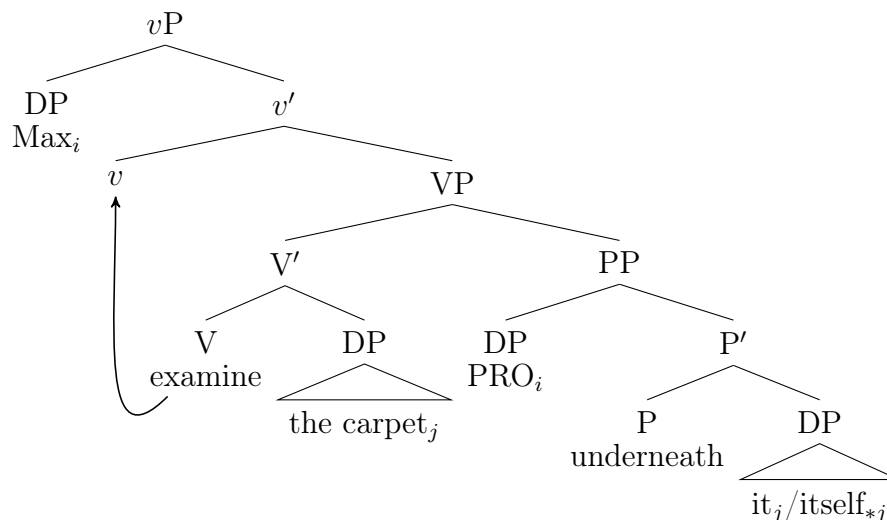
(48) Maxi rolled the carpet_{*j*} over *it_{*j*}/itself_{*j*}.

Note that a reflexive in replacement of the pronoun in (47) makes the sentence ungrammatical.

(49) *Max praised/examined the carpet_{*j*} underneath itself_{*j*}.

Taking the examples in (47)–(49) together, it can be concluded that the predicates certainly play a role. (48) follows from my analysis where [the carpet_{*j*} over *it_{*j*}/itself_{*j*}] is treated as a PP with *the carpet* in its specifier position. However, (47) and (49) are different. Given that *itself* in (49) cannot be bound by *the carpet*, we have to assume that *the carpet* is not in the specifier of the PP *underneath itself*. This seems to be on the right track, as grammatical version of the sentence in (47) cannot mean that the carpet is underneath itself. Instead, it means Max is underneath the carpet. I suggest that both the binding and semantic effects can be captured by the following analysis of (47) and (49):

(50)



Given (50), I predict that the binding relation between reflexive and the object *the carpet* (or its D head) cannot be established, given that minimal search into the *vP* and *PP* will either return *Max* or the *PRO* it controls.

A more serious problem in fact comes from the double object and dative constructions where the second object is a pronoun. We have seen that the proposed MS-agree analysis can handle the cases where the second object is a reflexive and it can be bound by either the subject or the first object, which the upward agree cannot handle. Therefore, the MS-agree analysis does not have any problem with the reflexive binding data in (51). However, when the second argument is a pronoun, the question is: why can't the pronouns occur in the second object position?

- (51) a. John_i wrote a letter to *him_i/himself_i.
 b. Mary_i sent a gift to *her_i/herself_i.
 c. Bill_i showed Mary *him_i/himself_i.
 d. Mike_i sold Mary *him_i/himself_i.

Similar constraints on the distribution of pronouns can be found in German. (52a) shows that when the reflexive and pronoun are inside a small clause/adjunct *PP*, the reflexive and pronoun are in a non-complementary distribution. However, (52b)

suggests that when the PP is not an adjunct, as Lee-Schoenfeld (2008) assumes that the P head in question does not assign θ -roles, pronouns are excluded in the same position.

(52) *German* (Lee-Schoenfeld 2008:292)

- a. Eri sah [PP direkt vor sich_i/ihm_i] eine Schlange auf dem
he saw directly in-front-of self/him a snake on the
Boden.
floor
‘He saw directly in front of himself/him a snake on the floor.’
- b. Die Frau_i interessiert sich nur [PP für sich_i(selbst) /sie_{*i}].
the woman interests self only for self(emphatic) /her
‘The woman is only interested in herself/*her.’

I do not have a good solution to this problem, but I believe the complementary distribution between reflexives and pronouns in (51) and (52) may receive an explanation if we have a more spelled out theory of pronouns and their syntactic distribution. Perhaps these examples suggest that the distinction between argument and adjunct PPs are important for Condition B, a position that is argued extensively by Reinhart & Reuland (1993) and many others. The question then is how to integrate the predicate-based analysis with an MS-agree analysis and meanwhile avoid the problems of the predicate-based analysis that I identified. A possible way out is that there are two types of Ps. One type of Ps does not assign θ -roles themselves, although they may assign case to its complement. An example of these P heads is the Ps in dative constructions such as those in (51). Given that these Ps do not assign θ -roles, they are not like predicates, and therefore cannot be considered phase heads (Lee-Schoenfeld 2008). By contrast, the other type of Ps is the one that assigns θ -roles. These are P heads that are in adjunct PPs, an example of which can be found in (48). The splitting of P heads into two types is standard, as argued by Reinhart & Reuland (1993). However, a serious problem of such an alternative analysis under MS-agree

is that now we lose our account of the fact that either the subject or the first object can be the antecedent of the second object (which is a reflexive) in double object and dative constructions, as these are PPs that are arguments of the predicate, and therefore we incorrectly predict that only the subject can be the antecedent in these constructions.

In fact, it is still unclear if modeling the distribution of pronouns and reflexives together in one theory is the best practice. Hicks (2009) and Charnavel (2018a) both point out that pronouns might observe different constraints from reflexives. If reflexive and pronoun are subject to different constraints, it is not well-motivated to consider sentences such as (51) as either counter-argument or evidence for a theory of reflexive binding. Since the MS-agree analysis is proposed primarily for reflexive binding, it may be justified that the theory should not concern itself (too much) about data like those in (51), in either a positive or negative way.

Another topic we have not even touched on is Condition C. There are good reasons for this, as Condition C may have something to do with semantic or pragmatic constraints on the use of referential expressions and pronouns in general (e.g., Safir 2004; Schlenker 2005a,b). The formalization of reflexive binding therefore might be independent from that of the distribution of referential expressions. I leave this for future research.

Finally, simplex reflexives and long-distance binding are intentionally avoided in the above discussion because they are orthogonal to the question at hand. For example, the simplex reflexive *ziji* in Chinese can be long-distance bound by an antecedent in a higher TP. Long-distance binding cannot be modeled with MS-agree. However, this is probably not a problem: Charnavel & Sportiche (2016) argue that plain anaphors must be locally bound, and long-distance binding is simply a consequence of the logophoric use of reflexives. Specifically, following Huang & Liu (2001) and Ke & Pires (accepted), I assume that the long-distance bound reading of *ziji* reflects its

logophoric use: being a logophor, a reflexive is converted to a pronominal (Reinhart & Reuland 1991), and therefore it can thus co-refer with a long-distance antecedent.

3.5 A remaining issue

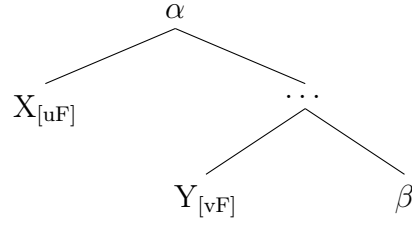
Under the current framework, since we do not assume feature percolation or projection, a question immediately arises: Is c-command lost with minimal search? I will address this question by considering the cases of MS-agree and MS-labeling I have discussed so far in Chapter II and III.

The answer seems to be positive. Traditionally, the c-command relation can be formalized as a relation between sets, as well as between heads. If, under the current assumptions, sets do not have features, then we cannot compute the c-command relation between sets in syntax; for example, the c-command relation between a DP antecedent and a reflexive cannot be established. However I will point out below that c-command can be a good approximation of the syntactic relation established by minimal search. This is mainly because when a set headed by $X_{[vF]}$ c-commands another set headed by $Y_{[uF]}$, $X_{[vF]}$ can be returned by minimal search triggered by $Y_{[uF]}$ in plausibly all known cases. Let us use some examples to make the discussion more concrete.

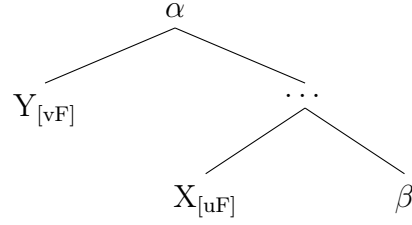
First, let me introduce the symbols I will use below. Capital letters are used to refer to lexical heads whose category is unspecified, with C, D, N, P and T (as well as Adj and Adv) reserved for heads with their syntactic categories specified. So X simply means a head with an unknown syntactic category, whereas C means the C head of a CP set/phrase.

As we mentioned, there are three configurations (53) where minimal search is initiated to find matching heads for the trigger with uFs. Minimal search may either search down from the sister set or the next built phase above the trigger.

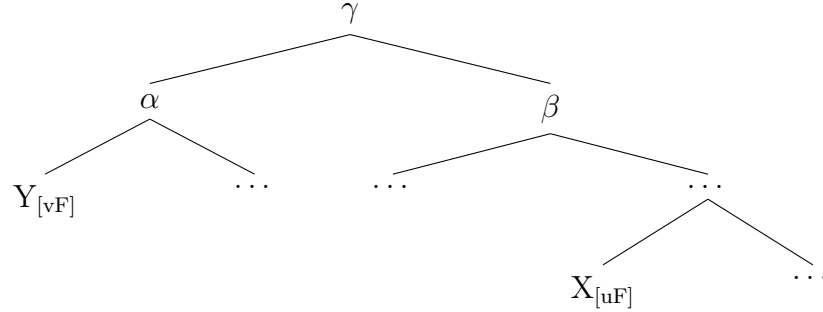
- (53) a. $X_{[uF]}$ merges with $Y_{[vF]}$ or a set containing $Y_{[vF]}$: $X_{[uF]}$ c-commands $Y_{[vF]}$;



- b. $Y_{[vF]}$ merges with $X_{[uF]}$ or a set containing $X_{[uF]}$: $Y_{[vF]}$ c-commands $X_{[uF]}$;



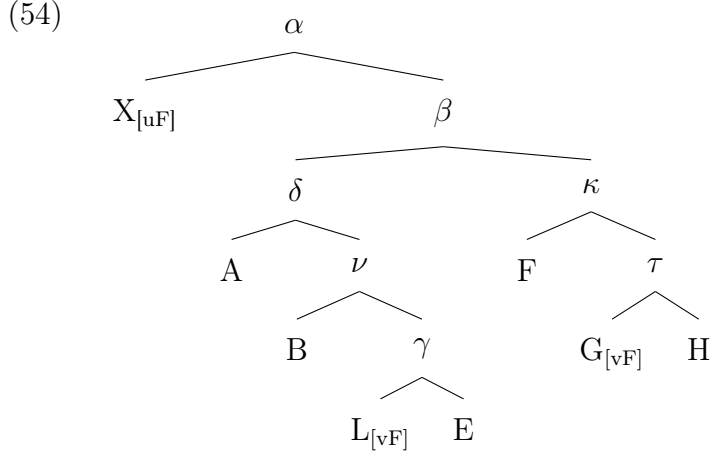
- c. A set containing $Y_{[vF]}$, α , merges with a set containing $X_{[uF]}$, β .



If a head X with a $[uF]$ enters into the derivation, a minimal search with the sister of X as the SD will be triggered, as shown in (43a). In this case, a head with a $[vF]$ contained in the sister of X , for instance, Y , may be returned. Since Y is contained in the sister of X , X always c-commands Y . Thus minimal search and c-command in this situation do not lead to different results.

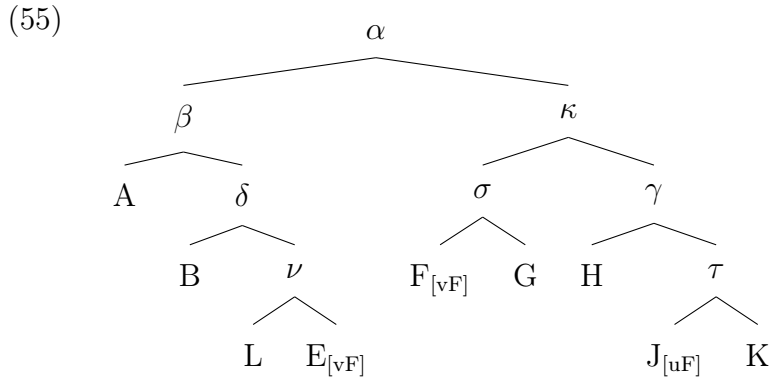
(43b) and (43c) are cases where minimal search takes the next built phase above the trigger as the SD and searches downward.

As for the configuration type in (43a), a problem may arise when there are two candidates contained in the sister of X :



Imagine that X in (54) is a T head with unvalued ϕ -features, and it agrees with the head of the subject (somewhere inside δ) rather than that of the object τ . Could the problem shown in (54) occur? That is, could the head of the object be returned by mistake if the head of the subject $L_{[vF]}$ is deeply embedded inside σ ?

Similarly, for the second cycle minimal search that takes the next built phase above the trigger as the SD, imagine a syntactic object as below.



If set α is the next built phase above J_{uF} , and the $[uF]$ on J is not valued when α is built, then a minimal search with $ST = [uF]$, $SD = \alpha$ is initiated. The head $E_{[vF]}$ is supposed to be returned, and it could be correctly returned when it is not that deeply embedded inside β . But could it not be returned when it is deeply embedded (as that in (55)), and $F_{[vF]}$ inside σ is accidentally returned instead?

To state the problem in another way, can we find a case where σ in (54) does not

contain any head with ϕ -features or an index feature above $L_{[vF]}$? Or is it possible not to have a head bearing ϕ -features or an index feature above $E_{[vF]}$ in (55)? I believe that as far as ϕ -feature or index feature agreement is concerned, it is very difficult to find a configuration that contains only one head with ϕ -features or an index feature such that this head is deeply embedded in a position like the item L in (54) or E in (55). In other words, empirically, we cannot find cases where (i) a head with ϕ -features or an index feature is embedded that deep and (ii) there is no head with ϕ -features or an index feature above it.

Let us assume that the β set in (54) is one of the major types of phrases: CP, DP, TP, vP , VP and PP,¹¹ and let us focus on ϕ -feature agreement. CP and DP have C and D as their heads, and these heads are widely assumed to bear ϕ -features (including default ones). Therefore, if δ in (54) is a CP or DP, when they do not have a specifier, the first head contained in CP and DP, i.e. the position of the item A in (54), would be a C or D head. Therefore, minimal search taking β as the SD will return their heads, i.e., the C or D head in the A position. That is, minimal search cannot go any deeper to access the item $L_{[vF]}$ as the relevant features on the item A would cause an intervention to those on the item L.

There must also be a head inside TP and vP that bears (underspecified or default) ϕ -features. For instance, the following examples show that there must be a head in TP and vP that can be targeted by minimal search, preventing other heads more deeply embedded in the TP or vP or the object from agreeing with the matrix T head.

(56) To sleep under the stars is the only thing Eli wants now.

The null head is likely a PRO for (56), as shown in the analysis below (PRO could raise to the specifier of the embedded TP):

¹¹At least in languages that have DPs, DPs but not NPs are used for ϕ -agreement purposes; in addition, VPs are also not independently used in argument positions for agreement purposes. Therefore, I do not consider NPs and VPs here.

(57) [_{TP} To PRO sleep under the stars] is the only thing Eli wants now.

The null head being returned gives rise to a default feature agreement on the matrix. Of course, assuming that it is a PRO that agrees with the matrix T head in (56) might put us in an odd situation, but this seems empirically not too problematic for TPs. Can we pose a similar analysis for *v*Ps, an example of which is shown below?

(58) [_{vP} Wasting resources] is the first thing I would avoid.

Interestingly, there is no contrast on the matrix T agreement between (59) and (60), which may suggest that the matrix T in fact agrees with *v* rather than Spec*v*P.

(59) David being the chairman has meant more work for all of us.

(60) These people being the committee members has meant more work for all of us.

However, there are also independent reasons why Spec*v*Ps in (59) and (60), namely *David* and *these people*, cannot agree with the matrix T head. This is because Spec*v*Ps are not assigned grammatical cases, as illustrated by the following sentences:

(61) Him/*He being the chairman has meant more work for all of us.

(62) Them/*they being the committee members has meant more work for all of us.

The Spec*v*Ps in these subjects cannot be nominative but must be accusative, which is a default case (Schütze 2001). As a general observation, a DP must be marked by a grammatical case for it to be involved in an agreement relation; otherwise it will give rise to a default agreement (see e.g., Koopman 2006 and citations therein), as frequently discussed regarding DPs that are not marked with an abstract case in Icelandic (Sigurðsson 1991), as well as various other languages we discussed in Chapter II, including Georgian (Rezac 2004), Hindi-Urdu (Keine & Dash 2017), and

Tsez (Polinsky & Potsdam 2001). Therefore, either it is the specifier (its head in fact) or the *v* head in the subject *v*P that agrees with the matrix T, the result is the default third person singular agreement. This is in fact predicted given MS-agree as defined in Chapter II, because either the D head of Spec*v*P or the *v* head can be found by minimal search, and therefore either of them may agree with T.¹²

When PPs are used as arguments, the analysis is similar to TPs, as I have previously discussed:

- (63) a. [pro behind the shed] needs mowing.
 b. [pro before ten] is a bad time to call me.
 c. [pro under the refrigerator] is disgusting!

The reader may wonder if some type of DPs may cause a problem, e.g., PossP (possessive DPs). For example, (64) shows that *these people* cannot agree with the T head and the reflexive.

- (64) [These people_{*i*}'s [boss_{*j*}]] was/*were angry with himself_{**i/j*}.

I have to assume that the possessor is somewhere at a position lower than the D head of the subject DP, as in the structure that Hicks (2009) argues for. Hicks (2009:153) assumes that there is an *n*P below the D head and it hosts the possessor in its specifier position before the possessor moves to the specifier of NumP, which is posited right below the D head. Accordingly, the structure of the subject DP in (64) is shown below:

- (65) [_{DP} D [_{NumP} [_{DP} these people] Num [_{*n*P} ⟨these people⟩ *n* [_{NP} boss]]]]

With the above structure assumed, the highest D head will always be returned by minimal search, no matter whether it is triggered by the T head for subject–verb agreement or by the reflexive for reflexive binding.

¹²This phenomenon is potentially very interesting, as it can explain why optionality sometimes arises for agreement with different heads. Coordinate phrases are a potential interesting empirical domain) at the same hierarchical level that can all be returned by minimal search.

In summary, I argue that the relations established by minimal search are sometimes different from c-command relations. Previously we could say that a DP binds a reflexive if it c-commands and co-refers with the reflexive, but now, given that we have excluded feature percolation or phrase projection from the system, we cannot establish such a c-command relation between the DP antecedent and the reflexive. What we can establish, by minimal search, is a relation between the D head of the DP antecedent and the D head of the reflexive. No c-command relation holds between these two D heads. However, I suggested that this does not cause an empirical problem, as the relations returned by minimal search give essentially the same results as we want with c-command. This is because when a c-commanding set is relevant for agreement or binding, it is almost always that its head can be returned by minimal search. In a sense, minimal search establishes the so-called c-command relations in a different way. Since minimal search has two cycles which involve different types of SDs, we see that sometimes a set labeled by a head with [uFs] c-commands a head with [vFs]; this is a relation returned by the first cycle minimal search. Other times, however, we see that a head with [uFs] is c-commanded by a set whose head is with [vFs]; this is a relation returned by the second cycle minimal search. I leave for future research a more comprehensive study of whether we can reduce c-command completely to minimal search.

As a final note, I assume that in sentence processing, given that labels can be used, the c-command relation is a good approximation of the syntactic relations that have been formed by minimal search. For instance, in syntax there is no c-command relation between the D head of an antecedent DP and the D head of a bound reflexive; nevertheless, after the structure is successfully labeled through the labeling algorithm, we can formalize the c-command relation between those two labeled sets. I assume that sentence processing can use labels to refer to sets. In this sense, the computation of the c-command relation can be important for sentence processing, an issue I will

deal with in Chapter V.

CHAPTER IV

The Interpretability of Φ -Features

4.1 Introduction

In Chapter II and III, I made the claim that with minimal search-based agree, various types of agreement and reflexive binding can have a unified analysis. This captures the similarities between subject–verb agreement and reflexive binding. One of the goals of this chapter, however, is to show that subject–verb agreement and reflexive binding are different in a non-trivial way. A crucial difference resides in the semantic interpretability of the ϕ -features involved. That is, the number, person and gender features on a T head (e.g., *was* vs. *were*) that agree with a subject are not semantically interpretable (Chomsky 2000a, 2001b). By contrast, these features on a reflexive (e.g., *himself* vs. *themselves*) are semantically interpretable. The semantic uninterpretability of the ϕ -features on T is relatively uncontroversial, or at least not as controversial as the semantic interpretability of the ϕ -features on reflexives and pronominals in general. This chapter therefore concentrates on the semantic interpretability issue associated with the ϕ -features on reflexives and more generally, bound pronominals. I argue that the ϕ -features on reflexives and bound pronouns are in fact semantically interpretable.

As far as I know, there is not much discussion on the semantic interpretability of the ϕ -features on reflexives. However, influential theories have been proposed con-

cerning the semantic (un)interpretability of ϕ -features on bound pronouns (e.g., Heim 2008; Kratzer 2009). These theories claim that the ϕ -features on bound pronouns are semantically uninterpreted. Although these studies are not mainly about reflexives, they are directly relevant to the question at hand as both bound reflexives and bound pronouns are standardly treated as bound variables in formal semantics. As I will argue below, the conclusions of these studies should in principle be applicable to all bound variables, bound reflexives included.

Let us start with a few examples:

- (1) a. only I took a picture of **myself** →
 others did not take pictures of **themselves**
- b. only I did **my** homework →
 others did not do **their** homework
- (2) a. only Mary took a picture of **herself** →
 others did not take pictures of **themselves**
- b. only Mary did **her** homework →
 others did not do **their** homework

The (b) examples in (1) and (2) are heavily discussed in the literature regarding bound pronouns (e.g., Heim 2008), and the (a) examples are similar examples for reflexives. A comparison of bound pronouns and reflexives suggests that both are subject to the same issue of semantic interpretability. The examples in (1) show that in the implied meanings of these *only*-sentences (the meanings represented after the arrows), the person and number features are not carried over from the regular meaning of these sentences. The examples in (2) suggest that the same holds for the gender feature. Take (1a) as an example. Suppose that the ϕ -features on the reflexive are semantically interpreted and thus cannot be ignored, the regular and implied meanings of the *only*-sentence (1a) should be something like that in (3a,

b), respectively, if the context concerns two individuals only, {I, John}. The implied meaning is equivalent to the plain English in (4), which capture the referential reading but not the bound reading of the reflexive in (1a):

(3) a. [only 1ST-SG₁] 2 [t₂ took a picture of 1ST-SG₁]

b. \neg [John₃] 4 [t₄ took a picture of 1ST-SG₄]

(4) John did not take a picture of myself/me.

Therefore, if the ϕ -features on the bound pronouns and reflexives are interpreted, we cannot compute the implied meaning correctly. In other words, we are unable to derive the bound reading of reflexives/pronouns here, as the only available meaning in the above *only*-sentences is their referential reading.

Before we discuss more about the ϕ -features on reflexives and bound pronouns, I would like to first define the concept “interpretability” and explain how semantic ϕ -features are interpreted, which should be in turn embedded in the background of the feature system I am assuming. Then in the second section of this chapter, I will review some previous accounts addressing the question as to why the ϕ -features in sentences such as (1) and (2) are not interpreted. I then provide novel data from English, Chinese, Arabic and Russian to support the argument that ϕ -features, the number feature specifically, on bound pronouns are in fact semantically interpretable. Finally, an account is provided to explain why ϕ -features are seemingly ignored in some particular environments such as focus contexts.

4.1.1 The feature system

Following the multi-valent feature theory proposed by Adger (2010) (also see Adger 2003; Adger & Svenonius 2011), the term “feature” is defined as in (5).

(5) Definition of feature

A feature is an ordered pair $\langle \text{Att}, \text{Val} \rangle$ where

- a. Att is drawn from the set of attributes, A, B, C, ...
- b. Val is drawn from the set of values, +, −, ∅, A, B, C, A/B, ...

Each feature is composed of two levels, with the first level the names of the attributes, and the second level their values. The values are drawn from a mixture of binary feature values (+ or −) and descriptive feature values (e.g., A, B, C). The binary values capture features that are either present or not present, and the descriptive features are designed for features that have more than two values. ∅ indicates null feature values, and “/” means ambiguous feature values. When a feature is unvalued, or the feature value is underspecified, the Val slot is left empty. I adopt a slight modification of Adger’s (2010) system so that it is compatible with the feature system proposed by Lewis & Vasisht (2005) (LV05). For instance, the categorical features such as “T, D, P” are treated as attributes by Adger (2010). However, this dissertation (and LV05) assumes that these category names are secondary feature values under the attribute “Cat(egory).”

In this dissertation, I do not use “interpretable” as “interpretable to the CI/LF interface,” which is what “interpretable” means in Chomskyan system (e.g., Chomsky 2000a). Such a definition of interpretability is what is normally found in the literature. Instead, When I say a feature is “interpretable,” I would also mention whether the feature is interpretable to the CI/LF interface or to the SM/PF interface. The former is also called semantic interpretability, and the latter phonological interpretability. A semantic feature is interpretable to the CI/LF interface, and a phonological/morphosyntactic feature is interpretable to the SM/PF interface. Some features may be interpretable to both interfaces (cf. Epstein 2007). Therefore, the interpretability of a feature is relevant to both the CI/LF and SM/PF interfaces. This conception of interpretability is close to the “readability” of the features to the interfaces (Chomsky 2000b:17).¹

¹I use CI and LF interchangeably, and SM and PF interchangeably. This is necessary as especially

(6) uses *was* and *himself* as examples to show how features of lexical items are represented before they enter into the syntactic derivation; that is, before agreement and binding apply.

- (6) a. *be*: {Cat:T, Tense:PAST, Num: , Person: }
 b. *himself*: {Cat:N, Case: , Num:SG, Person:3RD, Gender:MAS, Animacy:+}

Note that on *himself*, there are ϕ -features with values taken from the lexicon, whereas no such values are present on *be*. The ϕ -features on *be* are not semantically interpretable. In addition, the ϕ -features on *be* do not bear values. The feature values depend on and co-vary with those of the ϕ -features on a D head (e.g., of a DP subject) that *be* agrees with. By contrast, the ϕ -features on the reflexive *himself* are semantically interpretable. But one may still wonder what exactly the semantic interpretability means. The next subsection addresses this question: Φ -features are semantically interpreted as presuppositions.

4.1.2 Semantic ϕ -features as presuppositions

In formal semantics, pronouns are usually analyzed as a combination of ϕ -features and an index (Cooper 1979; Heim 2008; Charnavel 2018b), as shown below in (7):

- (7) pro_i : [Person [Gender [Number [pro_i]]]]

- (8) a. I_2 : [1ST [SG [pro_2]]]
 b. she_3 : [3RD [FEM [SG [pro_3]]]]

On the other hand, reflexives are either assumed to bear both ϕ -features and an unvalued index feature (Hicks 2009), or only unvalued ϕ -features (Reuland 2001; Heinat 2006, 2008). As discussed in Chapter III, I basically follow Hicks's analysis, so I assume that a reflexive bears both ϕ -features and an unvalued index feature. After

in the semantics literature, LF and PF are used instead of CI and SM.

a binding relation is established, the unvalued index feature will be valued, resulting in a feature bundle similar to bound pronouns:

(9) refl_i : [PERSON [GENDER [NUMBER [refl_i]]]]

(10) a. myself_4 : [1ST [SG [refl₄]]]

b. herself_5 : [3RD [FEM [SG [refl₅]]]]

A comparison between (7, 8) and (9, 10) confirms that in terms of the semantics of the ϕ -features and the index value, bound pronouns and bound reflexives in fact behave essentially in the same way semantically.

Starting from Cooper (1983), the gender feature and other ϕ -features are analyzed as presuppositions (i.e., Heim & Kratzer 1998; Schlenker 2003; Sauerland 2003; Heim 2008; Charnavel 2018b, among others). Under this analysis, ϕ -features denote partial identity functions, introducing presuppositions to restrict the domain of pronouns' index. As exemplified by (11), the singular number feature presupposes that the referent of the pronoun's index includes only one atom (11a), and the first-person feature presupposes that the speaker must be included in the referent (11b). The combination of these two presuppositions is the presupposition carried by the singular first-person pronoun (11c).

(11) a. $\llbracket \text{SG} \rrbracket = \lambda x : x \text{ is an atom. } x$

b. $\llbracket \text{1ST} \rrbracket^{g,c} = \lambda x : x \text{ includes the speaker in } c. x$

c. $\llbracket \text{I}_j \rrbracket^{g,c} = g(j)$, where $g(j)$ is defined only when $g(j)$ is an atom and it includes the speaker in c .

It is obvious that the ϕ -features on verbs or T heads cannot be interpreted as presuppositions, although they give us information about the ϕ -features on the nominals they agree with. It is in this sense that the ϕ -features on reflexives and bound pronouns are semantically interpretable, but those on verbs or T heads are not.²

²I assume that the ϕ -features on T heads or verbs are phonological ϕ -features, not semantic ϕ -

4.2 A challenge: the PF Feature Transmission Approach

Data in (1b) and (2b), repeated below in (12), nevertheless cause a problem to the presupposition analysis of ϕ -features.

- (12) a. only I did **my** homework \rightarrow
 others did not do **their** homework
 b. only Mary did **her** homework \rightarrow
 others did not do **their** homework

Recall that the problem is, if we keep the ϕ -features on bound pronouns in the computation of the implied meanings of *only*-sentences, we would not be able to have *their* in (12). Instead, we should always have the same pronouns as those in *only*-sentences, i.e., *my* for (12a), as is demonstrated by (13).

- (13) If bound pronouns bear semantic ϕ -features:
 [only 1ST-SG₁] 2 [t₂ did 1ST-SG₂'s homework] \rightarrow
 \neg [others₃] 4 [t₄ did 1ST-SG₄'s homework]

The following example suggests that the same problem arises in the context of VP ellipsis:

- (14) a. I_i did my_i homework. Mary_j didn't ([_{VP} do her_j homework]).
 b. I_i did my_i homework. [Mary and Mike]_j didn't ([_{VP} do their_j homework]).
 (15) a. Mary₅ didn't 6 [_{VP} t₆ do 1ST-SG₆'s homework]
 b. [Mary and Mike]₅ didn't 6 [_{VP} t₆ do 1ST-SG₆'s homework]

The examples in (14) indicate that the elided VPs, which are parallel to the antecedent VPs in the antecedent sentences, include a bound pronoun. The examples in (15)

features. I will return to this point later when I discuss the ϕ -features on collective nouns. Interested readers are referred to Shen & Smith (2019) for a recent discussion on the split of the semantic and phonological ϕ -features, the latter of which they refer to as morphological features.

suggest that if the antecedent VPs are copied to the elided VPs, and if the ϕ -features on bound pronouns are semantically interpreted, the only reading available is the referential interpretation of the pronouns. In that case, the elided pronouns cannot be properly bound by their local antecedents, *Mary* or *Mary and Mike*, as the ϕ -features of the pronouns and their local antecedents are different. In other words, the bound reading of the pronouns is not available under VP ellipsis if the ϕ -features on the bound pronouns are semantically interpreted.

Heim (2008) (see also Kratzer 1998, 2009; Stechow 2003 and Rullmann 2004) argues that a possible solution to the above interpretability problem caused by examples in (12) and (14) is to assume that bound pronouns do not bear semantic ϕ -features. Bound pronouns are pronounced as if they have ϕ -features because they inherit the ϕ -features from their binders. (16) gives an illustration of this solution.

(16) If bound pronouns do not bear semantic ϕ -features:

- a. [only 1ST-SG₁] 2 [t₂ did \emptyset_2 's homework] \rightarrow
 \neg [others₃] 4 [t₄ did \emptyset_4 's homework]
- b. I₁ did my₁ homework. Mary₅ didn't 6 [VP t₆ do \emptyset_6 's homework]

(16) indicates that if bound pronouns do not have semantic ϕ -features, the problem is resolved, as now the interpretation of the bound pronouns is determined by their local binders.

What we need then is an operation for the correct pronunciation of the bound pronouns. How does the SM interface know that the bound pronoun in (12) must be pronounced as a first-person singular? This is accomplished by the following Feature transmission rule operating at the phonological form (PF, which is SM in this dissertation):

(17) *Feature Transmission under Variable Binding* (Heim 2008:50)

In the derivation of PF, all features of a DP must be copied onto all variables

that it binds.

Such an analysis has been called the MORPHOSYNTACTIC ACCOUNT (Bassi & Longenbaugh 2017). The core proposal of this account is summarized in (18).

(18) *The morphosyntactic account*

- a. ϕ -features of bound pronouns are present only at PF, and are not semantically interpretable.
- b. ϕ -features of bound pronouns are different from those of referential pronouns which are semantically interpreted as presuppositions.

It is worth noting that PF Feature Transmission as in (17) is not compatible with the Y-shape language model I assume in this dissertation (see Chapter I), because it assumes the following: (i) Binding of the pronoun occurs in the logical form (LF, which is CI in this dissertation) (Heim & Kratzer 1998); (ii) PF feature transmission relies on the binding relation. This implies that the binding relation established at LF must feed feature transmission at PF. This is not available under the framework adopted in this dissertation, because after a syntactic object is assembled and is transferred to the interfaces (LF and PF), what occurs in one interface will not affect the other interface. As a result, the binding relation, if it is established at LF, cannot feed feature transmission at PF.

A simple way out is make a slight modification on Heim & Kratzer's (1998) system. That is, binding relations are not established at LF, but in the syntax, just as Hicks (2009) has proposed for reflexive binding. After a binding relation is established by agree, the index values of the binder and bindee, which are variables, must be identical. These variables are assigned specific index values for semantic interpretation at LF, requiring the identity relation between the index values of the binder and the bindee to be kept unchanged.

Now I have shown that the PF Feature Transmission Approach does constitute

a serious challenge to the presupposition analysis of ϕ -features. Although it should be well-acknowledged that the PF Feature Transmission Approach provides a simple and neat account, let us bring to the discussion some more data which argues against this approach.

4.3 Challenges to the challenge

In this section, I will first review a few studies which suggest that the PF Feature Transmission Approach encounters many difficulties, and that ϕ -features on bound pronouns are only ignored under particular contexts. I will then show that a revised version of the morphosyntactic account that keeps the core assumptions of the PF Feature Transmission Approach has the potential to overcome those challenges. Therefore, the debate has not been settled conclusively. Novel evidence in favor of the presupposition account but against the PF Feature Transmission Approach and its revised versions is thus provided in the next section.

4.3.1 Evidence

Bassi & Longenbaugh (2017) argue that the same problem of semantic interpretability is observed in the following example. The subscript “F” signifies a focus on a certain part of the sentence.

- (19) Only the students who took phonology_F thought it was cool.

In this example, the pronoun is not bound by an antecedent as it is not c-commanded by the noun *phonology* with which it co-refers. However, (19) has a familiar implication as we have seen in sentences such as those in (12).

- (20) only the students who took phonology_F thought it was cool \rightarrow
students who took [other classes]_j did not think they_j are cool

In the above example, the number feature on the pronoun appears to be not interpreted. Such a pattern can be easily extended to the person and gender features:

- (21) a. only the students who met you_F thought you were smart →
 the students who met other people_j did not think they_j were smart
 b. only the students who met [the girl]_F thought she was smart →
 the students who met [the boy and the man]_j did not think they_j were
 smart

Although the sentences in (21) have different readings in addition to the one indicated here, the intended reading listed in (21) does suggest that the ϕ -features on pronouns (so-called donkey pronouns) that are not bound can be not interpreted, just like those on bound pronouns. Bassi & Longenbaugh (2017) adopt the E-type analysis of the E-type pronouns (Parsons 1978; Cooper 1979; Heim 1990; Elbourne 2005) for (20). That is, donkey pronouns are taken as definite descriptions that contain a bound variable. For instance, the donkey pronoun *it* in (20) is equal to *the class they took*, where *they* is bound by the subject of the sentence *the students who...*:

- (22) only [the students who took phonology_{F,5}] 6 [t₆ thought [it₅ = the class they₆
 took] was cool →
 the [students who took [other classes]₅] 6 [t₆ thought [they₅ = the classes they₆
 took] was cool

Since the donkey pronouns in these examples are not bound pronouns themselves, Bassi & Longenbaugh (2017) argue that in order to account for the uninterpretability of the ϕ -features on these pronouns, something more than Heim (2008) would need to be proposed to deal with the semantically uninterpretability of unbound pronouns (see Sudo 2012, 2014 for a relevant argument along these lines).

Bassi & Longenbaugh's (2017) argument does suggest that the uninterpretability of ϕ -features may not be associated with binding at all, as such a phenomenon is

observed both with binding and without binding. However, it should be pointed out that this is a direct counterargument against the PF Feature Transmission Approach. This is because it does not say anything about bound pronouns. It remains a question as to whether bound pronouns should be minimal pronouns which do not have semantic ϕ -features. In short, Bassi & Longenbaugh (2017) provide only an indirect argument against the PF Feature Transmission Approach, because the PF Feature Transmission Approach is primarily motivated by the interpretability of ϕ -features on bound pronouns.

Spathas (2010), however, provides direct evidence against the PF Feature Transmission Approach based on mixed gender agreement in Greek. In Greek, a bound pronoun agrees with its binder in either its semantic/natural gender feature or its morphosyntactic gender feature. (23) indicates that the bound pronoun could optionally be *her* or *its*, matching the natural feminine gender or the neuter morphosyntactic gender feature of its binder ‘*the girl.NEU*,’ respectively. Note that the different choices of bound pronouns do not distinguish meaning.

- (23) Mono to koritsi mas dhiakosmise to dhomatio tis/tu.
 only the.NEU girl.NEU our decorated the room her/its
 ‘Only the girl decorated her room.’

Examples such as (23) are problematic for the PF Feature Transmission Approach. If ϕ -features are copied to the bound pronoun from its binder at PF, only the neuter gender feature should show up on the bound pronoun. Contrary to the prediction, however, the bound pronoun can be optionally pronounced with a semantic gender feature.

4.3.2 Possible responses to the challenges

There are two possible responses to the Spathas’s (2010) argument based on mixed gender agreement in Greek. The PF Feature Transmission Approach could assume

that nouns like *koritsi* ‘girl.NEU’ in Greek bear both the feminine natural gender [NatGender: FEM] and the grammatical neuter gender [GramGender: NEU] features as morphosyntactic features. If this is the case, then both of these features can be transmitted from the binder through PF feature transmission. The only thing we need now is some specific pronunciation restrictions to obtain the right spell-out for a specific language (cf. Kratzer 2009). For example, there may be a pronunciation rule determining whether either [GramGender: NEU] or [NatGender: FEM] must be pronounced on the bound pronoun in Greek.³

If the pronunciation rule discussed above can be added, and if we ignore any unfavorable consequences associated with this language and construction-specific stipulation, the PF Feature Transmission Approach cannot be ruled out by the mixed gender agreement data in Greek (23), particularly because the choice between ‘her’ and ‘its’ as a bound pronoun does not have different semantic consequences. Nevertheless, if the choice between bound pronouns had an effect on the interpretation of the bound pronouns or the sentences, then the PF Feature Transmission Approach could hardly be sustained, as it cannot explain why the pronunciation of different forms of pronouns at PF would lead to different interpretations of the pronouns or that of the sentences at LF. In the next section, I will provide data with exactly the following pattern: Choice of bound pronouns alters the semantic interpretation of the

³Spathas (2010) points out a potential difficulty of such a PF Feature Transmission Approach: In Greek, a predicate adjective agrees with the subject in gender, exemplified by (i), and the morphosyntactic gender feature but not the natural gender feature can surface up at the agreeing adjective.

(i) *Greek* (Spathas 2010:222)

To koritsi ine omorfo/*omorfi.
the.NEU girl.NEU is pretty.NEU/pretty.FEM

‘The girl is pretty.’

If the PF Feature Transmission Approach assumes that both the natural gender and grammatical gender features are morphosyntactic features that can be transmitted at PF, it would then need to stipulate a specific pronunciation rule for resolution of the two morphosyntactic features in pronunciation.

sentences. Let us refer to the response just discussed as the REVISED PF FEATURE TRANSMISSION APPROACH.

Another possible response to Spathas's (2010) argument is more interesting to us. It could be assumed that nouns like *koritsi* 'girl.NEU' in Greek bears a phonological feature [GramGender: NEU] and a semantic feature [NatGender: FEM]. This assumption is quite standard. One may then hypothesize that both phonological/morphosyntactic and semantic features of a binder are transmitted to a corresponding bound pronoun through co-indexation. Since both [GramGender: NEU] and [NatGender: FEM] are transmitted to the bound pronoun, this analysis could argue that either one of them may be pronounced on a pronoun. Meanwhile, this analysis would require that the phonological features but not the semantic features on the binder are obligatorily pronounced. This is different from that on a bound pronoun, where either the phonological or the semantic feature can be pronounced. Given that this analysis assumes that both the semantic and phonological features are transmitted from the binder to the bound pronoun, let us call it the FULL FEATURE TRANSMISSION APPROACH.

The Full Feature Transmission Approach is in fact motivated by Heim's (2008) solution to a problem due to the split binding examples discussed by Rullmann (2004), as exemplified by (24):

(24) **Context:** Imagine John is in one room with all of his ex-wives, one of the ex-wives says to John:

Sentence: All of us wanted to separate on peaceful terms, but only I wanted us to stay close after the divorce.

(25) only I_1 wanted [us = I_1 + John] to stay close after the divorce \rightarrow
for another ex-wife x , x did not want [x + John] to stay close after divorce

(25) indicates that only a variable inside the split pronoun co-varies with its binder. The PF Feature Transmission Approach encounter a serious difficulty accounting

for such examples. Heim (2008) thus suggests that for the proper pronunciation of the person feature on the split pronoun *us* in (25), PF feature transmission should be coupled with the following pronunciation rules in PF (*i* and *j* are indices of pronouns):

- (26) a. If *i* or *j* is unspecified for person, then leave *i+j* unspecified.
- b. Otherwise, if *i* or *j* is 1st person, then specify *i+j* as 1st person.
- c. Otherwise, if *i* or *j* is 2nd person, then specify *i+j* as 2nd person.
- d. Otherwise, specify *i+j* as 3rd person.

Going back to (25), if only the person feature is concerned, the split pronoun *us* is semantically equal to [*I* + *John*], whose person feature is [*1st* + *3rd*]. (26b) is therefore the rule that should be applied to this case, and [*1st* + *3rd*] is specified as 1st person in PF.

Heim (2008) acknowledges that the above pronunciation rules do not provide a satisfying solution to the split pronoun problem. The above pronunciation rules are specifically for the person feature. Specific rules must be created for other ϕ -features such as gender and number as well (although the rules are probably simpler). On the contrary, if ϕ -features on bound pronouns are semantically interpreted, the rules in (26) are no longer needed, and the computation of the person feature on a split pronoun would naturally follow from the semantics of ϕ -features, which are independently needed for referential pronouns. Everything will then follow from the presupposition approach to ϕ -features (see Sudo 2014 for a relevant discussion).

It must be noted that the Full Feature Transmission Approach provides a better analysis of the split pronoun data and at the same time avoids the above problems associated with Heim's (2008) analysis. Recall that the Full Feature Transmission Approach sustains the core hypothesis of the PF Feature Transmission Approach that ϕ -features on a bound pronoun are transmitted from their binder through co-indexation, but it goes a step further by assuming both the semantic and phono-

logical/morphosyntactic features are transmitted to a bound pronoun from its binder through co-indexation. This analysis solves the problem caused by Rullmann’s (2004) data as in (24) without resorting to the pronunciation rules as in (26). This is because the semantic ϕ -features transmitted to the index of the bound pronoun can be used for the computation of the ϕ -features of the split pronoun, and the computation results will be correctly pronounced. The phonological features transmitted to the pronoun index will simply be ignored, as there is no way to compute the ϕ -features based on the phonological features if pronunciation rules such as those in (26) are not stipulated.

Can the Full Feature Transmission Approach solve the interpretability problem in *only*-sentences, which motivates the PF Feature Transmission Approach? I think the answer is positive. The Full Feature Transmission Approach assumes that bound pronouns are minimal pronouns without base-generated ϕ -features, and this assumption applies to the bound pronoun in the implied meaning as well as in the ordinary meaning. Of course, the details of this analysis need to be worked out, and this analysis must in certain ways interact with the semantics of *only* and focus. However, it seems that this alternative does deserve serious consideration: It reserves Heim’s insight that bound pronouns must be minimal pronouns without ϕ -features; meanwhile, it resolves the problems encountered by Heim’s PF Feature Transmission Approach, coupled with some stipulations. Let us summarize the core hypothesis of the Full Feature Transmission Approach as in (27).

(27) *Full Feature Transmission Approach*

Both phonological and semantic features are transmitted to a bound pronoun from its binder through co-indexation.

We have shown that the revised PF Feature Transmission Approach and the Full Feature Transmission Approach are defensible in the face of Spathas’s (2010) argument based on mixed gender agreement in Greek. In addition, the Full Feature

Transmission Approach offers a natural account of the computation of the ϕ -features on split pronouns. Therefore, these two analyses compose critical challenges to the semantic or presupposition account and deserve a serious evaluation. In the next section, new evidence for the presupposition account of ϕ -features on bound pronouns will be provided, suggesting that the PF Feature Transmission Approach, its revised version and the Full Feature Transmission Approach all make incorrect predictions.

4.4 New evidence for the presupposition account

In this section, I will provide arguments for the general presupposition account of the ϕ -features on bound pronouns. I will be making heavy use of collective nouns in this and the next section. Therefore, it is helpful to briefly review the syntax and semantics of collective nouns before we dive into the construction of the arguments.

4.4.1 The syntax and semantics of collective nouns

Collective nouns such as *team*, *family*, *couple* and *clergy* are normally assumed to bear the following features, where “SM-Num” represents for “phonological/morphosyntactic number feature that is interpreted at the SM interface,” and “CI-Num” signifies “semantic number feature that is interpreted at the CI interface”:

(28) *Features on collective nouns*

- a. Phonologically singular: [SM-Num: SG]
- b. Semantically ambiguous: [CI-Num: SG/PL]

The first piece of evidence for (28a) comes from the contrast between the singular and plural forms of collective nouns, illustrated in (29).

- (29) a. *team* vs. *teams*
- b. *couple* vs. *couples*

If the phonological feature on collective nouns is singular, then it is easy to explain why they have a plural morphosyntactic counterpart.

The agreement between a collective noun and a demonstrative or the determiner *a* provides another piece of evidence for (28a). (30) indicates that collective nouns with a singular phonological feature agree only with a singular demonstrative, and (31) shows the singular determiner *a* agrees only with a phonologically singular collective noun.

- (30) a. this team vs. these team*(s)
b. this couple vs. these couple*(s)

- (31) a. a team vs. a team(*s)
b. a couple vs. a couple(*s)

Demonstrative and determiner agreement rather than subject–verb agreement is employed here as a test to determine the phonological feature of collective nouns. This is because, as noticed by Corbett (1979, 2006) (his agreement hierarchy), demonstrative and determiner agreement most likely targets the phonological/morphosyntactic features of a nominal, whereas in subject–verb agreement, T is less likely to target the phonological features of the subject DP, giving rise to certain percentage of agreement against the semantic features of the subject.

The evidence for the ambiguous semantic number features of collective nouns can be found in the agreement between a singular collective nominal and a verb that is either singular or plural:⁴

- (32) a. The team **is** currently in a weaker position.
b. The team **are** very dedicated to trying to get this right.
- (33) a. The couple **is** in a part-time mentoring program.
b. The couple **aren't** planning on leaving.

Similarly, a collective nominal can agree with either a singular or a plural pronoun, and the collective nominal is interpreted either as a single entity/group or as group members, respectively:⁵

- (34) a. The basketball team won **its** only state title in 1951.
- b. The team won **their** first Illinois Intercollegiate Championship this season.
- (35) a. Under ideal circumstances, a newly married couple establishes **its** own household at once.
- b. The couple spent **their** whole time in the old city in their hotel room, watching television.

Similar analyses of collective nouns can also be found in Levin (2001); Depraetere (2003); Wong (2013) and Smith (2017).

4.4.2 Argument One

We have seen that collective nouns are ambiguous in their semantic interpretations; that is, they can be interpreted as a single group or as group members. Let us call these two interpretations the GROUP READING and the MEMBER READING. There is a correspondence between the semantic features and the interpretations of the collective nouns:

- (36) a. [CI-Num: SG]: the group reading
- b. [CI-Num: PL]: the member reading

The crucial data for our first argument is shown in (37).

⁴The examples are taken from the Corpus of Contemporary American English (COCA). *Team* does not frequently occur with a plural DP; however, it is not difficult to find collective nouns that go more frequently with plural DPs. I did a corpus study on 109 collective nouns based on the American English data in Google Ngram. That study reveals that some collective nouns such as *faculty*, *staff*, *couple* and *clergy* agree with plural DPs more than 40% of the time. Agreement with plural verbs is claimed to be more frequent for collective nouns in British English than in American English (Levin 2001; Smith 2017).

⁵The examples in (34) are taken from Google Ngram, with searches being specified for American English.

(37) Only this team_j has finished its_j/their_j projects.

Perhaps the most interesting observation here is that either the plural or the singular pronoun can be bound by the collective nominal. Importantly, the choice of the bound pronouns distinguishes the two interpretations of the collective nominal binder: The singular pronoun induces the group reading of the collective nominal binder, and the plural pronoun the member reading.

The semantic effects of the different choices of bound pronouns in (37) can be clearly brought out with kinship nouns. Kinship nouns such as *sibling* require a syntactically projected implicit argument (cf. Ke & Pires 2018; Ke et al. 2019; Ke & Pires accepted), and this argument must be an animate individual. The contrast between (38a) and (38b) indicates that the interpretations of these two sentences are different, and the reason why (38a) is not possible is that the team under discussion, as an inanimate object, naturally does not have a sibling.

- (38) a. #Only this team invites its siblings to practice.
b. Only this team invites their siblings to practice.

Therefore, it is clearly the group reading rather than the member reading that is induced by the singular bound pronoun *its*. On the other hand, the plural bound pronoun *their* can access the member reading, and this is why (38b) is good.

To further confirm the analysis, if we replace the kinship noun with a relational noun that can take an inanimate group entity as its argument, then a sentence with a singular pronoun bound by a collective nominal is predicted to be possible. This is borne out: (39a) is possible because the competitors can be the team's competitors.

- (39) a. Only this team invites its competitors to practice.
b. Only this team invites their competitors to practice.

There is now no contrast between (39a) and (39b), although different readings of the collective nominal are still induced.

This is a problem for the PF Feature Transmission Approach. The PF Feature Transmission Approach predicts that only phonological features can be pronounced on the bound pronoun. Since only the [SM-Num: SG] feature on the collective nominal binder will be transmitted to the bound pronoun, the bound pronoun can only be pronounced as “its.”

Furthermore, note that this is a problem for the revised PF Feature Transmission as well. This case is different from Spathas’s (2010) Greek example in (23), where the choice of bound pronouns does not distinguish meaning. As mentioned in Section 4.3.2, with the assumption that both [NatGender: FEM] and [GramGender: NEU] are phonological features, Spathas’s argument based on mixed gender agreement in Greek is no longer an argument against the revised PF Feature Transmission Approach. Could such a revised PF Feature Transmission Approach be applied to (37) as well? My answer is no. A reanalysis of (37) along the lines of the PF Feature Transmission Approach needs to assume that a collective nominal bears both a singular and a plural phonological number feature. However, as we have shown clearly in the previous section, collective nouns do not bear a plural phonological number feature.

Even if we assume that a collective nominal bears an ambiguous [SM-Num: SG/PL] feature, the revised PF Feature Transmission Approach (and its original version) cannot be rescued. This is because the revised PF Feature Transmission Approach assumes feature transmission at PF, and however the bound pronouns are pronounced, the resulting LF representation should remain unaffected. Consequently, it predicts that different choices of bound pronouns should not alter the meaning of the sentence. This is contrary to what we have observed with regard to the interpretations associated with different bound pronouns in (37). Instead, we have motivated the following generalization:

- (40) a. A singular bound pronoun targets the collective nominal binder’s group reading.

- b. A plural bound pronoun targets the collective nominal binder’s member reading.

I have just shown that the revised PF Feature Transmission Approach (and its original version) generates a prediction at odds with the fact that different interpretations could be induced by different bound pronouns. Nevertheless, this argument does not shed much light on the Full Feature Transmission Approach defined in (27). This is because the Full Feature Transmission Approach assumes that the semantic [CI-Num: SG/PL] feature, as well as the phonological [SM-Num: SG] feature, is transmitted to the bound pronoun, and that either the semantic feature or the phonological feature can be pronounced. When the [CI-Num: SG] is pronounced, we should get “its”; when the [CI-Num: PL] is pronounced, we get “their.” It is also reasonable to say that when a pronoun is pronounced, the corresponding meaning of the collective nominal binder is accessed. Consequently, the choice of the pronouns results in different interpretations of the collective nominal binders. In other words, our first argument suggests that the revised PF feature transmission approach (and its original version) gives rise to wrong predictions; however, it does not allow us to give an evaluation to the Full Feature Transmission Approach.

In the next section, I will present another argument against the two versions of the PF Feature Transmission Approach. In addition, I will show that even the Full Feature Transmission Approach is not tenable.

4.4.3 Argument Two

Since the revised PF Feature Transmission Approach (and its original version) and the Full Feature Transmission Approaches both assume that bound pronouns do not bear ϕ -features, both approaches predict that bound pronouns, complex and simplex reflexives should all behave the same as bound variables. This is because both bound pronouns and bound reflexives are treated as bound variables in formal semantics.

In fact, Kratzer (2009) not only extends the PF Feature Transmission Approach to reflexives, but asserts that minimal pronouns are in fact reflexives. In addition, given that the Full Feature Transmission Approach further assumes that the semantic features of a binder are also transmitted to the corresponding bound variable, this approach predicts that both the member and group readings are available with any types of bound variables. I argue that these predictions cannot be correct, as long as bound variables in Chinese are concerned.

4.4.3.1 Background on bound variables in Chinese

A question may arise here concerning the plausibility of resorting to Chinese data to evaluate the PF Feature Transmission Approach, which is mainly motivated by English data. This is a reasonable question because it could simply be that these two languages are different in the syntax and semantics of bound variables, and that the application of the PF Feature Transmission Approach to Chinese does not necessarily shed light on the question whether this approach is plausible for English.

I suspect that Heim's (2008) intent is to provide a theory of bound pronouns only for English. This is obviously not the case for Kratzer (2009), as Kratzer tests her theory against languages other than English. Nevertheless, let us see if PF Feature Transmission Approach should be applied to Chinese.

Examples in Chinese corresponding to the *only*-sentences in English, like the one in (41), suggest that the same ϕ -feature interpretability problem also arises in Chinese, and that Heim's theory should in principle be applicable to Chinese.

- (41) Zhiyou wo zuole wo/ziji/wo-ziji de jiating-zuoye.⁶
 only I did I/self/myself DE homework
 ‘Only I did my homework.’ → Others did not do their homework.

As we can see in (41), the bound pronoun (*wo* ‘I’), simplex reflexive (*ziji* ‘self’) and complex reflexive (*wo-ziji* ‘myself’) behave in the same way in the sense that

they all give rise to the same ordinary meaning and implied meaning. Therefore, it is well-motivated to apply the PF Feature Transmission Approach to solve the interpretability problem with regard to the three types of bound variables in Chinese.

4.4.3.2 Evaluating the morphosyntactic account

Let us now test the predictions of the (revised) PF Feature Transmission Approach and the Full Feature Transmission Approach concerning the three types of bound variables, namely bound pronouns, the simplex reflexive and complex reflexives. Recall that both these two approaches treat these three types of bound variables identically: They are all minimal pronouns. Therefore, both approaches predict that these three types of bound variables should be interpreted in the same way with regard to their ϕ -features.

Let us first take a look at the bound simplex reflexive. The simplex reflexive *ziji* ‘self’ in Chinese is standardly assumed to bear no inherent ϕ -features from the lexicon (e.g., Huang & Tang 1991; Cole & Sung 1994; Xue et al. 1994, but see Cole et al. 1990 for a prominent exception). (42) shows that the simplex reflexive must be interpreted as a singular entity and cannot refer to group members; thus, it can be the possessor of the relational noun ‘competitor,’ which can take an argument that is either an animate individual or a group, but cannot be that of the kinship noun ‘father,’ which requires an argument that is an animate individual, not a group.

- (42) Zhiyou zhe-ge tuandui yaoqing-le ziji-de duishou/#fuqin lai canjia
 Only this-CL team invite-ASP self-POSS competitor/father come attend
 wanyan.
 dinner
 ‘Only this team invited its competitor/#father to the dinner.’

On the other hand, the complex reflexive *tamen-ziji* ‘themselves’ in (43) allows a

⁶As shown in this example, Chinese allows a reflexive to occur as a possessor, different from English. This feature provides us a minimal pair context to test similarities and differences between a bound pronoun, simplex reflexive and complex reflexive.

member reading, similar to the bound pronoun *tamen* ‘they.’

- (43) Zhiyou zhe-ge tuandui yaoqing-le tamen-ziji-de/tamen-de
 only this-CL team invite-ASP themselves-POSS/they-POSS
 fuqin/duishou lai canjia wanyan.
 father/competitor come attend dinner
 ‘Only (the members of) this team invited their fathers/competitor(s) to the dinner.’

As a point of comparison, if we replace the plural reflexive and pronoun with their singular versions, similar to the case with the simplex reflexive, only the group reading is possible; the member reading is not accessible:

- (44) Zhiyou zhe-ge tuandui yaoqing-le ta-ziji-de/ta-de duishou/#fuqin
 only this-CL team invite-ASP itself-POSS/it-POSS competitor/father
 lai canjia wanyan.
 come attend dinner
 ‘Only this team invited its competitor(s)/#father(s) to the dinner.’

It seems plural variables with an antecedent x always have a reading denoting “the members of x ,” which is introduced by a MEMBERS function ($\text{MEMBERS}(x)$ = the members of x) (Ezra Keshet, p.c.). By contrast, singular and unmarked variables such as singular pronouns, singular complex reflexives and simplex reflexives do not have this reading. Assume that collective nominals are plural individuals in the sense of Link (1983), (45) shows the semantics of (43):

- (45) [only this team_[SG] [λ_3 invited [MEMBERS their₃] father/competitor...]]

The above results strongly indicate that the simplex reflexive *ziji* (which is generally assumed to have no inherent ϕ -features from the lexicon) is different from complex reflexives and bound pronouns. Only complex reflexives and bound pronouns bear ϕ -features and these ϕ -features must be semantically interpreted. Such results are not predicted by either the PF Feature Transmission Approach (and its revised version) or the Full Feature Transmission Approach.

Before we move on, a few questions about the simplex reflexive in Chinese must be addressed.

Firstly, it is helpful to mention the reasons why I believe *ziji* ‘self’ does not bear inherent ϕ -features from the lexicon. This is because differently from complex reflexives, it does not have the pronoun component. A widely accepted analysis of complex reflexives is that they consist of two parts: the pronoun and the *self* (Reinhart & Reuland 1993; Kayne 2002; Zwart 2002; Heinat 2006; Hicks 2009, among others). The pronoun is assumed to bear inherent ϕ -features, and therefore, it is natural to assume that the *self* part does not.

Empirically, unlike pronouns and complex reflexives, the ϕ -features of which presuppose only antecedents that bear matching ϕ -features, the simplex reflexive can be bound by an antecedent with any person and number features, consistent with the idea that simplex reflexives do not bear ϕ -features from the lexicon (or at least the ϕ -features, if there are, are underspecified):⁷

- (46) *Ziji can be bound by an antecedent with any person*

Wo/ni/ta piping-le ziji.
I/you/(s)he criticize-ASP self

‘I/you/(s)he criticized myself/yourself/himself.’

- (47) *Ziji can be bound by either a singular or a plural antecedent*

- a. Ta dui ziji hen you xinxin.
 (S)he toward self very have confidence
 ‘(S)he is very confident of herself/himself.’
- b. Tamen dui ziji hen you xinxin.
 They toward self very have confidence
 ‘They are very confident of themselves.’

⁷Since Chinese pronouns do not distinguish different genders morphologically, we will restrict our discussion to the person and number feature. It may be helpful to mention that the simplex reflexive can be bound by either an animate (regardless of its natural gender) or an inanimate antecedent, implying that it can go along with antecedents with any gender features.

I have shown that the simplex reflexive does not bear inherent ϕ -features from the lexicon. However, some researchers have made the argument that the simplex reflexive may acquire person and number features by agreeing with the T heads in its local and higher TPs, through a successive-cyclic movement of the simplex reflexive (e.g., Cole et al. 1990) to adjoin to these Ts. Therefore, a question regarding my analysis of the simplex versus complex reflexives is: If ϕ -features are acquired from a T head, then why do simplex and complex reflexives still behave differently?

The evidence for this agreement proposal comes from a blocking effect where a first or second person pronoun blocks third-person DPs from long-distance binding of *ziji* (Tang 1989), as illustrated by the following contrast:

- (48) a. *No blocking effects with intervening third person DPs*

Yuehan_i renwei **Yage**_j juede Tangmu_k dui ziji_{i/j/k} mei xinxin.
 John think **Jacob** feel Tom toward self no confidence
 ‘John_i thinks that Jacob_j feels that Tom_k is not confident in him/himself_{i/j/k}.’

- b. *Blocking effects with intervening first or second person pronouns*

Yuehan_i renwei **wo**_j/**ni**_k juede Tangmu_m dui ziji_{*i/*j/*k/m} mei
 John think **I/you** feel Tom toward self no
 xinxin.
 confidence
 ‘John_i thinks that I_j/you_k feel that Tom_m is not confident in him_{*i}/me_{*j}/
 you_{*k}/himself_m.’

However, such blocking effects are not detected with the number feature (cf. Tang 1989; Huang & Tang 1991):

- (49) Yuehan_i renwei **zhexie-yundongyuan**_j juede Tangmu_k dui ziji_{?i/j/k}
 John think **these-athlete** feel Tom toward self
 mei xinxin.
 no confidence

‘John_i thinks that these athletes_j feel that Tom_k is not confident in him_{?i}/them_j/himself_k.’

(49) suggests that previous account of blocking effects based on (i) successive-cyclic movement of the simplex reflexive and (ii) the agreement between the simplex reflexive and T heads cannot explain why blocking effects are not observed for the number feature. In addition, since the focus of my investigation here is the number features on the collective nominal binder and the simplex reflexive, the blocking effects with regard to the person feature in particular are thus irrelevant to this discussion. These blocking effects with regard to the person feature in fact receive an alternative account, as has been argued by Huang & Liu (2001) and Huang et al. (2009).

The last potential question I want to bring to the reader’s attention is whether the simplex reflexive’s syntactic status would affect our semantic analysis of it as a bound variable. It is still debated whether the simplex reflexive is a logophor, a long-distance reflexive, or a regular reflexive (see Charnavel et al. 2017 for a comprehensive review). I do not think this is an issue that should concern us here, because no matter whether the simplex reflexive is syntactically analyzed as a regular reflexive or a logophor, it is always treated as a bound variable as long as it co-varies with an antecedent.

In summary, we have shown that evidence from both English and Chinese supports the argument that ϕ -features, the number feature specifically, on bound pronouns must be semantically interpreted; otherwise, different choices of bound pronouns should not distinguish meaning and different types of bound variables should behave in the same way, contrary to fact.

4.4.4 Argument Three

An important concern regarding Argument One and Two is that the evidence shows only that the number feature on bound variables must be semantically interpreted. However, previous formulations of feature transmission allow the number feature to be base-generated and be semantically interpretable. If a morphosyntactic approach allows the number feature on bound variables to be optionally semanti-

cally interpretable, the two arguments above may miss the heart of the problem. To illustrate the concern, let us first consider Heim’s (2008) PF Feature Transmission Approach. For Heim (2008), PF feature transmission is obligatory, and pronouns are optionally generated with specified (semantically interpretable) ϕ -features (pp. 50–52). However, the restriction is: If the ϕ -features are base-generated on bound pronouns, then these features must not be in conflict with the ϕ -features transmitted from the pronoun binder. Assuming that the number feature is base-generated, are Argument One and Two still counterarguments to Heim’s (2008) PF Feature Transmission Approach?

I think the answer is positive. Let us take (38), repeated below as (50), as an example.

- (50) a. #Only this team invites its siblings to practice.
- b. Only this team invites their siblings to practice.

If the plural number feature on the bound pronoun in (50b) is base-generated, (50b) is still a problem for Heim’s (2008) PF Feature Transmission Approach. This is because PF feature transmission is obligatory if *their* is a bound pronoun, which implies that the phonological feature on *their*’s binder must be transmitted to *their*. As we have mentioned, the phonological number feature on the collective nominal binder is singular. The base-generated number feature, on the other hand, must be plural, as the bound pronoun is semantically interpreted as plural. This leads to a conflict between the base-generated number feature and the number feature transmitted from the binder, which is not allowed in Heim’s system. Therefore, no matter the number feature on the bound pronoun is assumed to be base-generated or not, Heim (2008) incorrectly predicts (50b) is ungrammatical.

What if Kratzer’s (2009) version of Feature Transmission under Binding is adopted? Kratzer (2009) proposes the following mechanism for feature transmission from a v head, which hosts the λ binder of the bound pronoun in (51) and bears base-generated

ϕ -features according to Kratzer (2009), to a bound pronoun in the internal argument/object position:

(51) *Feature Transmission under Binding* (Kratzer 2009:216)

The ϕ -feature set of a locally bound pronoun unifies with the ϕ -feature set of the head that hosts its binder.

Kratzer's Feature Transmission under Binding works together with another mechanism called Predication:

(52) *Predication* (Kratzer 2009:196)

When a DP occupies the specifier position of a head that carries a λ -operator, their ϕ -feature sets unify.

Let us take (50b) again as an example, with its structural representation in (53).

(53) [_{TP} Only this team [_{vP} <this team> *v_{invites}* their siblings to practice]]

Feature Transmission under Binding requires that the ϕ -features on the bound pronoun *their* to be unified with the ϕ -features on the local *v_{invites}* head, which carries a λ operator that in turn binds *their*. Then Predication states that the ϕ -features on *v_{invites}* and those on SpecvP, i.e., *this team* must unify. Taken together, the above two mechanisms require the binder and the bound pronoun to share identical ϕ -features inside the vP. The critical prediction of Kratzer's theory is summarized below in (54).

(54) *A prediction of Feature Transmission under Binding*

SpecvP shares its ϕ -features with those on *v* and the pronoun bound by *v*.

If the ϕ -features on the bound pronoun are not base-generated and thus are phonologically features, then Kratzer (2009) predicts that the ϕ -features on the bound pronoun must be identical to the phonological ϕ -features on the collective nominal binder, which is singular, implying that only a singular bound pronoun is possible in (53), contrary to fact.

Crucially, it must be noted that Kratzer (2009) assumes, following Rullmann (2004), the number feature is exempt from Feature Transmission under Binding. That is, the “ ϕ -features” in Feature Transmission under Binding (51) applies only to non-number features, such as gender and person features. It is unclear how Kratzer (2009) can derive the “ignorance” of the number feature in the implied meaning/focus value of the *only*-sentences that we have discussed extensively, including the example in (53). If the number feature is base-generated on the bound pronoun in *only*-sentences, we run into exactly the problem as pointed out clearly by Heim (2008). We do not have a solution to the interpretability problem associated with *only*-sentences such as (53) if we assume that feature transmission is optional. Since Argument One and Two are both built upon observations about *only*-sentences, they do impose a challenge to Kratzer’s analysis.

Now let us ignore all possible problems with regard to the splitting of ϕ -features into two categories: the number features versus other ϕ -features such as gender and person. Argument One and Two cannot then be generalized to the gender and person feature transmission. Below I present another argument from Standard Arabic and Russian to support the idea that the (semantic) gender feature is no different from the number feature in the sense that it must be semantically interpretable under binding.⁸

The following is an example with which we can test feature transmission of both the number and gender features.⁹ At first glance, it may be argued that the Arabic example (55) is consistent with Kratzer’s Feature Transmission under Binding because the ϕ -features on the *v*P-internal subject ‘these girls,’ *v_{invited}* and the bound pronoun ‘their’ are all 3RD.PL.FEM.

(55) *Standard Arabic*

⁸I should thank my consultant Fahad Alrashid for the Arabic examples, and Aliaksei Akimenka for the Russian examples, presented in this subsection.

⁹I thank Yasutada Sudo for his suggestion to look for this type of examples.

faqaT ha?ula? l-banaat_j daʃa-u
 only these.3RD.PL.FEM the-girls.PL.FEM invited.3RD.PL.FEM
 ?aTiba?-hon_j li-l-tadreeb
 doctors.PL.MAS-their.3RD.PL.FEM to-the-practice

‘Only these girls invited their doctors to the practice.’

Recall that according to Kratzer (2009), the sharing of the ϕ -features between the binder and the bound pronoun is a result of two feature unifications: the unification between the ϕ -features of ‘these girls’ and $v_{invited}$, and that of the ϕ -features between $v_{invited}$ and ‘their.’ In addition, the T head agrees with the higher copy of ‘these girls,’ as shown by the agreeing features present on the verb.

However, the following sentences with collective nominal binders seem to be at odds with Kratzer’s (2009) prediction in (54) based on her Feature Transmission under Binding¹⁰:

- (56) a. faqaT hada l-fareeq_j daʃ-a
 only this.3RD.SG.MAS the-team.SG.MAS invited.3RD.SG.MAS
 munaafisi/abaa?-h_j li-l-tadreeb
 competitors/fathers-its.3RD.SG.MAS to-the-practice
 ‘Only this team_j invited its_j competitors/#fathers to the practice.’
- b. faqaT hada l-fareeq_j daʃ-u
 only this.3RD.SG.MAS the-team.SG.MAS invited.3RD.PL.MAS
 munaafisi/abaa?-him_j li-l-tadreeb
 competitors/fathers-their.3RD.PL.MAS to-the-practice
 ‘Only this team_j invited their_j competitors/fathers to the practice.’

The problem is, at least at surface, when the collective nominal binds a singular pronoun, ϕ -feature unification is not found between the v P-internal subject, v , and the bound pronoun. This is of course a familiar problem that we have observed with the number feature in English and Chinese.

¹⁰According to my consultant, when the collective nominal subjects in (56) and similar examples from Standard Arabic in this subsection are interpreted as a single group, the singular agreement number on the verb is the only option; however, when the collective nominal is interpreted as group members, the plural agreement number on the verb is most natural.

What if the number feature is exceptional and must be handled in a way different from other ϕ -features, as assumed by Kratzer (2009)? According to Kratzer (2009), the gender feature must be shared between the v P-internal subject, v and the bound pronoun. A broader survey of relevant data shows that this is not the case. There is in fact no real dependency between the phonological/morphosyntactic ϕ -features of the collective nominal binder and those of the bound pronoun. Consider the contrast between (57a) and (57b).

- (57) a. faqaT hadih l-ʕaʔilah_j daʕa-t
 only this.**3RD.SG.FEM** the-family.SG.FEM invited.**3RD.SG.FEM**
 munaafis-ha_j li-l-tadreeb
 competitor-its.**3RD.SG.FEM** to-the-practice
 ‘Only this family_j invited its_j competitor to the practice.’
- b. faqaT hadih l-ʕaʔilah_j daʕa-u
 only this.**3RD.SG.FEM** the-family.SG.FEM invited.**3RD.PL.FEM**
 munaafisi-him_j li-l-tadreeb
 competitor-their.**3RD.PL.MAS** to-the-practice
 ‘Only this family_j invited their_j competitor to the practice.’

(57a) is included only for comparison purposes, as it does not give us any new information beyond what previous Arabic examples have told us. What is interesting to us is (57b), where the bound pronoun is plural. Crucially, the bound pronoun in (57b) bears a masculine gender feature, whereas the collective nominal binder bears a feminine gender feature. Where does the masculine gender feature on the bound pronoun come from?

It seems the masculine gender feature on the bound pronoun in (57b) is a default gender feature. The default gender feature is used only when the bound pronoun refers to the members of the collective nominal binder (not to the collective nominal as a singular entity), given that the gender of the members is unknown or most likely mixed.¹¹ In this case, the bound pronoun cannot bear the same morphological feature with the collective nominal binder, as that is associated with the group reading of

the collective nominal. In light of this interpretation, the identification of the gender feature between the collective nominal binder and the bound pronoun in (56b) is simply a coincidence; that is, the masculine gender on the bound pronoun in (56b) is a default gender, not a gender that is inherited from the collective nominal binder.

Given that the masculine feature is a default gender feature, evidence for the disassociation between the collective nominal binder and the bound pronoun must come from cases where the collective nominal binder bears a feminine gender feature. The contrast between (58a) and (58b) illustrates the point, similar to the contrast between (57a) and (57b):

- (58) a. faqaT hadih l-Hokomah_j daʃa-t
 only this.**3RD.SG.FEM** the-government.SG.FEM invited.**3RD.SG.FEM**
 xaSm-ha_j
 enemy-its.**3RD.SG.FEM**
 ‘Only this government_j invited its_j enemy.’
- b. faqaT hadih l-Hokomah_j daʃa-u
 only this.**3RD.SG.FEM** the-government.SG.FEM invited.**3RD.PL.FEM**
 xuSuum-hom_j
 enemy-their.**3RD.PL.MAS**
 ‘Only this government_j invited their_j enemy.’

Again, in (58b) we see that both the number and gender features on the bound pronouns are different from those on the collective nominal binders.

As previously mentioned, the default masculine gender is used when the bound pronouns refer to the members of the collective nominal binders (e.g., in (58b)). This is probably because the natural gender of the members are unknown or most likely mixed. In cases where the natural gender of the members are known and not mixed, the gender feature on bound pronouns should be consistent with the natural gender of the members denoted by the collective nominal binders. (59a) is the same as (58b),

¹¹If the bound pronoun is singular, it must share the same gender feature with its binder. This is understandable because the gender should be in accordance with the formal features on its binder based on its own reference.

where default gender is used, and this is because the gender of the referent of the bound pronoun, i.e., the members of the school, is unknown or most likely mixed in normal situations. When it is known that the bound pronoun refers to females only, as in (59b), its gender feature must be feminine; by contrast, when the bound pronoun refers to males only, as in (59c), its gender feature must be masculine.

- (59) a. faqaT hadih l-madrasah_j daʕa-u
 only this.**3RD.SG.FEM** the-school.SG.FEM invited.**3RD.PL.FEM**
 munaafisi-him_j
 competitors-their.**3RD.PL.MAS**
 ‘Only this school_j invited their competitors_j.’
- b. min bayn madaaris l-banaat, faqaT hadih
 among schools the-female only this.**3RD.SG.FEM**
 l-madrasah_j daʕa-u
 the-school.SG.FEM invited.**3RD.PL.FEM**
 munaafisi-hin_j
 competitors-their.**3RD.PL.FEM**
 ‘Among the all-girls schools, only this school_j invited their_j competitors.’
- c. min bayn madaaris l-ʔawlaad, faqaT hadih
 among schools the-male only this.**3RD.SG.FEM**
 l-madrasah_j daʕa-u
 the-school.SG.FEM invited.**3RD.PL.FEM**
 munaafisi-him_j
 competitors-their.**3RD.PL.MAS**
 ‘Among the all-boys schools, only this school_j invited their_j competitors.’

The contrast between (59b) and (59c) strongly supports the argument that the gender feature on the bound pronoun tracks the gender feature associated with the reference of the bound pronoun, rather than the morphosyntactic features of its binder. This contrast leads to the conclusion that the gender feature on bound pronouns in Standard Arabic must be semantically interpretable, instead of being transmitted from the binders.

Similar sentences from Russian reveal, with direct inflectional evidence, that the phonological number and gender features on possessive bound variables also do not

agree with the phonological number and gender features on the *v* heads or the collective nominal binders. Instead, such phonological number and gender features agree with the syntactic complement of the possessor, i.e., the possessee. I list below sentences from reflexive possessor rather than pronominal possessors, because Russian allows bound reflexive possessors but not 3rd person bound pronominal possessors. Sentence (60) below shows that the number feature of the bound reflexive does not match the number feature on its collective nominal binder, the D head of the collective nominal binder, more accurately.

(60) *Russian*

Tol'ko ehta komanda_j priglasila
 only this.SG.FEM.NOM team.SG.FEM.NOM invited.3RD.SG.FEM
 svoikh_j sopernikov/ottsov na trenirovki
 REFL.3RD.PL competitors/fathers.PL.MAS.ACC to the-practice

Meaning One: 'Only this team_j invited its_j competitors/#fathers to the practice.'

Meaning Two: 'Only this team_j invited their_j competitors/fathers to the practice.'

It is important to note that the number feature on the bound reflexive in (60) does not interact with the semantic interpretation of the bound reflexive. As shown by the English translation, the bound reflexive in (60) is ambiguous between a singular and a plural interpretation. If it is interpreted as singular, then it refers to the team as a single group, enabling it to go with *competitors* but not *fathers* (Meaning One). If the bound reflexive is interpreted as plural, it refers to the members of the team, thus it can go with either *competitors* or *fathers* (Meaning Two). The number feature is thus a phonological feature, not a semantic feature. Therefore, Russian, different from Chinese, allows a disassociation between the phonological and semantic ϕ -features on bound reflexives.¹²

¹²This supports our feature specification on pronouns in Chapter VI

The phonological number features on the bound pronoun are in fact a result of an agreement relation between the reflexive possessor and the possessive bound pronoun's syntactic complement, i.e., the possessee 'competitors/fathers.' This is supported by the following sentence where the bound variables are singular in number but can take the kinship noun 'father' as its complement, as long as the team members have the same father.

- (61) Tol'ko ehta komanda priglasila
 only this.SG.FEM.NOM team.SG.FEM.NOM invited.3RD.SG.FEM
 svoego sopernika/ottsa.
 REFL.3RD.SG.MAS.ACC competitor/father.SG.MAS.ACC
 Meaning One: 'Only this team_j invited its_j competitor to the practice.'
 Meaning Two: 'Only this team_j invited their_j competitor/father to the practice.'

Recall that if the bound reflexive must refer to the team as a group, then it cannot take a kinship noun complement, because a team does not have a father. This implies the semantic number feature is different from the phonological number feature on the bound reflexive in Russian. The phonological number feature on the bound variables are in fact copied from its complement rather than from the local *v* head or the collective nominal binder.

Another observation regarding the above sentences is that the natural gender of the team members can also be different from the morphosyntactic gender feature on the collective nominal. That is, 'this team' in (60) can refer to a team with all male members although the morphosyntactic gender is feminine. In such a situation, the bound reflexive can still refer to these male members, although the bound reflexive is marked as feminine. This observation indicates that the gender feature that surfaces up on the collective nominal binder is a phonological feature, not a semantic feature. To further support the analysis, it would be great to see if the bound reflexive can have the same phonological feature as its collective nominal binder, but be interpreted

with a different semantic gender.

Unfortunately, the bound reflexive in (60) is plural, and in Russian, plural bound reflexives are underspecified in gender. We therefore need singular bound reflexives to investigate potential disassociation between the phonological features on the bound reflexives and their binders. (61) provides such an example, as the bound reflexive bears a masculine gender feature whereas the binder bears a feminine gender feature.

Where does the masculine feature on the bound reflexive in (61) come from? (62) shows that, like the phonological number feature, it is also a result of the agreement between the possessor and its NP complement.

- (62) a. Tol'ko ehta komanda priglasila
 only this.**SG.FEM.NOM** team.**SG.FEM.NOM** invited.**3RD.SG.FEM**
 svoego ?druga.
 REFL.**3RD.SG.MAS.ACC** friend(male).**SG.MAS.ACC**
 'Only this team_j invited its_j (male) friend.'
- b. Tol'ko ehta komanda priglasila
 only this.**SG.FEM.NOM** team.**SG.FEM.NOM** invited.**3RD.SG.FEM**
 svoju ?podrugu.
 REFL.**3RD.SG.FEM.ACC** friend(female).**SG.FEM.ACC**
 'Only this team_j invited its_j (female) friend.'

The contrast is between the masculine- and feminine-marked words for 'friend.' If the masculine version is used, i.e., *druga*, the corresponding possessive reflexive must bear a masculine gender feature (62a). However, if the feminine version is used instead, i.e., *podrugu*, the corresponding possessive reflexive must bear a feminine feature (62b). This contrast suggests that the phonological gender feature on the bound reflexives has nothing to do with the collective nominal binders. Instead, they are determined syntactically by a syntactic agreement between the bound reflexive possessor and the possessee. In other words, the phonological gender feature on the bound variables is not copied from the collective nominal binder, contrary to the prediction of the morphosyntactic account.

What if bound variables are used in θ -role positions, so that the features on bound variables cannot be copied from another nominal through agreement? (63) shows that a bound variable in a θ -role position (boldfaced) bears ϕ -features consistent with their binder/referent.

- (63) Tol'ko ehta shkola priglasila
 only this.SG.FEM.NOM school.SG.FEM.NOM invited.3RD.SG.FEM
 sopernikov kotorykh ona
 competitors.PL.ACC which.PL.ACC **it.3RD.SG.FEM.NOM**
 znala
 knew.3RD.SG.FEM
 'Only this school invited the competitors that it knew.'

The examples in (64) further illustrate that the number feature on the emphatic complex reflexives is different from that on the binder, suggesting that the semantic number feature of the bound variables in Russian are not copied from the binders; instead, the number feature on the boldfaced bound variables in (64) is semantically interpretable, as it must be consistent with the number of the bound variables' referents.¹³

- (64) a. Tol'ko ehta para kotoraya
 only this.SG.FEM.NOM couple.SG.FEM.NOM which.SG.FEM.NOM
 priglasila svoikh družej vinila
 invited.3RD.SG.FEM REFL.PL.ACC friends.PL.ACC blamed.3RD.SG.FEM
 ?**samikh-sebya** za proishestvie
PN.PL.ACC-REFL.ACC for accident.SG.NEUT.ACC
 'Only this couple that invited their friends blamed themselves for the accident.'
- b. Tol'ko ehtot klass kotoryj
 Only this.SG.MAS.NOM class.SG.MAS.NOM which.SG.MAS.NOM
 priglasil svoikh družej vinil
 invited.3RD.SG.MAS REFL.PL.ACC friends.PL.ACC blamed.3RD.SG.MAS

¹³Individual variations among collective nouns exist: some collective nouns can readily bind a plural variable but some others are less plausible binders. For instance, if collective nouns such as 'army, team, committee' serve as binders of the plural variable in (64), the sentence will be degraded; whereas 'couple, family, group, class' are more plausible binders.

?samikh-sebya za proishestvie
 PN.PL.ACC-REFL.ACC for accident.SG.NEUT.ACC
 ‘Only this class that invited their friends blamed themselves for the accident.’

Again, note that plural nominals in Russian are not specified in gender; therefore, we cannot test whether the gender feature on the plural bound variables could be different from that on the binders in (64).

In sum, in this subsection I examined the relation between the number and gender features on bound variables and those on their collective nominal binders in Standard Arabic and Russian, two languages that exhibit morphosyntactic features on bound variables. For Arabic, evidence indicates that both the number and gender features on bound pronouns are determined by the number and gender features of the bound pronouns’ referent. Note that this cannot be a result of feature transmission from the binders to the bound pronouns. Feature transmission accidentally works for cases where the binder’s morphosyntactic/phonological features are consistent with the semantic features. As in this case, matching either type of features will lead to the same ϕ -feature specification on the pronouns. The really revealing case is when the phonological features and semantic features on the binder are dissociated. Collective nominal binders provide exactly such a case. With collective nominal binders, Arabic informs us that the number and gender features on bound pronouns are all together determined by the pronouns’ semantic reference, a characteristic that is only consistent with the semantic approach: The number and gender features are semantically interpreted as presuppositions, and therefore if a bound pronoun refers to a certain referent, it must bear the same semantically interpretable ϕ -features as those associated with the referent.

Russian data further advance our understanding regarding a way that the phonological/morphosyntactic features on bound variables can be specified in a language with rich inflection. Different from Chinese, English and Standard Arabic, the phono-

logical number and gender features on bound reflexives in Russian can be disassociated from the semantic number and gender features. We found that the phonological number and gender features on the possessive bound reflexives in Russian is determined by its complement, as they are involved in a formal agreement relation. By contrast, the semantic number and gender features are not determined by the phonological features on the bound possessive reflexives or by those on the possessees.

Taken together, the Arabic and Russian data provide us a strong counter-argument against the Feature Transmission under Binding approach, as the semantic interpretable ϕ -features on bound variables are determined by their reference, which is completely in the semantic realm. Similarly, the phonological number and gender on bound reflexives in Russian are also not determined by their binders; instead, they are determined by a syntactic agreement relation with the complement of the bound variable possessor. In sum, both the semantic and phonological ϕ -features on bound variables are not copied from the binders, contrary to the prediction of all versions of the morphosyntactic account.

However, it is important to note that the above arguments support the idea that the ϕ -features on bound pronouns must be interpreted in the ordinary meaning of *only*-sentences, illustrated by (65a). But my arguments assert nothing about how the bound pronoun in the implied meaning should be interpreted, and thus this remains a question (65b).

- (65) a. The ϕ -features on *its* must be interpreted in the ordinary meaning:
 only this team did **its** projects
- b. Are the ϕ -features on *their* in the implied meaning not interpreted?
 the other teams did not do **their** projects

We now turn back to the original examples that motivate the PF Feature Transmission Approach, and will try to provide an explanation of why the ϕ -features are

seemingly not semantically interpreted under some particular contexts, namely, the focus contexts.

4.5 Toward an explanation

Why are ϕ -features on bound variables not interpreted in the implied meaning of *only*-sentences? Again, an example of *only*-sentences is repeated below in (66). A general answer to this question is that ϕ -features are not interpreted at the focus value associated with *only* (Spathas 2010; Jacobson 2012; Sauerland 2013).

- (66) only I did **my** homework \rightarrow others did not do **their** homework

The same interpretability issue emerges with other focus contexts such as *even*-sentences and *also*-sentences, given a context where only the speaker and Mary are salient:

- (67) a. even I_F have finished my homework \rightarrow
 (It is expected that) Mary has finished her homework
 b. I_F have also finished my homework \rightarrow
 Mary has finished her homework

Therefore, it is likely that the “invisibility” of ϕ -features on bound variables has an important connection with the semantics of focus. In order to have a proper understanding of the nature of the interpretability issue, it is necessary to briefly review Rooth’s (1992b) theory of focus, which is revised from Rooth (1985).

4.5.1 Rooth’s theory of focus

Rooth (1992b) proposes an influential theory of focus. According to this theory, each expression has two values: an ordinary semantic value and a focus value. The ordinary semantic value is the denotation of a given expression, and the focus value is

the expression's secondary value that takes focuses and the alternative sets induced by focuses into account. For sentences without focuses, its ordinary semantic value and focus value are identical. When a focus is involved in a sentence, the focus value is computed by replacing the focus-marked element with the alternative set induced by the focus. The ordinary value is indicated by $\llbracket \cdot \rrbracket^o$, and the focus value $\llbracket \cdot \rrbracket^f$, with the former a member of the latter. It is $\llbracket \cdot \rrbracket^f$ that interprets focus-marked elements as alternatives. For instance, given a sentence (68) where *Mike* in (68) is focus-marked, its ordinary and focus value are shown as in (69):

(68) Mike_F won the election.

(69) a. *Ordinary value*

win(Mike, the election)

b. *Focus value*

$\{\text{win}(x, \text{the election}) \mid x \in D_e\}$

Rooth (1985, 1992b) points out that pragmatic factors must also contribute to the interpretation of a sentence with a focus. He thus imposes a contextually defined set C . C is defined through a pragmatic process fixing the quantifier domain associated with *only*. To combine the focus value and the contextual restriction, Rooth (1992b) invents an operator “ \sim .” This operator introduces the presupposition that the contextually provided set C is a subset or a member of the focus value of \sim 's sister constituent. Again, taking (68) as an example; if only Mike and Sarah are salient in the context, the C set and the corresponding relation between C and the focus value of the sentence (68) are as shown below:

(70) a. $[_S [_S \text{Mike}_F \text{win the election}] \sim C]$

b. $C \subseteq \llbracket \text{Mike}_F \text{win the election} \rrbracket^f = \{\text{win}(x, \text{the election}) \mid x \in D_e\}$

c. $C = \{\text{win}(\text{Mike}, \text{the election}), \text{win}(\text{Sarah}, \text{the election})\}$

(70a) shows that a \sim operator is inserted to connect the contextually determined set C with the interpretation of the sentence with a focus. In (70b), C is restricted as a subset of the focus value of the sentence. (70c) specifies the value of C given a particular context.

The ordinary value, focus value and C are input to the semantics of *only*:

$$(71) \quad \llbracket \text{only } \alpha \rrbracket = 1 \iff \forall \rho \in C \ (\rho \neq \llbracket \alpha \rrbracket^o \rightarrow \rho \text{ is false})$$

$$\text{presupposition}^{14}: \llbracket \alpha \rrbracket^o \in C \subseteq \llbracket \alpha \rrbracket^f$$

Let us take the *only*-sentence in (66), namely, *only I_F did my homework*, as an example. The ordinary value and focus value can be found in (72), and the set C in (73), given that the salient individuals in the context are {I, Mike}.

$$(72) \quad \text{a. } \textit{Ordinary value}$$

$$\text{do}(\text{I}_2, \text{my}_2 \text{ homework})$$

$$\text{b. } \textit{Focus value}$$

$$\{\text{do}(x_2, \text{my}_2 \text{ homework}) \mid x \in D_e\}$$

$$(73) \quad C = \{\text{I}_2 \text{ did my}_2 \text{ homework, Mike}_3 \text{ did my}_2 \text{ homework}\}$$

If we fit these values in the semantics of *only*, which requires only the ordinary value to be true, and any proposition in C that is not equal to the ordinary value must be false, then we get the following truth value for the sentence:

$$(74) \quad \text{a. } \text{I}_2 \text{ did my}_2 \text{ homework} = 1 \ \&$$

$$\text{b. } \text{Mike}_3 \text{ did my}_2 \text{ homework} = 0$$

We see that we have spelled out the interpretability problem of the ϕ -features on bound pronouns under the semantics of focus, assuming that only the binder is focus-marked: (74b) is not what we intend. Instead, we want the ϕ -features on the bound

¹⁴The ordinary value, of course, must be a member of the focus value, although this is not always formally stated in the literature.

pronoun to match with those on its binder. Below, I will review previous solutions to the interpretability problem related to bound pronouns in the light of Rooth’s theory of focus.

4.5.2 Previous solutions

Spathas (2010) argues for a solution according to which the presupposition associated with ϕ -features does not enter into the computation of the focus value, following Heim (2005) (who credits Benjamin Spector for this idea). This is to say that for sentences such as *only I_F did my homework*, the focus value will ignore the ϕ -features on the bound pronouns (75b); however, for the calculation of the ordinary value, the presupposition meaning of ϕ -features are preserved (75a):

- (75) a. *Ordinary value*
 $\text{do}(I_2, [1\text{ST } [\text{SG } \text{pro}_2]]\text{'s homework})$
 b. *Focus value*
 $\{\text{do}(x_2, x_2\text{'s homework}) \mid x \in D_e\}$

A similar analysis is proposed by Jacobson (2012) and Sauerland (2013). For example, Jacobson (2012) also suggests that ϕ -features are not interpreted in the focus value, but she takes a step further, by suggesting that the focus value is computed in a compositional way. That is, the focus value of an expression is the focus value of its parts. As a result, when a sentence includes a bound pronoun, the focus value of that sentence will be partially computed based on the focus value of the bound pronoun. Jacobson assumes that the focus value of a bound pronoun is different from its ordinary value: The ϕ -features are not interpreted in the focus value. Taking again the sentence *only I_F did my homework* as an example; the ordinary and focus values of the bound pronoun *my* are shown below in (76):

- (76) a. *Ordinary value*

$$\llbracket \text{my homework} \rrbracket^o = \lambda x \in \text{1ST.SG. } x\text{'s homework}$$
b. *Focus value*

$$\llbracket \text{my homework} \rrbracket^f = \{ \lambda x. x\text{'s homework} \}$$

(76a) indicates that the domain of the bound pronoun is constrained by the pre-supposition induced by the ϕ -features on the bound pronoun; however, such domain restriction is not operative in the focus value (76b).

Although technically, solutions along the lines that ϕ -features are ignored in the computation of focus value may be valid, they do not provide an explanation as to why ϕ -features must be ignored in the computation of the focus value but not in the ordinary value of *only*-sentences.

4.5.3 Current solution

If we assume that the bound pronoun in *only*-sentences is also focus-marked, like its binder, then we can explain why ϕ -features must be interpreted for the ordinary value but not for the focus value (Ezra Keshet p.c.). The leading idea is that when a bound pronoun is focus-marked, exemplified in (77), the ϕ -features on the bound pronouns can be ignored in the focus value. Of course, we would need to allow a focus-marked pronoun to be optionally unaccented.

- (77) Only I_F did my_F homework.

However, more work needs to be done before we actually solve the interpretability problem. Let us see how exactly this idea could be worked out. If we assume a focus on the bound pronoun, according to Rooth's (1992b) theory of focus which I briefly introduced in Section 4.5.1, the ordinary value of that bound pronoun will not be affected by the focus; however, the focus value of that bound pronoun will be an alternative set. This is exactly what we want. That is, in the ordinary value,

ϕ -features on bound pronouns need to be interpreted, but in the focus value the ϕ -features are not interpreted because we compute a set of alternatives to the pronoun. But it may be not obvious how the ϕ -features are not interpreted when an alternative set of the bound pronoun is computed.

In order to address this question, we must have a better understanding of the interaction between the interpretability of ϕ -features and the focus value. Let us take a look at the alternative set induced by the focus on the proper name *John* in (78), which will give us a clear idea about such an interaction:

(78) Only John₂ did his₂ homework.

(79) a. $\llbracket \text{John}_F \rrbracket^f = \{x \mid x \in D_e\}$

b. $C = \{\text{John, Mary, Bill, Tom, Beth, } \dots\}$

(79a) indicates that the focus value of the focus-marked proper name *John*, and (79b) an alternative set C that is given by the context. If we assume that proper names bear ϕ -features as well as an index, for example, *John* bears 3RD.SG features (let us ignore the stereotype gender), then what happens in the alternative set is that both the person and number features on *John* are not operative as presuppositions or domain restrictions. This is because the alternative set may include a name with any person or any number features, as long as its semantic type is $\langle e \rangle$.

Returning to the alternative set induced by a focus-marked bound pronoun, in order to compute its alternative set, an important question we need to address is what the semantic type a bound pronoun is. If we assume that a bound pronoun is also of type $\langle e \rangle$, following Sauerland (2000) and Mayr (2012), among many others, then the alternative set will be identical to that of a proper name. That is, the alternative set will be a list of type $\langle e \rangle$ individuals. What then can distinguish the focus value of a bound pronoun from that of other constituents with an $\langle e \rangle$ type? Let us again use proper names for comparison's sake. In order to answer this question,

we need to spell out the internal structure of a proper name and a bound pronoun. (80) indicates that the proper name *John* bears the same ϕ -features as the bound pronoun co-indexed with it (ignoring the gender feature).

- (80) a. $John_5$: [3RD [SG [John₅]]]
 b. he_5 : [3RD [SG [pro₅]]]

To put it another way, the proper name *John*, like a pronoun, can also be interpreted by its index, coupled with presuppositions triggered by its ϕ -features (Elbourne 2005). So the interpretation of *John* under an assignment function g should be as in (81).

- (81) $\llbracket John_5 \rrbracket^g = \lambda x. x$ is 3RD-SG. x 's name is "John"

The proper name *John* in (81) is in fact a predicate, indicating an individual whose name is "John". Comparing (81) to the interpretation of the pronoun *him* in (82), we find that the only difference between the proper name *John* and a bound pronoun co-indexed with *John* is on the predication: The proper name requires that the referent its index refers to has a certain name, whereas a bound pronoun does not have this requirement.

- (82) $\llbracket he_5 \rrbracket^g = g(5)$ if $g(5)$ is 3RD-SG, otherwise undefined.

This is probably because, as proposed by Elbourne (2005) and Matushansky (2008) (see also Sauerland 2000, 2008), proper names have a phonologically null determiner head and the name is just the predicate, whereas bound pronouns are at the D head with a null NP so that no predicate is available for interpretation, as illustrated by (83a) and (83b), respectively.

- (83) a. $[\emptyset_D [John]]$
 b. $[he [\emptyset_{NP}]]$

Based on the semantic and syntactic differences between a proper name and a pronoun, we are ready to turn back to the question regarding the focus value of

bound pronouns, as well as the question whether the focus value of bound pronouns is different from that of proper names. As indicated by the structures in (83), the focus on the proper name and the bound pronoun should in fact be on different syntactic positions, given the assumption that null elements cannot bear focus. That is, the focus on the proper name is in fact on the NP part, whereas that on the bound pronoun is on the D head. The focus on the NP part will replace the name *John* with different names, and thus may alter the gender, person and number features associated with the names, as we have seen in (79). However, for the bound pronoun the focus is on the D head; the predicate NP is not affected. This means that the alternatives to the bound pronoun will be a set of variables, since it is the null NP that determines the variable status of pronouns (Elbourne 2005). On the other hand, since the pronoun is focused, alternatives to the pronouns should be induced. What should be the alternatives? Given that the pronoun includes a set of ϕ -features, the alternative set will be a set of alternative ϕ -features (the gender feature is not represented here but can be easily added):

$$(84) \quad \llbracket [\text{he}_5]_F \rrbracket^f = \{3\text{RD} - \text{SG}, 2\text{RD} - \text{SG}, 1\text{RD} - \text{SG}, \\ 3\text{RD} - \text{PL}, 2\text{RD} - \text{PL}, 1\text{RD} - \text{PL}\}$$

Correspondingly, the focus value of the whole bound pronoun construction will be a set of pronominal variables:

$$(85) \quad \llbracket \llbracket [\text{he}_5]_F \emptyset_{NP} \rrbracket \rrbracket^f = \{\text{he}, \text{you}_{\text{SG}}, \text{I}, \text{they}, \text{you}_{\text{PL}}, \text{we}\}$$

Note that each of the pronouns in the alternative set in (85) is supposed to be assigned an index, and therefore, there may be multiple occurrences of *he* as long as they have different indices.

In sum, we have now derived the focus value of a focus-marked bound pronoun based on the syntactic structure of bound pronouns (and pronouns more generally). The focus value and the contextually provided C of the *only*-sentence in question are

presented in (86).

- (86) a. $\llbracket \text{John}_F \text{ did his}_F \text{ homework} \rrbracket^f = \{ \llbracket \text{John}_F \rrbracket^f \text{ did } \llbracket \text{his}_F \rrbracket^f \text{ homework} \}$
 b. $C = \{ \text{John did his homework,}$
 Mary did her homework,
 Bill did his homework,
 Beth did her homework, $\dots \}$

In principle, the C set may also include propositions with pronouns that are not bound, such as those shown in (87).

- (87) $\{ \text{John did her homework,}$
 Mary did his homework,
 Bill did her homework,
 Beth did his homework, $\dots \}$

These propositions can be filtered out by presupposition failure, due to the ϕ -feature mismatch between the binders and the pronouns (Ezra Keshet, p.c.), as long as a binding relation is imposed.

4.5.4 Addressing remaining questions

A remaining problem is that nothing forces the pronouns in C to be bound in (86b). This is because such a requirement is not imposed on the focus value in (86a). In other words, undesired alternatives such as those in (88) are in principle possible. I think this problem could be addressed, and a few potential solutions are suggested below.

- (88) $\{ \text{John } \lambda_1 \text{ did her}_2 \text{ homework, Mary } \lambda_2 \text{ did her}_3 \text{ homework} \}$

Recall that I adopt the syntactic approach to binding; that is, I assume that binding relations are established in syntax, and semantics will make use of the binding

relations.¹⁵ Thus, binding in semantics is computed based on syntactic binding relations (see Chapter III, also see Reuland (2011, ch. 3)). In syntax, the antecedent and the bound variable bear the same index variable as they stand in an agreement relation. At the CI/LF interface, these index variables are assigned specific index values, referring to certain referents in the assignment model. If this framework is adopted, then it can be suggested that the binding relation is established before LF (i.e., in the syntax) and the binding relation must be also held in the focus value. If in syntax a pronoun is bound by an antecedent, in the computation of the ordinary value, the pronoun must still be bound by the antecedent; in the focus value, the alternatives of the pronoun must be bound by the alternatives of the antecedent.

How is this possible? Why does the binding relation remain unchanged after both the binder and the bound pronoun being focused? Recall that in Chapter III, I proposed that (at least) every D head receives an index variable when they are selected from the lexicon. Note that these are index variables rather than specific index values. Reflexive Binding is an agreement between index variables, and the bound reading of pronouns is a result of accidental index variable identity (e.g., two variables $x = y$) (Hicks 2009). What focus operates on is therefore the representation with index variables rather than specific values. Focus on a variable will not change the variable. What focus value changes is in fact the predicate, as shown by (89a). The corresponding focus value for the simple sentence *only John_F likes himself* can be represented in (89b).

- (89) a. $[D_x \text{ [NP John}_F]] = \{[D_x \text{ [NP John]}], [D_x \text{ [NP Mary]}], [D_x \text{ [NP Bill]}]\}$
- b. $[\text{only John}_F \text{ likes [himself}_F]] = \{[D_x \text{ [NP John]}] \text{ likes [him}_x \text{ [NP self]}], [D_x \text{ [NP Mary]}] \text{ likes [him}_x \text{ [NP self]}], [D_x \text{ [NP Bill]}] \text{ likes [him}_x \text{ [NP self]}]\}$

The propositions in the focus value then must be interpreted separately, allowing

¹⁵Partee (2008) reconciles the paradox between syntactic and semantic binding, proposing that syntactic binding can be a prerequisite for semantic binding: only when a syntactic binding relation is successfully established could the corresponding semantic binding be licensed.

the index variable to be assigned different values across propositions. Inside a single proposition, however, the same index variables must be assigned the same index values, which will give us the effect that binding relations are unaffected in the focus value. Details and other consequences of this approach awaits future research.

Another possibility is that the focus value may include undesired cases such as those in (88), but these cases are filtered out by the context. That is, the burden is shifted from the computation of a proper focus value to the contextually determined C set in Rooth's (1992b) theory. As the ordinary value involves a binding relation, irrelevant propositions that do not involve such a binding relation can be filtered out from the alternatives by the context. The constraint that requires binding relations in the alternative C set to be parallel to that in the ordinary value could be imposed if we assume a restriction similar to the parallelism condition in theories of VP ellipsis (Sag 1976; Williams 1977; Rooth 1992a; Takahashi & Fox 2005): The VP antecedent and the elided VP must be semantically identical. This extra constraint imposed by the context will lead to a desired C set as in (86b), but excluding undesired propositions as in (88). A concern, nevertheless, is that this approach may result in a less formally restricted focus theory.

The last option that comes to mind is: It can be argued that the focus on the bound pronoun is to establish a co-variance (including binding) relation with its focus-marked antecedent. Such a strategy to setting up a co-variance relation has been observed across different phenomena. For instance, Sauerland (2000, 2008) finds that an accented focus (indicated in capitals) on bound pronouns is obligatory when the pronoun is interpreted as a bound pronoun in (90a). If, on the other hand, the pronoun is interpreted as a referential pronoun that bears the same reference as the pronoun in the discourse, the referential pronoun is not accented (90b).

(90) *Discourse*: On Monday, every boy λ_2 called **his**₂ mother.

a. On Tuesday_F, every TEACHER λ_3 called [**HIS**₃] mother.

- b. On Tuesday_F, every TEACHER called **his** mother.

I agree with Sauerland (2000, 2008) and Mayr (2012) that the focus on the bound pronoun is to make a contrast between the (bound) pronoun in (90a) and the pronoun in the discourse of (90), given that these two pronouns must be interpreted as bearing different indices. However, I think there is something more going on here. Intuitively, the accent on the pronoun in (90a) seems to have strengthened or led to the bound reading.

Such an intuition is confirmed by the data from Sauerland (2008), which is built upon experimental results from Hirschberg & Ward (1991). As (91) shows (the subscripts are modified to simplify the discussion), with a focus on the second pronoun, that pronoun is obligatorily interpreted as a bound pronoun (92a), rather than a co-referential pronoun (92b).

- (91) John likes his mother and Bill likes **HIS** mother.

- (92) a. John_i likes his_i mother and Bill_j likes his_j mother.

- b. *John_i likes his_i mother and Bill_j likes his_i mother.

Note that it is not implausible to assume that *Bill* in (92) bears contrastive focus because it is in contrast with *John*. Therefore, the empirical observation is that the focus on the pronoun helps establish a co-variance relation with its focus-marked antecedent.

I do not have a formal account of the above empirical observation, but this phenomenon is reminiscent of Kratzer’s (1991) analysis of “Tanglewood” constructions,¹⁶ exemplified in (93), where VPA stands for verb phrase anaphora, and VPE verb phrase ellipsis.

- (93) a. *Discourse*: The addressee was angry at the speaker and said the following. “What a copy cat you are! You went to Block Island because I did.

¹⁶I should thank Ezra Keshet for bringing “Tanglewood” constructions to my attention.

You went to Elk Lake Lodge because I did. And you went to Tanglewood because I did.” The speaker felt that the addressee exaggerated and replied:

- b. I only [_{VPA} went to Tanglewood_F] because you did [_{VPE} ~~went to Tanglewood~~_F]

Given a context where only {Tanglewood, Block Island, Elk Lake Lodge} are relevant, the C set for (93b) is as follows:

- (94) a. I went to Tanglewood because you went to Tanglewood
 b. I went to Elk Lake Lodge because you went to Elk Lake Lodge
 c. I went to Block Island because you went to Block Island

To obtain (94), Kratzer (1991) assumes that both the occurrences of *Tanglewood* are focus-marked. But it is important to note that if, following Kratzer (1991), both occurrences of *Tanglewood* are focus-marked and thus both induce relevant alternative sets, we should expect the following *C* set, instead of that in (94).

- (95) a. I went to Tanglewood because you went to Tanglewood
 b. I went to Tanglewood because you went to Elk Lake Lodge
 c. I went to Tanglewood because you went to Block Island
 d. I went to Elk Lake Lodge because you went to Tanglewood
 e. I went to Elk Lake Lodge because you went to Elk Lake Lodge
 f. I went to Elk Lake Lodge because you went to Block Island
 g. I went to Block Island because you went to Tanglewood
 h. I went to Block Island because you went to Elk Lake Lodge
 i. I went to Block Island because you went to Block Island

However, it is (94) rather than (95) that is consistent with the interpretation of (93b). Note that in (94) there is a co-variance relation between the two focus-marked

occurrences of *Tanglewood* whereas in (95) there is no such co-variance relation. In order to accurately derive such a co-variance relation between the focused phrases, Kratzer (1991) proposes that focused constituents bear distinguished focus indices, and the computation of the focus value is sensitive to the focus indices. In terms of the example in (93b), the two occurrences of *Tanglewood* bear the same focus index (the subscript *i*).

(96) I only [_{VPA} went to Tanglewood_{*F,i*}] because you did [_{VPE} ~~went to Tanglewood~~_{*F,i*}]

The two focused-marked occurrences of *Tanglewood* that bear the same focus indices are interpreted as distinguished variables that co-vary and range over all the members of the alternative sets:

Kratzer's (1991) analysis can be directly applied to *only*-sentences if we assume that both the binders and bound variables are focus-marked, as shown in (97):

(97) Only I_{*F,i*} did my_{*F,i*} homework.

With focus-indices, the co-variance relation required for (86) can be formally derived.

Note that an analysis that assumes the binding or co-variance relations are established in syntax and then these relations are kept unchanged in the semantic component seems also work with *Tanglewood*-construction. In addition, it could also be stipulated that the focus value of the *Tanglewood*-construction in (93b) is simply (95), but the *C* set can filter out the unintended propositions that are without a co-variance relation, based on the information provided in a discourse such as that in (93a).

Before concluding this chapter, it is worth noting that ϕ -features on bound pronouns are also claimed not to be interpreted under VP-ellipsis (Heim 2008), in examples such as (98).

(98) a. John did his homework;

b. but I_F didn't ~~do my homework~~.

I think it is arguable that in this case, the ϕ -features on the bound pronoun are also not completely semantically uninterpretable. Again, I assume that the bound pronoun in the elided VP is focus-marked, similar to the focus-marked constituent inside the elided VP in the *Tanglewood*-construction. The same kind of explanations I offered for the co-variance relation between bound pronouns and their binders in *only*-sentences and the *Tanglewood*-construction can be applied to account for the co-variance relation between *I* and the elided *my* in (98). That is, the ϕ -features on the bound pronoun are naturally “ignored” in the focus value of the sentence when an alternative set of the bound pronoun is computed. We can again resort to the three solutions I proposed in this section to address the question why in the focus value the pronoun must co-vary with its local antecedent: (i) The binding relation is established in syntax, and cannot be altered by any operations in semantics, including the computation of the focus value; (ii) although the focus value of (98b) may include cases where the object pronoun is not bound by the subject pronoun, the context can filter out these cases without a binding/co-variance relation, based on the fact that (98) involves a binding relation; and (iii) the subject and object pronouns, which are both focus-marked, must be assigned with the same focus indices, and thus must co-vary, following Kratzer (1991).

My solution to the “ignorance” of ϕ -features in the focus value of the VP ellipsis sentences (e.g., (98)) crucially hinges on the assumption that the elided object pronoun is focus-marked, the same as its antecedent. As we have mentioned, this assumption is also held by Kratzer (1991), as is demonstrated by (93). Such an assumption may suffer from criticism based on studies which suggest that focus-marked constituents cannot be elided (e.g., Heim 1997; Merchant 2001; Takahashi & Fox 2005, but see Bassi & Longenbaugh to appear; Sauerland 2000 for counterarguments). However, as we have seen, focus-marked constituents can be elided only in a particular

context: when it is in a co-variance relation with another focus-marked constituent. Although I do not know why this restriction exists; it is definitely an interesting topic for future studies.

4.6 Summary

In this chapter, I introduced the interpretability problem concerning the ϕ -features on bound pronouns in *only*-sentences, followed by the solution that the morphosyntactic account (the PF Feature Transmission Approach) offers: Bound pronouns are minimal pronouns that do not bear inherent semantic ϕ -features from the lexicon, and their ϕ -features at PF are transmitted from their binders. Previous studies have argued for an alternative approach, the presupposition account, which assumes that ϕ -features on both bound and referential pronouns are brought from lexicon and are semantically interpreted as presuppositions. I then pointed out that the evidence offered by previous studies against the morphosyntactic account, however, is mostly indirect, and does not address the interpretability problem of ϕ -features on bound pronouns directly.

Spathas's (2010) argument is an exception. Spathas (2010) indicates that in Greek, a pronoun bound by a DP with a morphosyntactic neuter gender and a natural feminine gender can be either pronounced as neuter or feminine. The morphosyntactic account predicts that only the morphosyntactic neuter feature can be transmitted and thus be pronounced on the bound pronoun. Nevertheless, I showed that Spathas's argument is not a knock down counterargument against the morphosyntactic account if the morphosyntactic account is revised by assuming that both the natural and neuter gender features are transmitted to the bound pronoun.

I considered two possible revised versions of the morphosyntactic account: the revised PF Feature Transmission Approach and the Full Feature Transmission Approach. The revised PF Feature Transmission Approach assumes that both natural

and morphosyntactic gender features are treated as phonological/morphosyntactic features that can be transmitted to the bound pronoun at PF, whereas the Full Feature Transmission Approach hypothesizes that both semantic (the natural gender feature) and morphosyntactic features are transmitted to the bound pronoun through co-indexation. Both approaches predict that either the semantic or phonological/morphosyntactic feature can be pronounced on the bound pronoun.

I then provided novel arguments based on new data from English, Chinese, Arabic and Russian to support the presupposition account. The arguments provide direct evidence against both versions of the PF Feature Transmission Approach and against the Full Feature Transmission Approach. The first argument is based on collective nominal binders in English, which are argued to be morphosyntactically singular but semantically ambiguous between a singular group reading and a plural member reading. I showed that when the bound pronoun is singular, only the group reading of the collective nominal binder is available; when the bound pronoun is plural, only the member reading is available. Kinship nouns and other relational nouns were used as a diagnostic to tease apart these two readings. The observed pattern strongly argues against the two versions of the PF Feature Transmission Approach, because the PF Feature Transmission Approach and its revised version predict that only singular bound pronouns are possible and that the choice of the singular and plural pronouns should not distinguish meaning, contrary to fact.

I developed another argument built upon data from Chinese bound pronouns, simplex reflexives, and complex reflexives, to argue against the PF Feature Transmission Approach, its revised version, as well as the Full Feature Transmission Approach. This is because all these morphosyntactic approaches predict that bound pronouns, the simplex reflexive and complex reflexives must behave similarly, for example, when they are bound by collective nominal binders. However, I showed that this prediction is incorrect with regard to the Chinese data, as the simplex reflexive accesses only

the group reading of the collective nominal binder whereas complex reflexives and pronouns can access either the group or the member reading.

To extend my argument to ϕ -features beyond the number feature, I presented an argument built upon novel data from Standard Arabic and Russian, where we can investigate the spell-ed out ϕ -features on bound variables. The Arabic data strongly support the argument that both the number and gender features on bound pronouns are semantically interpretable, as they are determined by the reference of the bound pronouns. Data from Russian further indicates that the morphosyntactic/phonological gender feature on possessive bound variables (the gender feature is phonological as it does not interact with the semantic interpretation of the bound variables) is determined by a syntactic agreement with the possessee, contrary to the prediction of the morphosyntactic account. As a result, I concluded that (all versions of) the morphosyntactic account I considered in this chapter is not defensible.

In the end of the chapter, I also provided a solution to the interpretability problem of the ϕ -features on bound pronouns, along the lines of the presupposition account. The idea is that if we assume an optionally unaccented focus on the bound pronoun in *only*-sentences, it naturally follows that the ϕ -features on the bound pronouns are “ignored” in computing the alternative set of the bound pronouns. The alternative set of the bound pronouns will then co-vary with the focus alternative set of the pronoun binder. This results in the seemingly “ignorance” of the ϕ -features on bound pronouns in the focus value, but not in the ordinary value. In other words, the ϕ -features on bound pronouns are not really semantically uninterpretable in any steps/computations in the current analysis.

I also argued that the analysis is applicable to bound reflexives, although I used bound pronouns primarily in the examples, given that previous discussions are mostly concerned about bound pronouns. This analysis thus suggests that the ϕ -features on bound complex reflexives are also semantically interpreted as presuppositions. The

ϕ -features on T heads, on the other hand, cannot be interpreted as presuppositions, although they may signal the ϕ -features of the subject DP they agree with. Consequently, we derive an important distinction between subject–verb agreement and reflexive binding: On the sites of the T head and the reflexive, the ϕ -features involved are different in their semantic interpretability. In Chapter VI, I aim to relate the syntactic and semantic analyses of these two constructions to the processing of them.

CHAPTER V

Encoding Syntactic Relations as Features

5.1 Introduction

In Chapters II and III, I provide a syntactic analysis of agreement and binding, arguing that these two constructions can be captured by a unified syntactic analysis, minimal search-based agree (MS-agree). In Chapter IV, I provide a semantic analysis of the ϕ -features on bound variables, and suggest that the ϕ -features on bound variables, including reflexives, are semantically interpretable, differently from the ϕ -features on T. The conclusion is that subject–verb agreement and reflexive binding are similar in their syntactic analysis, but the interpretability of the ϕ -features involved in these two types of dependencies differ. The interpretability difference may lead to processing differences between subject–verb agreement and reflexive binding, addressing the first challenge to the cue-based retrieval theory as mentioned in Chapter I.

In this chapter, I aim to address the second challenge to the cue-based retrieval theory: the problems of encoding syntactic dependencies as features. Specifically, this is a question of how long-distance syntactic relations between the antecedent and the reflexive are encoded in parsing. Therefore, this chapter is not written from the perspective of syntactic theory; rather, it is written from the perspective of parsing, which is in turn based on certain assumptions of syntactic analysis. This chapter

proposes an encoding theory for syntactic relations such as the c-command relation and the locality constraint in reflexive binding, setting up a basis for the cue-based retrieval theory of reflexive binding in parsing.

Let us use the following examples, similar to those discussed in Chapter I, to illustrate the research questions of this chapter.

- (1) a. The new executives who oversaw the middle managers apparently **were** dishonest about the company's profits.
b. The new executives who oversaw the middle managers apparently doubted **themselves** on most major decisions.

Imagine that we read the sentences word-by-word, not being allowed to re-read a word once we move to the next word. Now we are reading the word *were* in (1a) or *themselves* in (1b). The problem here is how we can retrieve the constituents that are in a syntactic dependency with *were* or *themselves* from our memory. For the examples in (1), the constituents to be retrieved are the subjects of the sentences. I call these constituents the target of memory retrieval, and *were* or *themselves* the retrieval sites.

Assuming the cue-based retrieval theory (Lewandowsky & Murdock 1989; McElree et al. 2003; Lewis & Vasishth 2005; Lewis et al. 2006), in order to correctly retrieve the target from the memory, a set of retrieval cues must be used to match against features on the target. Given that the targets in (1) are in a syntactic dependency with the retrieval sites, we have to use the syntactic dependency as a retrieval cue.

For subject-verb agreement, traditionally, the T head will initiate a memory retrieval for the subject DP. Under the MS-agree analysis proposed in Chapter II, the T head will need to agree in ϕ -features with the head of the subject DP, i.e., the D head, which is the label of the DP. In addition to the agreement relation, we have to establish a co-member relation between the subject DP and T' that is headed by T. Theoretically, all these have to be done to improve the possibility of successful

retrieval of the target. One way to implement the retrieval is to set the retrieval cues based on the ϕ -feature specifications on T and the syntactic position of the subject DP. Consequently, the retrieval cues for the retrieval at the retrieval site *were* should include [Num: PL, SyntacticPosition: SpecTP]. The number feature of the DP is the number feature on its label, the D head.¹ A more detailed discussion of the retrieval cue specification will be provided in the next chapter.

However, a serious problem arises if we want to implement the syntactic dependency between the reflexive and its antecedent in (1b) under the cue-based retrieval theory. This is because the syntactic dependency between the antecedent and the reflexive is not determined by the definite syntactic positions of either one of them or even both of them. Instead, the dependency is determined by their positions relative to each other. Previously, the dependency is claimed to be restricted by the c-command relation. Note that the MS-agree analysis of reflexive binding presented in Chapter III does not give us the c-command relation. However, since c-command provides a good approximation of relations established by MS-agree, I will address first the following question: How can we encode the c-command relation as features, so that these features can be used as retrieval cues in memory retrieval?

I apply a current syntactic theory to solve the c-command feature encoding problem. Following Epstein et al.'s (1998) derivational conceptualization of c-command in terms of sisterhood (and parenthood), I redefine the c-command relation as a c-commanded-by relation, from the perspective of c-commandees rather than c-commanders. Like Epstein et al.'s (1998) derivational c-command, derivational c-commanded-by is also based on merge, a fundamental operation in minimalist syntax theory. With merge, I reduce long-distance c-command relations to local relations. With the new definition, the c-command relation can be encoded as [C-commanded-by] features on

¹Note that we can use the label for retrieval purposes, as we assume labels are for interpretation purposes and memory retrieval is part of the parsing of the sentence. Labels are not available only for syntactic operations in the syntax module.

lexical items.

5.2 Representational c-command

C-command is an important component of contemporary syntactic theories. It plays a significant role in Binding Theory. Binding conditions are defined directly based on the c-command relation. The representational definitions of c-command, Condition A and Condition B are shown in (2), (3) and (5), respectively.

(2) *Definition of c-command (the representational version)*

A c-commands B iff

- a. neither A nor B dominates the other, and
- b. the first branching node that dominates A also dominates B.

(3) *Binding Condition A*

A reflexive must be bound within its governing category.

(4) A binds B iff

- a. A c-commands B, and
- b. A and B are co-indexed.

(5) *Binding Condition B*

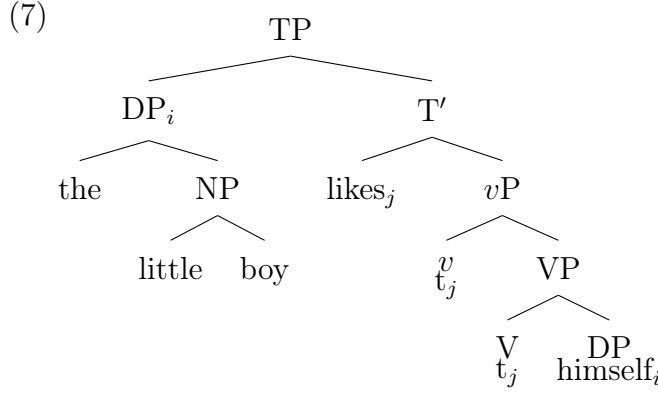
A pronoun must be free (not bound) in its governing category.

(6) *Definition of governing category*

X is a governing category for Y iff X is the minimal category (minimal TP or DP) containing Y, a governor of Y and a SUBJECT accessible to Y.

All the definitions are taken from Chomsky (1981, 1986). I identify the definition of c-command as representational because I will later contrast it with a derivational definition of c-command.

The most relevant definitions to our discussion here are the definition of c-command (2) and the definition of Binding Condition A (3). (7) shows how these definitions work.



In (7), *the little boy* c-commands *himself*, because neither *the little boy* nor *himself* dominates/contains the other, and the first branching node that dominates *the little boy*, that is, the TP, also dominates *himself*. In addition, *the little boy* is also co-indexed with *himself*. Given the definition in (4), *the little boy* binds *himself*.

However, the c-command relation cannot be properly formulated in minimalist program, mainly because under minimalism, the only legitimate operation for structure-building is merge. We mentioned the definition of merge in Chapter I, but for clarity, I repeat the definition below in (8).

(8) *Definition of merge*

Given any two distinct syntactic objects α , β , $\text{merge}(\alpha, \beta) = \{\alpha, \beta\}$.

The definition in (8) shows that merge takes two syntactic objects, α and β , which could be lexical items or sets of lexical items, and returns a set with α and β as its members.

Based on this definition, only two syntactic relations can be constructed with merge: immediate containment and co-membership. If α merges with β and the set κ is returned, κ immediately contains α and β . In addition, α is a co-member of β . In

other words, the c-command relation cannot exist except that it can be redefined in terms of immediate contain and co-membership. This is precisely the direction which some researchers are heading in: deriving the c-command relation from merge. This is usually called the derivational approach to c-command relations (syntactic relations broadly) (Epstein et al. 1998). The derivational approach to c-command relations considers c-command a result of merge. I will review this approach in detail, but before that, I would like to introduce an “intuitive” but unsuccessful way to encode the c-command relation as features, in the hope that the reader can appreciate more about why there is an encoding problem.

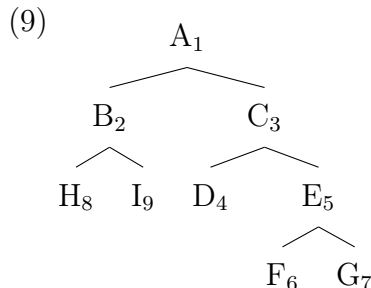
5.3 Previous attempts to encode c-command as a feature

5.3.1 An “intuitive” approach

A possible solution to the encoding problem regarding the c-command relation is one that immediately follows the definition of c-command in the last section: encoding the c-command feature on the c-commanders. That is, we may create a value slot on the feature list of the c-commanders, and then store there the ID of each chunk that the c-commander c-commands. For example, if A c-commands B₁, C₂ and D₃ (the numeric subscripts are IDs of the corresponding chunks), then the c-command feature slot on A will be [C-command: 1,2,3]. In terms of reflexive binding, the antecedent c-commands the reflexive, so the ID of the reflexive will be included in the value of the antecedent’s [C-command] feature. When the reflexive is processed in reading, a retrieval will be initiated and the chunk ID of a reflexive can be employed as a retrieval cue.

Kush (2013) and Kush et al. (2015) have an insightful discussion on the above intuitive approach and finally conclude convincingly that the intuitive approach is too inefficient. Under this approach, the c-commanding items bear a [C-command]

feature slot and the value of the feature needs to be updated consistently along the course of parsing. For instance, given the following structure tree (9), B₂ has a [C-command: 3,4,5,6] feature after B₂, D₄ and F₆ are parsed and are built into the structure.



The problem is, the value of the [C-command] feature on B₂ needs to be consistently updated when more words are processed and attached to the syntactic tree, if these words are c-commanded by B₂. For instance, when G₇ is parsed and is attached to F₆, given that G₇ is c-commanded by B₂, the [C-command] feature on B₂ needs to be updated to include G₇'s chunk ID. The same problem is present for other chunks such as D₄ and F₆, as their [C-command] value must also be updated when new c-commandees are attached to the structure.

Kush (2013) disfavors this approach because it puts too much burden on the parser: According to the cue-based retrieval theory under the ACT-R architecture (Anderson & Lebiere 1998; Anderson 2005), in order to update the [C-command] feature on a “distant” constituent, that constituent must be successfully retrieved. This is because ACT-R architecture puts a limit on the number of chunks that are accessible for manipulation at any stage. Taking again the structure in (9) for an example. When G₇ is under processing, only F₆ (and probably E₆, depending on the kind of parser adopted) are available for manipulation. All other chunks including B₂ and D₄ are stored in the working memory. Therefore, to update the [C-command] feature on B₂ and D₄ requires the successful retrieval of these chunks first. These chunks would need to be retrieved one-by-one and then their [C-command] feature value would be

updated in order. In other words, the retrieval and updating operations occur every time a new syntactic object that is c-commanded by a long-distance c-commander is processed. In terms of the structure in (9), when G_7 is processed and is attached to the structure, new c-command relations between B_2 , D_4 , F_6 and G_7 are formed. In addition to the updating of the [C-command] feature on F_6 , B_2 and D_4 must be retrieved first in separate retrievals and then their [C-command] features are updated, as shown below.

(10) *The update of the c-command feature values*

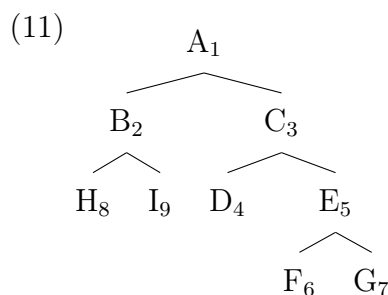
- a. $F_6: [] \rightarrow [G_7]$
- b. $D_4: [E_5, F_6] \rightarrow [E_5, F_6, G_7]$
- c. $B_2: [C_3, D_4, E_5, F_6] \rightarrow [C_3, D_4, E_5, F_6, G_7]$

As Kush (2013) points out, if the retrieval of the c-commanding chunks and the updating of the [C-command] features takes place in actual sentence processing, we expect that parsed chunks that are attached to the built structure later should be generally processed more slowly than those that attach to the structure earlier, because, in general, the later a syntactic object attaches to the structure, the more c-commanding syntactic objects need to be retrieved and have their [C-command] features updated, and therefore the longer time it will take to finish the process. However, no evidence shows that the reading times of syntactic objects increase continuously throughout the processing of a sentence. Instead, there is evidence that the reading rate increases as subjects move through a sentence, instead of slowing down. (Van Dyke & Lewis 2003).

5.3.2 The spine-based approach

Kush (2013) therefore concludes that an approach that explicitly specifies [C-command] as a chunk feature is computationally inefficient and empirically inadequate.

quate. He proposes an alternative: an approximation of the c-command relation through the “spine” feature in his system, following Alcocer & Phillips (2012). “A spine can be (informally) defined as a path through the right-most, non-adjunct daughters of a series of connected phrases.” (Kush 2013:120) Every time that a branching node branches leftward, a new dominance spine is created. Take the tree in (9) again for an example, repeated below in (11): It contains five spines, each with an index, as shown below.



(12) *Spines in (11)*

Spine 1: H₈

Spine 2: B₂ and I₉

Spine 3: A₁, C₃, E₅, G₇

Spine 4: D₄

Spine 5: F₆

The [C-command] feature, which Kush et al. (2015) call as ACCESSIBLE feature, on a node A relating to another node B is determined by the spine index of A and B’s parents: If their parents share the same spine index, then they are in a c-command relation. This is defined in (13), according to Alcocer & Phillips (2012).

(13) *The approximation of c-command using spines* (Alcocer & Phillips 2012:31)

A pair of nodes whose parents share the same spine index stand in a c-command relation.

For example, in (11), D₄ and F₆’s parents are C₃ and E₅ respectively. C₃ and E₅

are in the same spine, i.e., Spine 3 in (12). According to (13), given that D_4 and F_6 's parents share the same spine index, i.e., Spine 3, D_4 and F_6 are in a c-command relation.

In Kush's (2012) system, the spine index of a node's parent is encoded as a feature of that node. For instance, a reflexive will bear a spine index of its parent, and its antecedent will also have its own parent's spine index. As long as the reflexive and its antecedent share the same value on their parents' spine index, they are in a c-command relation. In memory retrieval, one of the retrieval cues will be something like [Parent_Spine: i], where i is the stored spine index of the parent node.

Note that Kush (2013) also tries to account for the distribution of pronouns (Binding Condition B in (5)). According to Binding Condition B, a pronoun must not be bound in its local clause (I follow the terms used by Kush here). However, a pronoun can be bound by a quantified phrase which is not in its local clause. To account for the distribution of pronouns, Kush resorts to two more features besides [C-command], LOCAL and ACTIVE. The [Local] feature is updated to reflect clause (CP) boundaries, by assigning each CP a different index. The [Active] feature is a binary feature to identify the scope of a quantified phrase, by encoding the quantified phrase [Active: 1] when chunks inside its scope are undergoing parsing, and encoding it [Active: 0] when the parser moves to a chunk that is outside of the quantified phrase's scope.

While the success that Kush's (2013) spine-based approach achieves in encoding the c-command relation as a retrieval cue should be fully appreciated, this approach faces several challenges, some of which have already been recognized by Kush himself and by Alcocer & Phillips (2012). First, this approach "provides a way of encoding a general proxy for surface c-command feature." (Kush 2013:120) Therefore, it is not a strict application of the c-command relation in parsing. For instance, the approximation in (13) only tells us if two chunks are in a c-command relation or not, and it cannot capture asymmetric c-command relations. As a result, (14a) is

predicted to be identical with (14b), which is incorrect.

- (14) a. John likes himself.
 b. *Himself likes John.

Similarly, the spine-based c-command approximation cannot capture the asymmetric c-command relation between the direct object and indirect object (15a, b): *Mary* c-commands *a picture of herself* in (15a) but not in (15b).

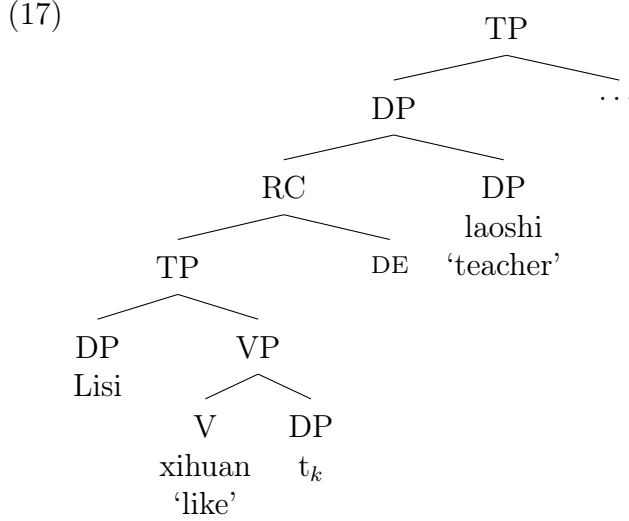
- (15) a. John gave Mary_i a picture of herself_i.
 b. *John gave a picture of herself_i to Mary_i.

Barss & Lasnik (1986) and Larson (1988) argue convincingly that the first object always asymmetrically c-commands the second object in double object or dative constructions. Since the spine-based approach only provides us a symmetric c-command relation between these two objects, it cannot account for the differences between (15a) and (15b).

In addition, the spine-based approach misses all cases where the c-commander is higher in a right spine of the tree and the c-commandee is deep down in a left spine (Alcocer & Phillips 2012). Counter-examples abound if cross-linguistic data are examined. For example, this approach cannot capture the c-command relations in all Chinese relative clauses. In (16), the head noun, *laoshi* ‘teacher,’ c-commands its trace inside the relative clause for the trace to be licensed (the Empty Category Principle, Chomsky 1981).

- (16) [_DP [_RC Lisi xihuan t_k de] laoshi_k] yijing likai xuexiao-le.
 Lisi like DE teacher already leave school-SFP
 ‘The teacher who Lisi likes has already left the school.’

However, the spine-based approach predicts that the spine index of the trace’s parent is not the same with that of its antecedent’s parent, as shown in the tree diagram in (17).



Generally, Generally, relative clauses will be a problem in a language that is head-final in its relative clause domain but is head-initial in some other domains such as VPs and TPs. English relative clauses are exception in not being a problem mainly because the relative clauses are head-initial, the same as other types of phrases such as VPs and TPs.

Another important concern is that the approximation of c-command is not motivated by any syntactic theory, neither it bears any psycholinguistic consequences. Therefore, it categorizes itself merely a computational algorithm that approximates the c-command relation.

5.3.3 The C-COMMAND-AS-CONTENT approach

Besides their pioneer work on the spine-based approach, Alcocer & Phillips (2012) in fact considers another algorithm which they do not peruse further in enough details. They call this alternative approach C-COMMAND-AS-CONTENT, and give the following algorithm to compute the [C-commanded-by] features on a chunk under parsing:

(18) *Alcocer & Phillips's (2012) C-COMMAND-AS-CONTENT approach*

1. Retrieve the current node N's parent node P and copy the list of P's c-commanders onto N, if it has any.

2. Verify whether P has a child other than N. If it does, this is N's sister node S.
3. Add S's label to N's list of c-commanders.
4. Project structure based on N and the state of the goal buffer.
5. Add nodes in projected structure to a processing queue.
6. Apply steps 1-4 to each of the nodes in the processing queue.
7. Receive the next word and repeat steps 1-6.

Step 1 is to copy the [C-commanded-by] feature of a parent node to its daughter node. Step 2 and 3 are to include the index of a node in its sister's [C-commanded-by] feature.

This approach is different from both the intuitive approach and the spine-based approach we reviewed above, because instead of encoding the c-command feature on the c-commanding chunks, the C-COMMAND-AS-CONTENT approach switches the burden to the c-commandee being parsed, making use of parent-child relations (Step 1) and sisterhood relations (Step 2 and 3).

I will argue later that the C-COMMAND-AS-CONTENT approach is actually preferable because of its theoretical advantages. But before we proceed, I would like to mention some disadvantages of this approach, as pointed out by Alcocer & Phillips (2012). I will show in the next section that with a small but crucial modification, the problems go away.

First, the [C-commanded-by] feature on a c-commandee usually contains a list of chunk IDs of its c-commanders. This violates an assumption of the ACT-R model; namely, feature slots do not usually use lists as their value. In addition, sometimes the list of the c-commanding chunks is very long, which will significantly increase the size of the memory buffer, conflicting with the idea that a buffer should take limited memory resource.

Another relevant concern is that if a list of c-commanding chunk IDs functions as retrieval cues, then no antecedent can match all the features, which results in partial match penalty, and the association strength between the target and the retrieval cues will be significantly reduced. Consequently, the C-COMMAND-AS-CONTENT approach predicts that sometimes the target cannot be retrieved.

It must be noted that the second problem is mainly caused by the relatively large number of c-commanding chunks being used as retrieval cues. Notice that similar problems also occur in other approaches such as the spine-based approach. As long as the c-command relations are encoded as item features, there will be a list of c-commanders encoded somewhere in the system. For the spine-based approach, suppose the spine index of a reflexive parent is i ; the retrieval cue will match all the chunks whose parent spine index is i . This gives rise to a bunch of distractors competing with the target. Therefore, the second problem is shared by the spine-based approach and the C-COMMAND-AS-CONTENT approach, and thus cannot be used as an argument for one approach over the other.

I will show later in the next section, however, that the C-COMMAND-AS-CONTENT approach is most consistent with a recent development in syntax theory regarding c-command, and the two problems raised above can be solved based on a simplification of the system.

5.4 Proposal: derivational c-commanded-by

5.4.1 Derivational c-command

Recall that a fundamental issue of the “intuitive” approach is that it assumes a representational c-command for parsing. In order to provide a plausible solution to the encoding problem of the c-command relation, we need a derivational conceptualization of c-command instead. A current theory in syntax, the derivational approach to

syntactic relations, as proposed by Epstein et al. (1998), is exactly what we need.

Epstein et al. (1998) argue that the independently motivated merge operation can capture representational c-command relations. The leading idea is that only when X merges with Y are X and Y involved in a c-command relation; that is, X c-commands Y as well as every syntactic objects that Y dominates, and vice versa.

The derivational approach offers an explanation of c-command by deriving c-command from merge, the only independently motivated, structure-building operation. To put it simply, if X and Y are merged, they enter into what we call “c-command relations” (Epstein et al. 1998:30); and if X or Y dominates the other, then X does not merge with Y, explaining the restriction in the definition of c-command in (2a); namely, A c-commands B requires neither A nor B dominates the other.

We now can say that the reason why there is a c-command relation between X and Y is because X and Y are merged at some derivational step. But this only explains symmetric c-command relations, that is, when X c-commands Y and meanwhile Y c-commands X. What about the cases where X asymmetrically c-commands a syntactic object Z that is dominated by Y, i.e., X c-commands Z but Z does not c-command X? Epstein et al. (1998) propose the following definition to capture all representational c-command relations (see a revised formalization of this definition in Epstein et al. 1998:170; see also Epstein 1999). The terms of Y include Y and all the syntactic objects in Y. The definition of TERM is adopted from Chomsky (1995:5).

(19) *Derivational C-Command* (Epstein et al. 1998)

X c-commands all and only the terms of the category Y with which X was paired/concatenated by Merge or Move in the course of the derivation.

(20) *Definition of TERM*

For any structure K,

- i. K is a term of K
- ii. If L is a term of K, every member of L (i.e., every element in L’s domain)

is also a term of K.

According to (19), if Z is not Y but a term of Y, and X merges with Y, then X c-commands Z. In this case, however, Z does not c-command X, as at no point of the derivation does Z merge with X or any set containing X. This explains how asymmetric c-command relations are established.

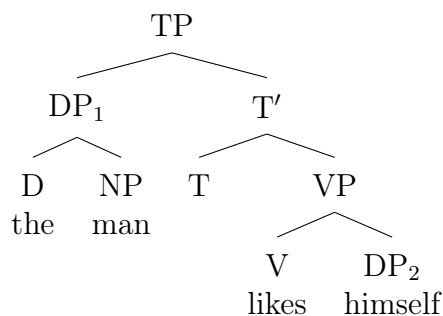
Epstein et al. (1998) point out that, by contrast, the representational c-command relation is a stipulated relation because we cannot explain why there must be a relation between X and Y if “the first branching node that dominates X also dominates Y,” and why it must be “neither X or Y dominates the other.”

In sum, the derivational definition of c-command, if on the right track, sheds light on where the c-command relation comes from. As noticed by Chomsky (2000a), merge provides two relations directly: sisterhood (= co-member) and immediate containment (= is a member of). C-command is a combination of these two relations. Both co-member and immediate containment are local relations. This conceptualization of c-command opens a way to encode the c-command relation as a local feature. However, a very important question here is how X can relate to the terms of Y when X merges with Y. In the next section, I will show that this is done by inheritance of the c-command feature from Y to its terms. In order to get this inheritance of the c-command feature into place and implement the idea in parsing, we need to reinterpret derivational c-command from a different perspective.

5.4.2 Relating derivational c-command to parsing

Let us now relate derivational c-command to parsing. Given the syntactic tree below, with irrelevant nodes omitted, the definition of derivational c-command says that DP₁ *the man* c-commands DP₂ *himself*, because DP₁ merges with T' and DP₂ is a term of T'.

(21) The man likes himself.



DP₂ does not c-command DP₁ because at no point during the course of derivation does DP₂ merge with DP₁ or a syntactic object containing DP₁. Therefore, derivational c-command correctly predicts that there is an asymmetric c-command relation between DP₁ and DP₂.

How can we now compute the c-command relation between DP₁ and DP₂ in sentence (21) in parsing? There are several possible ways, but first we want to exclude all non-incremental hypotheses that assume the parser computes the c-command relation only until the completion of the whole sentence. These approaches are highly implausible, as we have seen in the previous section. It seems quite obvious that comprehenders interpret the reflexive right when the reflexive is encountered.

An approach that assumes incrementality in parsing but meanwhile keeps the system simple is the following: The parser computes the c-command relation between a possible antecedent and a reflexive only when a reflexive is encountered. In order to compute the c-command relation, the parser needs to either (i) search through the structure tree parsed so far to find a potential antecedent or (ii) use cue-based retrieval to retrieve each node that dominates the reflexive and check their sisters to see if their sisters are possible antecedents. The former approach has been proposed by Phillips et al. (2011). Such an approach requires an extra task in addition to the LV05 system: a special, independent searching algorithm. It is unclear how such a search algorithm can be implemented in the ACT-R model. The latter approach is a version of cue-based retrieval, and as we have also mentioned in Section 5.3,

previous studies have identified two possible implementations, Kush’s (2013) spine-based approach and Alcocer & Phillips’s (2012) C-COMMAND-AS-CONTENT approach. We have seen that the spine-based approach is an approximation of the c-command relation and cannot capture the empirical data in many ways. Importantly, it is not based on any syntactic theories and thus lacks a theoretical basis. I thus want to explore the C-COMMAND-AS-CONTENT approach a little bit more below.

How can we combine the insights from the C-COMMAND-AS-CONTENT approach, which I briefly reviewed in Section 5.3.3, with those of the derivational c-command approach? The derivational c-command approach has reduced the c-command relation to merge (and in fact also immediate containment). Merge (and immediate containment) creates local relations, which is consistent with the goal of the C-COMMAND-AS-CONTENT approach, i.e., to encode the c-command relation as a local feature on lexical items. In other words, the derivational approach to c-command relations actually provides the C-COMMAND-AS-CONTENT approach a theoretical basis. This is an interesting direction that is worth further exploration, especially in comparison to the spine-based approach, which does not have a theoretical basis. Unfortunately, a closer examination of the assumptions of the two theories reveals important incompatibilities between them.

The derivational approach to c-command relations, the definition of which is repeated below in (22), assumes a bottom-up structure building strategy (as standardly assumed in minimalism): when the c-command relation between two syntactic objects X and Y is computed, the structures of X and Y have been completed.

(22) *Derivational c-command* (Epstein et al. 1998:32)

X c-commands all and only the terms of the category Y with which X was paired/concatenated by Merge or by move in the course of the derivation.

However, in real time parsing most of the terms of Y may not be available when X merges with Y, if the LV05 left-corner parser is assumed. A possible solution is the

“intuitive” approach we reviewed, namely, encoding the c-command features on c-commanders, and later updating the features whenever new c-commandees are added into the structure. As we have mentioned, this solution has been disfavored due to Kush’s (2013) convincing illustration of its inefficiency.

On the other hand, the C-COMMAND-AS-CONTENT approach encodes the c-command features on c-commandees, rather than on c-commanders, which avoids overwhelming retrievals. This is because whenever new elements are parsed and added to the structure, they will simply inherit the [C-commanded-by] features from their parents and add their sister’s ID to their [C-commanded-by] features.

Assume that the C-COMMAND-AS-CONTENT approach is promising, I will now show that with a revision made to the derivational approaches to c-command relations, the C-COMMAND-AS-CONTENT approach can be compatible with the derivational theory of c-command. The idea is to shift our focus from the c-commanders to the c-commandees, as required by the C-COMMAND-AS-CONTENT approach. This turns out not to be a significant change to the derivational approach to c-command relations in (22), as shown in (23).²

(23) *Derivational c-commanded-by*

Given syntactic objects X, Y and Z, X is c-commanded by Y iff

- a. X and Y are merged (co-member), or
- b. Z is c-commanded by Y, and Z contains X (containment).

(24) The containment relation is transitive:

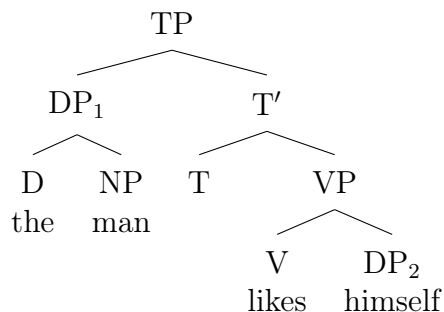
If A contains B, and B contains C, then A contains C.

The leading idea underlying (23) is to switch the c-command feature from c-commanders to c-commandees.³ I thus change the name of the feature from [C-command] to [C-commanded-by]. (23a) is easy to understand. It means that when X

²Note that since merge and move have been unified as a single operation, i.e., merge (Chomsky 2001a; Kitahara 1995), I will not distinguish them here.

and Y are merged, X is c-commanded by Y and vice versa. (23b) is an operation that can be recursively applied, based on the general transitive property of the containment relation (24). For instance, given the syntactic tree in (21), repeated below as (25), what we are most interested in is the c-command relation between DP₁ *the man* and DP₂ *himself*.

(25) The man likes himself.



As I mentioned above, the computation of the [C-commanded-by] feature must start from the c-commandees. Let us take DP₂ as a starting point. Since DP₁ and DP₂ are in a long-distance c-command relation (not in a co-member relation), we should apply (23b) and (24). That is, DP₂ is c-commanded by DP₁ because T' contains DP₂ and T' is c-commanded by DP₁. Why does T' contain DP₂? This is because T' contains VP and VP contains DP₂. According to (24), the containment relation is transitive; it can be therefore derived that T' contains DP₂.

With the definition of derivational c-commanded-by in (23), we can encode the c-command relation as a local [C-commanded-by] feature. The next subsection gives a demonstration of how to apply the derivational c-commanded-by approach in parsing.

5.4.3 C-commanded-by in parsing

Corresponding to the two conditions in derivational c-commanded-by (23), i.e., the co-member and the containment relations, we need two operations in parsing to

³See Chametzky 1996:26 for a similar idea under the framework of representational c-command.

encode the [C-commanded-by] feature on the c-commandees, PASS and INHERIT, as defined in (26).

(26) *Operations for encoding the [C-commanded-by] feature in parsing*

Given K, the parent of Y, and X being attached to Y:

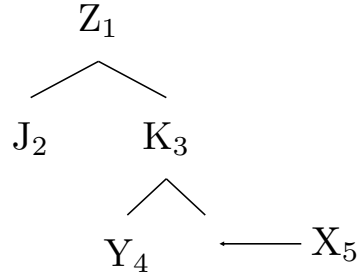
- a. **Inherit** (X, K): The value of K's [C-commanded-by] feature is added to that of X's.
- b. **Pass** (X, Y): The ID of Y is added to X's [C-commanded-by] feature, and vice versa.

For convenience of discussion, I derive the operation SHARE from the operation Inherit:

(27) **Share** (X, Y): Copy Y's [C-commanded-by] feature value to X.

In practice, using Share is sometimes more convenient than using Inherit because the value of the [C-commanded-by] feature on the parent of Y has already been Inherited by Y when X attaches to Y, given LV05's left-corner parser. Sharing the [C-commanded-by] feature from Y is the same to Inheriting the feature from Y's parent. Sharing must occur before Passing, otherwise when attaching X to Y, the ID of Y will pass to X first and then return to itself by Sharing. All of the three operations are triggered by new attachments in parsing. (29) is a demonstration of the implementations of the three operations, given the syntactic tree shown in (28), where subscripts indicate chunk IDs.

(28) *X attaches to Y as a daughter of K*



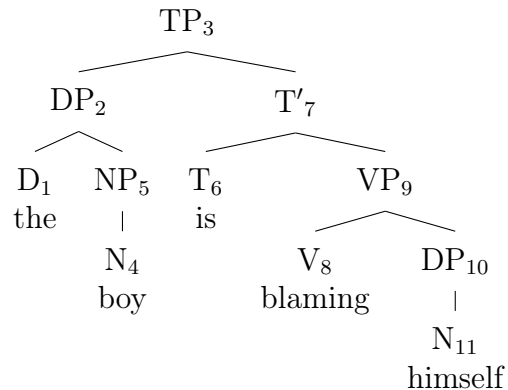
(29) Current state: Y_4 and its parent K_3 bear the [C-commanded-by: 2] feature, given J_2 c-command both of them.

- Step 1: X_5 attaches to Y_4
- Step 2: X_5 Shares the [C-commanded-by: 2] feature from Y_4 (or X_5 Inherits the feature from K_3)
- Step 3: X_5 and Y_4 Pass their ID to each other's [C-commanded-by] feature

The steps in (29) suggest that K_3 , Y_4 and X_5 all have a [C-commanded-by: 2,4] feature. This correctly captures the fact that all these constituents are c-commanded by J_2 and Y_4 .

Let us now use a real linguistic example to illustrate how the new system works. Consider (30), where the subscripts are IDs of the chunks.

(30) The boy is blaming herself.



The steps of parsing are shown below, given a left-corner parser adopted by LV05 under ACT-R.⁴ I present only the most relevant steps, highlighting steps that directly modify the [C-commanded-by] feature with boldface.

- (31) Expect for TP_3
- INPUT: *the*
- Expand *the* to D_1
- Create DP_2 with D_1 as its head
- Retrieve TP_3
- Attach DP_2 as a daughter of TP_3
- Expect for T'_7
- Pass**(T'_7 , DP_2):
- T'_7 : [C-commanded-by: 2]
- DP_2 : [C-commanded-by: 7]
- Retrieve D_1
- Inherit**(D_1 , DP_2):
- D_1 : [C-commanded-by: 7]
- INPUT: *boy*
- Expand *boy* to N_4
- Create NP_5 with N_4 as its head
- Retrieve D_1
- Attach NP_5 to D_1

⁴The algorithm adopted here is the same with the traditional left-corner parser (Resnik 1992) and that in LV05, except that I assume that attachment of a newly parsed word should target its sister node if its sister is already there, rather than its mother node. For instance, in (30), when *boy* is parsed and NP_5 is constructed, NP_5 attaches to D_1 as its sister rather than to DP_2 as its second daughter. Nothing more is needed for this revised algorithm but a new way to construct the whole tree. The gain is that we can apply **Pass** directly upon attachment, rather than using an additional retrieval to access the sister of the newly parsed chunk.

Share(NP₅, D₁):

NP₅: [C-commanded-by: 7]

Pass(NP₅, D₁):

NP₅: [C-commanded-by: 7, 1]

D₁: [C-commanded-by: 7, 5]

Inherit(N₄, NP₅):

N₄: [C-commanded-by: 7, 1]

INPUT: *is*

Expand *is* to T₆

Retrieve T'₇

Attach T₆ as a daughter of T'₇

Inherit(T₆, T'₇):

T₆: [C-commanded-by: 2]

Expect for VP₉

Share(VP₉, T₆):

VP₉: [C-commanded-by: 2]

Pass(VP₉, T₆):

VP₉: [C-commanded-by: 2, 6]

T₆: [C-commanded-by: 2, 9]

INPUT: *blaming*

Expand *blaming* to V₈

Attach V₈ as a daughter of VP₉

Inherit(V₈, VP₉):

V₈: [C-commanded-by: 2, 6]

Expect for DP₁₀

Share(DP₁₀, V₈):

DP₁₀: [C-commanded-by: 2, 6]

Pass(DP₁₀, V₈):

DP₁₀: [C-commanded-by: 2, 6, 8]

V₈: [C-commanded-by: 2, 6, 10]

INPUT: *himself*

Expand *himself* to N₁₁⁵

Attach N₁₁ as DP₁₀'s daughter

Inherit(N₁₁, DP₁₀):

N₁₁: [C-commanded-by: 2, 6, 8]

INPUT: Done.

As we can see at the end of the calculation, *himself* has the correct feature, i.e., [C-commanded-by: 2, 6, 8]. Importantly, the left branching NP *boy*, which the spine-based approach predicts not to be c-commanded by T'₇ (as its parent spine index is not the same to that of T'₇), now bears the right feature, i.e., [C-commanded-by: 7, 1].

It is clearly shown in the steps of the computation of the [C-commanded-by] features in (31) that features are assigned and computed step by step, and that the steps are quite evenly distributed in the parsing of each word. Words that are parsed later in the sentence do not bring in more retrieval steps or more parsing steps, contrary to the incorrect predictions generated by previous approaches to the computation of the [C-command] feature.

⁵The structure of *himself* is much more complicated than an N, but since this is only for an illustration of a left-corner parser that integrates Inherit, Pass and Share, I simplify the representational issue here.

5.5 Simplifying the encoding system

Let us now address the problems of the [C-commanded-by] approach raised by Alcocer & Phillips (2012). Recall that one problem is that the value slot [C-commanded-by] may contain a list of chunk IDs. This is contrary to an assumption of ACT-R: No such lists should be used as a slot value. Nevertheless, this may not be a serious issue because we will not use the list as a target of retrieval. Instead, we use the chunk IDs in the list only as separate retrieval cues. For instance, if a reflexive bears a [C-commanded-by: 2, 6, 8] feature (which is an abbreviation of [[C-commanded-by: [ID: 2, ID: 6, ID: 8]], then the retrieval cues should include: [ID: 2], [ID: 6], and [ID: 8].

We mentioned that the other problems identified by Alcocer & Phillips (2012) are caused by a single issue closely related to the problem discussed above; that is, the value of [C-commanded-by] is a list of chunk IDs and the list may be very long. The [C-commanded-by] value may contain too much information for the buffers, if a chunk with such a value is to be held in a buffer. In addition, when the [C-commanded-by] value of a reflexive is used as retrieval cues, all the chunks c-commanding the reflexive will be matched. Imagine that the minimal TP clause where the reflexive is located is very long, and that there are many chunks that c-command the reflexive: all the c-commanding chunks will be activated by the [C-commanded-by] retrieval cues. This causes partial match against many irrelevant chunks that are not the antecedent, which reduces the association strength between the real target and the retrieval cues. This issue exists in the spine-based approach, and it remains unresolved in my system.

A possible solution to this issue is to simplify the system by keeping track of IDs of specific, relevant chunks, and ignoring the rest of the chunks. In other words, we want to consider only the possible relevant antecedents only in the encoding process of the [C-commanded-by] feature. In order to see if such a simplification is possible, we need to know what chunks are possible to be c-commanders. It seems the c-command

relation has been argued to be relevant in the following domains (with examples in parentheses in italics), to the best of my knowledge.

- (32) a. Binding Condition A (*Jane_k likes herself_k*): a DP/QP (quantified phrase)/wh-phrase c-commands a reflexive.
- b. Binding Condition B (*Jane_k likes her_{*k/j}*): a DP/QP/wh-phrase cannot c-command and co-refer with a pronoun.
- c. Binding Condition C (*He_{k/*j} likes John_j*): a DP/QP/wh-phrase c-commands a referential expression.
- d. Bound pronoun (*Everybody_k likes his_k mother*): a QPs/wh-phrase c-commands a possessive pronoun.
- e. NPI licensing (*No teacher will ever cancel an important exam*): an NPI is licensed by a c-commanding negator, a conditional, a modal or a question operator.
- f. Proper Binding Condition (*I urged Bill to find out [who_k Mary saw t_k]*): the higher copy of a mover must c-command all its lower copies.

The list in (32) shows the cases where the c-command relations are claimed to be crucial for the identification of the relevant syntactic objects.⁶ If we can get a collection of all the possible c-commanders in languages, then [C-commanded-by] features can include only these possible c-commanders. Consequently, the system can be significantly simplified. All the chunk IDs of irrelevant syntactic objects will be ignored and not be part of the [C-commanded-by] features, for example, of a reflexive.

For the Binding Conditions A, B, C and pronoun binding, the categories of c-commanders are limited to DPs, QPs and wh-phrases. NPI (Negative Polarity Item)

⁶The relevance of c-command in all these cases are recently challenged (e.g., Barker 2012 and Bruening 2014). However, my purpose here is to find all possible cases where c-command maybe relevant. As I mentioned in Chapter III, c-command can be potentially excluded from syntax because some important syntactic relations that people previously claimed to be established by c-command now can be derived by MS-agree. Nevertheless, I still think c-command is important in parsing, as it provides a good approximation of the syntactic relations established by MS-agree.

licensing is more complicated, because NPIs can be licensed in a broad set of conditions, including negative expressions (negators), conditionals, questions and other downward entailment contexts (Ladusaw 1980) and non-veridical environments (Giannakidou 1999). Much effort has been dedicated towards attaining a unified theory of NPI licensing, suggesting that NPI licensing is more complicated than item-to-item relations (Giannakidou 1999; Israel 2004; Chierchia 2006; Giannakidou 2011). If this is on the right track, then NPI licensing may be excluded from the c-command feature computation.

The Proper Binding Condition states that traces must be bound (Fiengo 1977). That is, any trace that is left by a movement must be bound by the mover. This has a broad data coverage because it actually applies to most (arguably all) instances of movement, including topicalization, clefting, wh-movement, quantifier raising and scrambling (in Japanese), etc. The situation is similar to bound pronouns, except that the movers are distributed to a wider range of categories.⁷ For instance, many categories can undergo topicalization, making it unpredictable what c-commanders should be kept track of. There are two arguments supporting the idea that the c-command relation may not be computed concerning the phenomena covered under the Proper Binding Condition. First, these constructions are roughly identical to filler-gap constructions, and as Phillips et al. (2011) point out, the mover as a filler can be kept in the memory and be inserted into a coming gap as soon as possible. If Phillips et al.'s hypothesis is right, then c-command is not necessary for the identification of the filler as the antecedent of the trace (= gap) (see Stabler 2013 and Hunter 2018 for left-corner parsers that do not resort to c-command for filler-gap constructions). Furthermore, many authors have argued that Proper Binding Condition is dispensable as an independent principle because it can be derived from other constraints on derivation or externalization, without resorting to c-command (see e.g., Collins 1994; Cecchetto 2001; Hiraiwa 2003; Takita 2009). If these authors are on the right track,

then we do not need to worry about the Proper Binding Condition either.

To sum up, I have listed cases where c-command is plausibly involved, and found that the potential c-commanders come from a limited range of categories. Therefore, it is reasonable to assume that the parser only keeps track of this limited range of categories as c-commanders, namely DPs, QPs and wh-phrases, ignoring all other categories. If this simplified approach is adopted in parsing, then the Pass (X, Y) operation is activated only for these limited number of arguments: DPs, QPs and wh-phrases. That is, Pass (X, Y) in (26b) copies Y's ID to X's [C-commanded-by] feature only when Y is a DP, QP or wh-phrase, and X's ID is copied to Y's [C-commanded-by] feature only when Y is a DP, QP or wh-phrase. With this simplification, the value of the [C-commanded-by] feature on the reflexive *himself* in example (30) ends up as only [C-commanded-by: 2], which does not lead to any problem of extensive partial matches.

5.6 Extending to the locality feature

In this section, I apply the encoding mechanism developed in the previous sections to another restriction on reflexive binding, namely the locality constraint. Recall that in Chomskyan classical binding theory, in order to have an antecedent bind a reflexive, the antecedent of a reflexive must c-command the reflexive in the governing category of the reflexive (roughly equal to a minimal DP or TP with a subject). That is, the minimal DP or TP (with a subject) that dominates the reflexive must also be the minimal DP or TP dominating its antecedent. This is the locality requirement on reflexive binding.

⁷See Barker (2012) for arguments suggesting that c-command is not required for a pronoun to co-vary with a quantified phrase. Nevertheless, experimental evidence show that the c-command relation does matter in the interpretation of the bound pronouns (Cunnings et al. 2015; Kush et al. 2015; Moulton & Han 2018). What I want to note here is that if c-command is relevant for bound pronouns (for most cases this does seem right), the types of possible c-commanders for a bound pronoun are very restricted.

In our formalization of reflexive binding with MS-agree, we have reduced this locality requirement to phase theory, by assuming that DP (with a subject) and *v*P are phases, and that minimal search for the antecedent of the reflexive is triggered when the next phase above the reflexive is built. Such a phase-based account of the locality constraint on reflexive binding has been discussed and supported in many previous studies, including Lee-Schoenfeld (2004, 2008); Quicoli (2008); Despić (2015) and Charnavel & Sportiche (2016), among others.

In parsing, in order to successfully retrieve a correct antecedent for the reflexive in reflexive binding, the parser needs to retrieve a local antecedent. However, as pointed out by Dillon et al. (2014), the current cue-based retrieval theory cannot encode this syntactic relation as a feature. I argue that the algorithm I developed for derivational c-commanded-by can derive the locality requirement on reflexive binding, without resorting to any additional mechanisms. Specifically, I will demonstrate in this section that Inheritance, Passing and Sharing can be used for the encoding of the locality constraint as a feature.

Let us use a structure with two TPs (33) as an example.

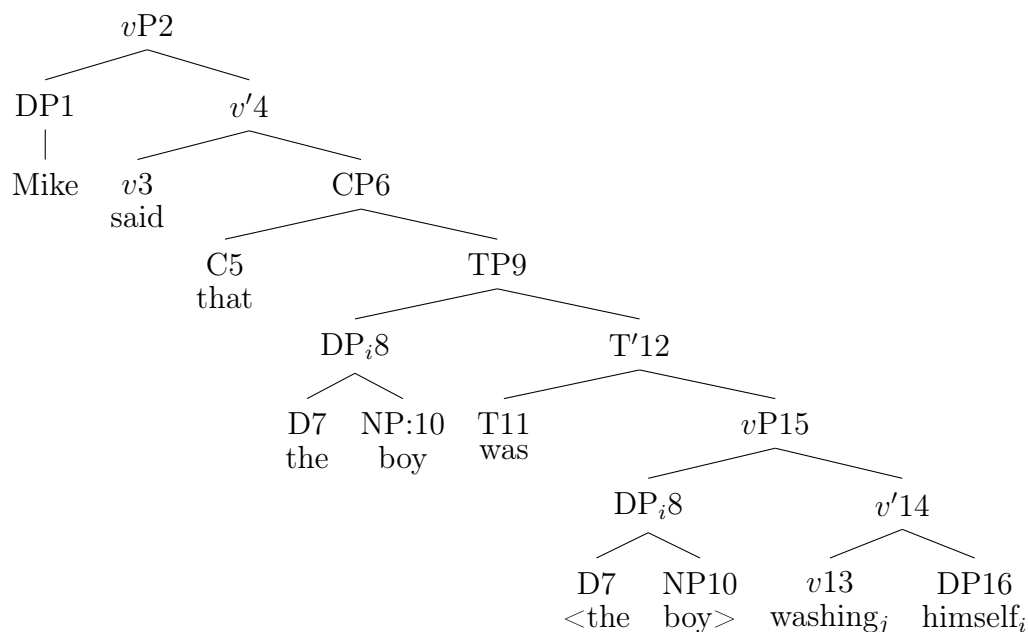
- (33) [TP₁ Mike [_vP <Mike> said that [TP₂ the boy was [_vP <the boy> washing himself]]]].

In the matrix clause TP₁, *Mike* c-commands the reflexive *himself*, but *Mike* cannot be the antecedent of the reflexive because *Mike* is not within the next phase above the reflexive (the lower *v*P). Instead, only *the boy* is. How can we capture this locality requirement and prevent *Mike* from being a legal antecedent in the cue-based retrieval system?

The leading idea is that before the Inheritance, Passing and Sharing of the [Local] feature apply, the label of the node at hand is checked. This checking procedure has been used to restrict these operations to a subset of syntactic categories in the computation of the [C-commanded-by] feature, as was discussed in the last section.

For the [Local] feature, the phasal categories (e.g., CP and *v*P) cue a new local domain. Therefore, whenever a phasal category is parsed, the parser clears the [Local] feature, deletes its content, and initiates its value to empty. I call this operation *Clear()*. The simplified structure of the example sentence (33) is represented in (34),⁸ and the steps of the computation of the [Local] feature are shown in (35). To simplify the discussion, I include only the main steps, specifically those involving the Inheritance, Passing and Sharing operations.

(34)



- (35)
- a. $\text{Pass}(\text{DP1}, v'4) \longrightarrow v'4$: [Local: 1]
 - b. $\text{Inherit}(v3, v'4) \longrightarrow v3$: [Local: 1]
 - c. $\text{Share}(\text{CP6}, v3) \longrightarrow \text{CP6}$: [Local: 1]
 - d. **Clear**(CP6) $\longrightarrow \text{CP6}$: [Local:]
 - e. $\text{Inherit}(\text{C5}, \text{CP6}) \longrightarrow \text{C5}$: [Local:]
 - f. $\text{Share}(\text{TP9}, \text{C5}) \longrightarrow \text{TP9}$: [Local:]

⁸In this structure, I delete the higher TP and CP, and build the structure only up to the matrix *v*P. I also put the lower copy of *the boy* in pointed parentheses but encode other less relevant copies with a *t* and an index.

- (Store DP8 as the filler)
- g. $\text{Inherit}(\text{DP8}, \text{TP9}) \longrightarrow \text{DP8: [Local:]}$
 - h. $\text{Share}(\text{T}'12, \text{DP8}) \longrightarrow \text{T}'12: [\text{Local: }]$
 - i. $\text{Pass}(\text{T}'12, \text{DP8}) \longrightarrow \text{T}'12: [\text{Local: } 8]$
 - j. $\text{Inherit}(\text{T11}, \text{T}'12) \longrightarrow \text{T11: [Local: } 8]$
- (Identify the trace and fill in DP8)
- k. $\text{Share}(v\text{P15}, \text{T11}) \longrightarrow v\text{P15: [Local: } 8]$
 - l. **Clear** $(v\text{P15}) \longrightarrow v\text{P15: [Local: }]$
 - m. $\text{Inherit}(\text{DP8}, v\text{P15}) \longrightarrow \text{DP8: [Local: }]$
 - n. $\text{Share}(v'14, \text{DP8}) \longrightarrow v'14: [\text{Local: }]$
 - o. $\text{Pass}(v'14, \text{DP8}) \longrightarrow \text{T}'14: [\text{Local: } 8]$
 - p. $\text{Inherit}(v13, v'14) \longrightarrow v13: [\text{Local: } 8]$
 - q. $\text{Share}(\text{DP16}, v13) \longrightarrow \text{DP16: [Local: } 8]$

Note that the simplified approach has been applied as suggested in the previous section for the computation of the c-command relation. That is, we consider only DPs/QP/wh-phrases for the [Local] feature. As the last step, Step q, indicates, this computation gives us the correct result: Only the ID of the DP chunk *the boy* is in the local binding domain of the reflexive, i.e., the $v\text{P15}$ phase.

In sum, this chapter has developed a method to encode syntactic relations such as [C-commanded-by] and [Local] as features, couched in a revised version of the derivational approach to syntactic relations. The content of the features for syntactic relations, [C-commanded-by] and [Locality] in particular, will be used as retrieval cues in the modeling of reflexive binding under the cue-based retrieval theory, as is assumed in the next chapter.

CHAPTER VI

Experiments

6.1 Introduction

Chapter I has emphasized that illusions of grammaticality may reveal how cue-based retrieval works. In addition, illusions of grammaticality/facilitatory interference effects occur when competent language users fail to detect ungrammaticality (Clifton et al. 1999; Pearlmutter et al. 1999; Wagers et al. 2009; Phillips et al. 2011). That is, a reader/listener may perceive a sentence as well-formed despite knowing upon conscious reflection that it is an ungrammatical sentence. For instance, the following sentence in (1) sounds acceptable to some comprehenders although it breaks the subject–verb agreement relation.

- (1) [DP₁ The new **executive** who oversaw [DP₂ the middle **managers**]] apparently **were** dishonest about the company's profits.

An interesting explanation of illusions of grammaticality is provided by the cue-based retrieval theory (Lewis & Vasishth 2005; Wagers et al. 2009). The idea is that when the T head *were* is encountered, the parser needs to retrieve the correct subject from the memory system, by using retrieval cues to match against the chunks stored in the working memory. The retrieval cues in this case include [Num: PL, Cat: DP, SynP: SpecTP] (Cat = Category, SynP = SyntacticPosition), since the correct subject must

be a plural DP and it must be in the specifier position of the T head/TP. The chunk that matches more retrieval cues will receive more activation boost, everything else being equal. Then the chunk with the highest activation level is most likely to be retrieved. In (1), both the DP1 and DP2 partially match two of the retrieval cues. DP1 matches [Cat: DP, SynP: SpecTP], and DP2 [Num: PL, Cat: DP]. Therefore, DP1 and DP2 may gain similar amount of activation boost. In a stochastic, noisy memory retrieval system, misretrieval of the non-subject DP2 happens, which fools the comprehender who may judge the sentence as grammatical, leading to a speed-up in the reading measures, which is called facilitatory effects.

Previous studies find a crucial processing difference between subject–verb agreement and reflexive binding: Facilitatory effects are detected in subject–verb agreement cases such as that in (1); however, no grammatical illusions are detected in reflexive binding. This motivates an examination of the representational differences between these two types of constructions, as I have done in Chapter II to Chapter IV, in the hope of providing an explanation for the processing difference.

In Chapter II and III, I have argued that subject–verb agreement and reflexive binding are syntactically similar in the sense that the syntactic dependency in these two cases are both instances of minimal search-based agree (MS-agree). That is, these two types of dependency are both established through minimal search. Consequently, although in parsing, I continue to use different syntactic retrieval cues for the memory retrieval of the subject or antecedent in subject–verb agreement and reflexive binding, the nature of the syntactic dependencies are the same. In other words, although I use [SynP: SpecTP] to retrieve the subject in subject–verb agreement, whereas use [C-commanded-by] to retrieve the antecedent in reflexive binding, it must be noted that the syntactic relations involved are not different in nature. Therefore, I do not expect that they function very differently in sentence processing.

In Chapter IV, I claim that the ϕ -features on bound pronouns and reflexives are

semantically interpretable, whereas those on Ts or verbs are not. Taking the three chapters together, we find that subject–verb agreement and reflexive binding form an interesting comparison. We may expect that these two constructions, although syntactically similar, may be processed differently because the nature of the ϕ -features involved differ.

The goal of this chapter is to construct an explanation regarding the processing difference based on the semantic interpretability of the ϕ -features involved, and then report experimental results that may help further evaluate the explanation. However, before we work on such an explanation, we need to clean our ground first. While researchers agree that facilitatory effects are found in subject–verb agreement, previous studies in fact render controversial and mixed results regarding facilitatory effects in reflexive binding (e.g., Dillon et al. 2013 versus Jäger et al. 2019). In what follows, a review of previous experiments is provided to evaluate whether facilitatory effects exist in reflexive binding. Such a review would also help us get a better understanding of the conditions under which facilitatory effects occur.

6.1.1 Literature review

6.1.1.1 Sometimes facilitatory effects do not occur

The results concerning the facilitatory effects in the processing of reflexive binding are mixed. Studies such as Sturt (2003) (Experiment 2) and Dillon et al. (2013), using eye-tracking methods, do not find facilitatory effects in either the early reading time measures (including first-pass and first-fixation) or the later reading time measures (second-pass and regression) on both the reflexive position and the spill-over positions. Specifically, Dillon et al.’s (2013) eye-tracking experiments find that the critical position (the reflexive *themselves* and the preposition *on*) and the following three spill-over positions in the target-mismatch, distractor-match condition exemplified in (2a) are not read faster (mainly in the total fixation times) than the corresponding

positions in target-mismatch, distractor-mismatch condition exemplified in (2b).

(2) *No facilitatory effects with reflexive binding*

a. *Target-mismatch, distractor-match*

The new executive who oversaw **the middle managers** apparently doubted themselves on most major decisions.

b. *Target-mismatch, distractor-mismatch*

The new executive who oversaw **the middle manager** apparently doubted themselves on most major decisions.

Similarly, in a cross-modal priming study, Nicol & Swinney (1989) do not detect priming effects of the grammatically inaccessible antecedent right after the reflexive is read, which supports that the syntactically inaccessible distractor is not retrieved in the processing of the reflexive.

Based on this string of studies, Phillips et al. (2011) suggest that “argument reflexives are immune to interference from structurally inaccessible antecedents because antecedents are retrieved using only structural cues.” Phillips et al. (2011) consider the c-command constraint a syntactic filter even before cue-based retrieval is triggered. Such a syntactic filter approach is recently supported by studies on reflexive binding (Andrews et al. 2016) and those on bound pronouns (Kush et al. 2015).

6.1.1.2 Sometimes there is instead a late signal of inhibitory interference

Some other studies, although they also did not observe interference in the early measures, found signals or trends of late interference effects that sometimes did not reach statistical significance. What is unexpected is that when the distractor matches the retrieval cues, and the interference effects are usually inhibitory rather than facilitatory, leading to an increase of processing cost. Evidence for such an increase of processing cost comes from Sturt’s (2003) first experiment and its follow-up study.

In his first experiment, Sturt (2003) adds a context (3) before the presentation of a critical test sentence (4) to highlight the distractor by putting it in the discourse focus.

(3) *Context*

{Jennifer / Jonathan} was pretty worried at the City Hospital.

(4) *Test sentence*

{She / he} remembered that the surgeon had pricked {herself / himself} with a used syringe needle. There should be an investigation soon.

A pronoun is used in the subject position of the test sentence to refer back to the topic in the context. Note that although the distractor c-commands the reflexive, it does not share the same minimal binding domain/phase with the reflexive. The locality requirement of reflexive binding excludes the long-distance distractor as a syntactically possible antecedent (Chomsky 1981, 1986, see also Chapter III).

Sturt (2003) indicates that when the gender of the reflexive matches the gender of the distractor, the participants slow down in the two spill-over positions with late measures, including the second-pass and regression reading time. In other words, for both the target-match and target-mismatch conditions, whenever the distractor matches the reflexive, participants experience a slow-down. However, these interference effects are not observed for the early reading time measures. Sturt (2003) thus concludes that the distractor is not considered in the first stage but interferes with the target in a later stage. These results are opposite to the prediction of Lewis & Vasishth's (2005) (LV05) computational cue-based retrieval model (see Patil et al. 2016b for a detailed comparison between Sturt's results and the predictions of LV05's cue-based retrieval model).

Sturt (2003) then does a follow-up study, where a sentence-by-sentence self-paced reading task is followed by a question that targets the interpretation of the reflexive.

The results show that when the distractor matches the retrieval cues, it can be misretrieved as the antecedent of the reflexive, with 17% of misretrieval for target-match condition, and 31% of misretrieval for target-mismatch condition. These results further suggest that the distractor does sometimes lead to interference and misretrieval. However, it is still unclear why misretrieval have led to inhibitory rather than facilitatory effects, if misretrieval is underlying the late interference effects observed in Sturt's first experiment.

Note that the interference effects in Sturt's (2003) first experiment disappear in his second experiment, where the distractor is no longer in a c-commanding, subject position of the whole sentence, but is embedded in a relative clause, similar to the examples shown in (2). The results from the two experiments in Sturt (2003), if taken together, seem to suggest that a syntactically inaccessible distractor matching the retrieval cues can cause late interference effects when it is in a syntactically prominent position, e.g., the subject position.

Partially resonating Sturt's (2003) results, Xiang et al.'s (2009) ERP study also does not find early effects of interference in reflexive binding, but they observed a strong trend in late interference effects, with a lasting greater positivity during 800-1000ms interval after the presentation of the critical word, which suggests an increase of processing cost in the later stage of processing. Furthermore, it also seems possible that the distractor may have been brought into consideration in an early stage, but interference effects emerge only later. More recently, Clackson et al. (2011) employ the eye-tracking method with an audio presentation of test sentences exemplified in (5), coupled with images associated with the target and the distractor.

- (5) {Peter / Susan} was waiting outside the corner shop. {He / She} watched as Mr. Jones bought a huge box of popcorn for himself over the counter.

Clackson et al. (2011) find that adults are not distracted by the image associated with the distractor when the distractor matches the gender of the reflexive (e.g., when the

distractor is *Peter* in (5)), except a brief rise in looks to the distractor in early stage of processing, i.e., between 200 and 500 ms, although this rise does not reach statistical significance. Taken together, it seems distractor may be brought into consideration in the early stage, where the interference effects, if there are any, emerge in the later stage.

It is also worth mentioning that in Clackson et al. (2011) the data from the pronoun condition with Condition B violation, where *himself* in (5) is replaced by *him*, clearly shows that adult participants are affected by the syntactically inaccessible distractor, i.e., the local DP *Mr. Jones* in (5), although *Mr. Jones* can in principle be excluded by a syntactic constraint on pronouns such as the Binding Condition B (Chomsky 1981).¹ Therefore, Clackson et al. (2011) reject the idea that the syntactic constraints filter out the distractor in the first stage processing.

6.1.1.3 Sometimes facilitatory effects do occur

A third camp of studies supports that interference effects exist for reflexive binding (Badecker & Straub 2002; Chen et al. 2012; Clackson & Heyer 2014; Cunnings & Felser 2013; King et al. 2012; Patil et al. 2016a; Jäger et al. 2019). For example, Chen et al. (2012) argue that there is no priori reason to believe that the parser should give priority to syntactic cues such as c-command over other information, contrary to what has been proposed by Phillips et al. (2011). The question is that if other information is available, why does the parser intentionally ignore them? Chen et al. (2012) further point out that previous studies take null results as indications of no effects, but an equally possible alternative hypothesis is that the previous studies may simply not have enough power to detect the effects. Therefore, Chen et al. (2012) recruit a larger sample (120 participants) for their experiments. Chen et al. (2012) take advantage of the special syntactic characteristics of simple reflexive *ziji* in Mandarin Chinese,

¹Also see Choy & Thompson 2010, where more interference from the distractor are found in the pronoun condition than in the reflexive condition.

which allows either a local or a long-distance c-commanding antecedent. In a self-paced reading task accompanied with a comprehension question after each sentence, exemplified in the Mandarin example (6), Chen et al. (2012) find that in (6a), the percentage of mis-interpretation of the sentence is higher with an animate distractor than with an inanimate one (*kanyizhe* ‘protester’ versus *kanyi* ‘protest’), although both of the distractors are at a non-c-commanding position because they are located inside the adjunct PP and thus are syntactically inaccessible to the antecedent of the reflexive.

- (6) a. **Fanduipai** lingxiu biao shi [_{CP} zhe-ge shengming [_{PP} zai {**kangyi** / opposition leader said this-CL announcement at {protest **kangyizhe**} shikong de shihou] gaojie-le **ziji** de dangyuan] protester} out-of-control DE time warn-ASP self DE party-member
 ‘The opposition leader said that this announcement warned his party members when the {protest/protesters} was out of control.’
- b. Zhe-ge shengming biao shi [_{CP} **fanduipai** lingxiu [_{PP} zai {**kangyi** / this-CL announcement said opposition leader at {protest **kangyizhe**} shikong de shihou] gaojie-le **ziji** de dangyuan] protester} out-of-control DE time warn-ASP *ziji* DE party-member
 ‘This announcement said that the opposition leader warned his party members when the {protest/protesters} was out of control.’

These results indicate that more misretrieval of the distractor results when the distractor is animate, which may be because animate noun phrases are more prominent (Barker et al. 2001), or *ziji* somehow prefers to take an animate antecedent. On the contrary, no such misretrieval difference is found for the local dependency condition in (6b), where the target is no longer a long-distance antecedent but a local one. The target is in the same local CP where the reflexive is located. Taken together, Chen et al.’s results again remind us that the relative prominence of the target versus the distractor is an important factor that might have contributed to the (non-)occurrence of misretrieval and interference effects.

Jäger et al. (2019) conduct a large-scale replication of Dillon et al.’s (2013) original eye-tracking study, with 181 subjects included in the analysis. The results confirm that, with the total fixation time on the critical position and a following spill-over position, facilitatory effects are detected in both reflexive binding and subject–verb agreement. Jäger et al. (2019) argue that Dillon et al.’s null finding regarding facilitatory effects in reflexive binding is a result of low power. However, Jäger et al. (2019) also find that the first-pass regression fixation time does show a difference between subject–verb agreement and reflexive binding, as facilitatory effects are only observed for the former. Jäger et al. (2019) interpret such results as a possible indication of the less availability of the non-commanding distractor in the early stage of processing, resonant with Sturt (2003).

Patil et al. (2016a) employ slightly different materials to evaluate the hypothesis whether syntactic constraints function as an initial filter in processing. The materials are similar to what we have seen in Sturt (2003) (Experiment 2) and Dillon et al. (2013), except that the distractor is put in the subject position of the relative clause. The distractor is thus in a more prominent syntactic position and has more features overlap with the target, as the target is also a subject. Examples of the two interference conditions are shown below.

- (7) a. The tough soldier that {Fred / Katie} treated in the military hospital introduced himself to all the nurses.
- b. The tough soldier that {Fred / Katie} treated in the military hospital introduced herself to all the nurses.

Patil et al.’s (2016a) results indicate that no early interference effects are detected for the target-mismatch, distractor-match condition (i.e., ...Katie ...himself ...in (7a)), but there are some signals of later facilitatory interference effects with the re-reading probability measure, although this effect was only marginal ($p = .063$). Interestingly, based on the cumulative progression data, which quantifies the dis-

tance (in pixels) a reader travels from the area of interest, Patil et al. (2016a) find that readers speed up after entering into the reflexive region the first time in the target-mismatch, distractor-match condition, compared to their reading the target-mismatch, distractor-mismatch condition. This is consistent with the prediction of the LV05’s cue-based retrieval model regarding facilitatory effects.

Dan Parker and colleague’s study of facilitatory effects render clearer evidence for the existence of facilitatory effects in reflexive binding (Parker et al. 2015; Parker & Phillips 2017). The studies successfully switch on and off facilitatory effects by making minimal changes on the testing sentences. Parker & Phillips (2017) robustly elicit facilitatory effects with conditions where the target mismatches two retrieval cues, based on Sturt’s (2003) materials. Nevertheless, when in 1-feature mismatch conditions, no facilitatory effects are detected in any region with any measure; in addition, no interaction of grammaticality within distractor-match is observed. Examples of 1-feature (gender) and 2-feature mismatch (gender and animacy) are shown in (8a) and (8b), respectively.

(8) a. *1-feature mismatch*

The strict {librarian / father} said that the studious schoolboy reminded herself about the overdue book.

b. *2-feature mismatch*

The strict {librarian / father} said that the brief memo reminded herself about the overdue book.

Similar results are replicated with other 2-feature mismatch patterns, including number and animacy, number and gender, with the interaction between grammaticality and distractor match also detected in an early measure, namely, the first-pass reading times, in addition to later measures. Parker and colleagues’ study therefore confirms that the facilitatory interference effects with 1-feature mismatch design are not able to achieve statistical significance and consequently not detectable. In other

words, the interference effects, although in principle may be there, are too weak to be observable with certain methods. This explains why facilitatory effects are not present in most of previous studies, as all of them use a 1-feature mismatch design.²

To summarize, subject–verb agreement is subject to facilitatory interference, which can be explained by the LV05’s cue-based retrieval model. However, previous studies obtained controversial results regarding the interference effects in reflexive binding. A set of studies do not find any interference effects, and another set of studies find that sometimes the inhibitory interference effects instead emerge in late reading time measures, and still another set of studies point out that interference effects can be detected in both the early and late measures. Finally, Parker and colleagues’ study sheds light on the nature of the debate: The interference effects in reflexive binding are too weak with a 1-feature mismatch design, and thus are not observable most of times, if not always. The conclusion is consistent with Cummings & Felser’s (2013) observation that interference effects are found only among participants who have a low memory capacity but not those with a high memory capacity. This is because both suggest that facilitatory interference happens only under certain specific conditions and is not prevented in a principle way. However, these studies are not compatible with an approach that takes syntactic relations such as the c-command relation as filters to prevent syntactic inaccessible distractors from being retrieved.

6.1.1.4 A meta-analysis

To settle down the debate, a meta-analysis on the processing of subject–verb agreement and reflexive binding is conducted recently by Jäger et al. (2017). Jäger et al. (2017) do a Bayesian meta-analysis to synthesize and evaluate empirical evidence for LV05’s computational model. The analysis is based on 28 experiments (some of them

²A potential complication of the 2-feature mismatch design is that the local subject is not possible at all to be an antecedent of the reflexive. This may trigger the logophoric reading of the reflexive (Reinhart & Reuland 1991, 1993; Charnavel 2018a), which allows the reflexive to take a long-distance antecedent (Sloggett & Dillon 2017).

are experiments from the same studies) on subject–verb agreement, and 21 experiments on reflexive/reciprocal binding. Only the mean reading times on the critical region, including first-pass fixation for eye-tracking experiments and the reading time on the reflexive region from other reading tasks such as self-paced reading tasks, are included in the model. The meta-analysis results show that, for subject–verb agreement, with target-mismatch conditions, there is strong evidence of facilitatory effects, in line with the prediction of LV05. However, for reflexive/reciprocal binding, there shows (moderate) evidence of inhibitory interference with target-mismatch conditions, not consistent with LV05. But as previously mentioned, the meta-analysis is based on either the first-pass reading time from eye-tracking experiments or the reading time on the reflexive region from other reading tasks; the results therefore may just indicate that in the early stage, inhibitory instead of facilitatory effects are observed in reflexive binding, which leaves open the question as whether the same inhibitory interference is observed in the later stage.

Jäger et al.’s (2017) meta-analysis thus reveals an inhibitory interference in reflexive binding under the target-mismatch conditions. This is surprising given that we have seen many studies that found null results and even facilitatory interference effects. Jäger et al. (2017) acknowledge this discrepancy between their results of qualitative and quantitative analyses and hypothesize that one of the reasons could be that it is misleading to draw conclusions based on multiple studies using statistical significance as the criterion, especially when the power of the studies is relatively weak, for example, when the sample size is small.

Bringing the qualitative review of the previous studies and Jäger et al.’s (2017) meta-analysis together, we find that the followings are correctly predicted by LV05: (i) Facilitatory effects are observed for subject–verb agreement under the target-mismatch, distractor-match condition; (ii) facilitatory effects are observed in reflexive binding under the target-mismatch, distractor-match condition—this is true at least

under Parker & Phillips’s (2017) 2-feature mismatch conditions. However, there are also challenging results that are not well-predicted, one of which concerns us most: There was no facilitatory interference for reflexive binding under the target-mismatch, distractor-match condition with the 1-feature mismatch paradigm; instead, first-pass reading times in such conditions showed moderate inhibitory effects. The results from the studies reviewed can be summarized as below:

- (9) Facilitatory effects are less observable in reflexive binding compared to subject–verb agreement.

A question then immediately comes up: Why is memory retrieval in reflexive binding less affected by interference from a distractor? In the following section, I will propose an account that is based on the syntactic and semantic analyses of agreement and reflexive binding developed in previous chapters.

6.1.2 Accounting for the absence of facilitatory effects

How to account for the finding that facilitatory effects are less observable in reflexive binding compared to subject–verb agreement, as summarized in (9)? It has been proposed that the syntactic relation involved in reflexive binding, i.e., the c-command relation, should be weighted heavier than the syntactic relation involved in subject–verb agreement, e.g., the Spec-head relation (or the subject-hood).³ This is called the weighting approach (see more discussion of the weighting approach in Chapter I, as well as Jäger et al. 2017 and Parker et al. 2017).

I argue in Chapter II and III that the agreement relation and the binding relation can both be formalized with MS-agree. That is, no essential difference should be expected between these two syntactic relations. If the syntactic analyses of these two constructions are viable, the weighting approach loses its theoretical motivation.

³C-command and SpecTP are relevant retrieval cues here because a reflexive needs a c-command antecedent, whereas a verb/T head looks for a subject at SpecTP.

A crucial difference between subject–verb agreement and reflexive binding is highlighted in Chapter IV. The main idea is that the ϕ -features on the T head in agreement are semantically uninterpretable features, whereas the ϕ -features on the reflexive in reflexive binding are semantically interpretable. It has been argued that the ϕ -features on reflexives are interpreted as presuppositions, whereas the ϕ -features on T do not have such semantic content; instead, they are simply phonological features (usually called morphosyntactic features in the literature). Since semantic features are interpreted at the CI interface, whereas phonological features are interpreted at the SM interface, I will represent the semantic number feature as [CI-Num: (SG/PL)], and the phonological number feature [SM-Num: (SG/PL)] (cf. Chapter IV).

But why should such interpretability difference be relevant to the occurrence of facilitatory effects? Here is a hypothesis that could potentially account for the asymmetry in the memory retrieval with semantic features and phonological features as retrieval cues:

(10) *Asymmetry of Interpretability in parsing*

Mismatch of semantically interpretable retrieval cues will be more likely detected than that of phonological retrieval cues.

With the following examples in (11), (10) means that in (11a), given that the number feature on *were* is semantically uninterpretable, and thus the relevant retrieval cue is [SM-Num: PL], the target DP1 mismatches only a phonological retrieval cue.

(11) *Facilitatory effects in subject–verb agreement and reflexive binding*

a. *Target-mismatch, distractor-match*

[_{DP1} The new executive who oversaw [_{DP2} the middle managers]] apparently were dishonest about the company's profits.

b. *Target-mismatch, distractor-match*

[_{DP3} The new executive who oversaw [_{DP4} the middle managers]] appar-

ently doubted themselves on most major decisions.

However, in (11b), given that *themselves* includes both a CI-Num and an SM-Num, the relevant retrieval cues are [CI-Num: PL, SM-Num: PL]. The target DP3 mismatches not only a phonological retrieval cue, but also a semantically interpretable retrieval cue, according to the Asymmetry of Interpretability Hypothesis (10), the latter of which is more likely to be detected. As a result, grammatical deviance will be more likely detected in reflexive binding, compared to subject–verb agreement, explaining the processing difference summarized from previous studies in (9).

The intuition behind the Asymmetry of Interpretability Hypothesis is quite simple: A semantically deviance/error is more likely to be detected than a phonological deviance/error.⁴

6.1.3 Asymmetry of interpretability in cue-based retrieval

This intuitive hypothesis can be formally implemented under the context of LV05’s cue-based retrieval model. In order to demonstrate how this implementation is possible, I would need to introduce, with minimal details, how LV05 compute facilitatory and inhibitory effects mathematically. Whether a chunk will be retrieved is determined by the total activation of that chunk. And if two chunks are in competition, only the one with a higher activation level will be retrieved. The activation level of a chunk i , A_i , is determined by the equation in (12), adopted from the ACT-R architecture (Anderson & Lebiere 1998; Anderson 2005).

(12) *Total activation*

$$A_i = B_i + \sum_j W_j S_{ji} - D_{ip} + \epsilon$$

where A_i is the total activation level of a chunk i , B_i is the base activation of i , and S_{ji} represents the strengths of association from an element j to i . W_j

⁴A similar asymmetry is reported in the field of second language acquisition (Slabakova 2006, 2008, 2009).

is (attentional) weight associated with j (slot values) of the goal chunk i . D_{ip} is the partial match penalty. ϵ is the noise added to the activation value of a chunk at each retrieval.

I added the partial match penalty, D_{iP} , according to Anderson & Lebiere (1998:77). This is because the total activation level should be adjusted by degree of mismatch between the chunk and the production condition. In the case of perfect match, D_{iP} is 0. Otherwise, the D_{iP} is computed by the formula in (13).

(13) *Partial match penalty*

$$D_{ip} = \textit{penalty} * \sum_{i=1}^n (|p_i - f_i|)$$

where *penalty* is the mismatch penalty parameter, with a default = 1, p_i is the i th slot of the production condition (i.e., the i th retrieval cue), and f_i is the corresponding slot of a chunk. If one slot of p_i and f_i mismatch, the absolute value, $|p_i - f_i|$, is 1.

The equation in (12) allows for partial match, because an important term of the total activation is the sum of all matches between the feature slots of a chunk and the retrieval cues, being blind to whether this is a case of full match or partial match. Imagine a case where a partially matched distractor has a higher total activation, it will be misretrieved, leading to grammatical illusions/facilitatory effects (e.g., in the case of subject–verb agreement).

The Asymmetry of Interpretability Hypothesis can be implemented in the cue-based retrieval theory as follows:

- (14) a. Mismatch of semantic retrieval cue is costly and leads to a higher mismatch penalty (the *penalty* term in (13)), compared to that of phonological retrieval cue.

- b. Higher mismatch penalty on the target reduces the activation level of the target (see (12)).
- c. The lower activation level of the target leads to an overall higher retrieval latency of the retrieval, in both the distractor match and distractor mismatch conditions (cf. Logačev & Vasishth 2015).
- d. The lower activation level of the target in both the distractor match and distractor mismatch conditions lead to a higher possibility of longer retrieval and retrieval failure.
- e. The test sentences in both the distractor match and distractor mismatch conditions will be more likely judged as ungrammatical.
- f. Once the test sentences are judged ungrammatical, the benefit brought by the distractor in the distractor match condition is concealed.

It is important to note that an overall higher retrieval latency leads to a longer reading time, everything else being equal; however, this does not mean that the misretrieval of the distractor disappears or is reduced. In fact, if the activation level of the target is lower, the distractor is more likely to be misretrieved. Therefore, given this specific implementation of the Asymmetry of Interpretability Hypothesis as in (10), we in fact predict that the misretrieval of the distractor may still happen in the processing of reflexive binding sentences such as (11b). However, the higher mismatch penalty on the retrieval of the target brings down the activation level of the target generally, and thus increases the reading retrieval overall, leading to a higher probability of ungrammatical judgments. This may cancel out the benefits that the misretrieval of the distractor brings, as the test sentences under both the distractor match and distractor mismatch conditions are judged very ungrammatical. Consequently, in reflexive binding, we are less likely to observe facilitatory effects.

Another way to implement the Asymmetry of Interpretability Hypothesis is to assume that the parser allocates more attention resources to memory retrieval when

the retrieval involves a semantically interpretable retrieval cue. This will lead to a lower noise rate, i.e., a smaller standard deviation for the ϵ term in (12)), which will potentially decrease the misretrieval rate; that is, facilitatory effects due to the misretrieval of the distractor will be moderated.

Since the experimental results do not shed light on the distinction between the above two possible implementation of the Asymmetry of Interpretability Hypothesis, I do not explore potential distinctions between them.

Similarly, the lower activation level of the target increases the retrieval difficulty and thus may also bring down the acceptability of the sentences overall.⁵

Let us discuss a concrete example. In the retrieval of the antecedent in reflexive binding, as shown in (11b), repeated below as (15b), the retrieval of the target DP3 involves a mismatch of both the semantic retrieval cue [CI-Num: PL] and the phonological cue [SM-Num: PL], but the retrieval of the distractor DP4 does not.

(15) *Subject-verb agreement versus reflexive binding*

a. *Target-mismatch, distractor-match*

[_{DP1} The new executive who oversaw [_{DP2} the middle managers]] apparently were dishonest about the company's profits.

b. *Target-mismatch, distractor-match*

[_{DP3} The new executive who oversaw [_{DP4} the middle managers]] apparently doubted themselves on most major decisions.

By contrast, in the retrieval of the subject in agreement, as shown in (15a), the retrieval of the target DP1 involves a mismatch of the phonological retrieval cue

⁵I do not discuss another important factor that may contribute to the processing differences between subject-verb agreement and reflexive binding. For reflexive binding, the reflexive needs to retrieve the *v*P-internal subject as its antecedent, according to the syntactic analysis proposed in Chapter III. The existence of the *v*P-internal subject will increase the overall activation level of the subject/antecedent, and therefore the target will be more likely to be correctly retrieved. By contrast, in subject-verb agreement, the T head will retrieve the subject directly, without an intermediate activation of the *v*P-internals subject, and therefore the target is less likely to be retrieved compared to that in reflexive binding. I leave the possible effect of the *v*P-internal subject hypothesis to future research.

[SM-Num: PL], whereas the distractor DP2 does not. Consequently, the retrieval of the antecedent in reflexive binding in (15b) receives a higher mismatch penalty than that of the subject in a corresponding subject–verb agreement sentence (e.g., (1)), according to the first implementation in (14). Therefore, the activation level of the antecedent in reflexive binding (15b) is lower than that of the subject in subject–verb agreement (15a). According to the second implementation, if misretrieval is less likely to occur whenever semantically interpretable retrieval cues are involved, misretrieval rate in (15b) will be lower, compared to (15a). That is, facilitatory effects due to misretrieval is smaller in (15b). Therefore, no matter whether the first or the second implementation is adopted, the overall activation level of the finally retrieved antecedent in reflexive binding is lower than that of the retrieved subject in subject–verb agreement. Mapping the activation level of a chunk to the retrieval latency of that chunk, the retrieval latency associated with the antecedent in reflexive binding is bigger than that with the subject in subject–verb agreement. In general, the resulting reading time on the retrieval site in reflexive binding is predicted to be longer than that in subject–verb agreement.

Note that the Asymmetry of Interpretability Hypothesis in (10) and the implementation in (14) suggest that the latency associated with the retrieval of the antecedent in reflexive binding is bigger than that associated with the retrieval of the subject in subject–verb agreement. However, this does not take into account the misretrieval of the distractor. According to the implementation in (14), the misretrieval of the distractor in reflexive binding is predicted to be higher than the misretrieval of the distractor in subject–verb agreement, as the activation level of the target in reflexive binding is lower than that in subject–verb agreement. However, such facilitatory effects might be canceled out by the bigger retrieval latency associated with the target in the case of reflexive binding. This explains why in reflexive binding case, rather than in the subject–verb agreement, facilitatory effects have not been consistently

observed.

6.1.4 Predictions of the account

The Asymmetry of Interpretability Hypothesis in (10) predicts that when the mismatch of semantic retrieval cues is not involved in reflexive binding, grammatical illusions are more possible to occur. This motivates the using of collective nouns as testing materials. Collective nouns can help us dissociate SM-Num from CI-Num, since they bear a singular SM-Num feature and two (or ambiguous) CI-Num features, as I have argued for in Chapter IV. The feature specification of collective nouns such as *government* can be found in (16).

- (16) *government*: [SM-Num: SG, CI-Num: SG/PL]

Relevantly, the number features on non-collective nouns such as (linking) verbs and reflexives are shown below with examples.

- (17) a. *managers*: [SM-Num:PL, CI-Num:PL]
b. *were*: [SM-Num:PL]
c. *themselves*: [SM-Num:PL, CI-Num:PL]

Accordingly, with regard to different types of number features, I assume Accurate Matching in cue-based retrieval:

- (18) *Accurate Matching*

An SM-Num retrieval cue matches only an SM-Num and not a CI-Num, and vice versa.

The usage information is encoded as probabilities associated with each feature value. For instance, according to my corpus study of 109 collective nouns in American English with Google books, *government* occurs 99% of the time with a singular reflexive, and 1% of the time with a plural reflexive. The probabilities associated with the singular and plural CI-Num features on *government* are presented as below:

(19) *government*:

In reflexive binding: [CI-Num: SG (.99)/PL (.01)]

Note that the information about how frequent a collective noun co-occurs with a plural or singular T is less relevant here, given the assumption that memory retrieval triggered at the T head will not look for a [CI-Num] feature.⁶ The probability associated with certain feature will be used to weight the activation gain resulted from the match between this feature and the corresponding retrieval cue (i.e., the *W* term in (12)).

So why are collective nouns useful to test the predictions of the Asymmetry of Interpretability Hypothesis? Let us make a comparison between the retrieval configuration in examples involving a collective noun and that in examples without a collective noun. The examples and their retrieval configurations at the critical retrieval sites are shown in (20); I have underlined the most relevant retrieval cues and their matching features in the retrieval configurations:

(20) a. *Reflexive binding: collective head noun*

[_{DP1} The local government that supported [_{DP2} the tough detectives]] embarrassed themselves with a recent allegation of serious corruption.

b. *Subject-verb agreement: collective head noun*

[_{DP3} The local government that supported [_{DP4} the tough detectives]] were embarrassed with a recent allegation of serious corruption.

A comparison of the above two sentences indicates that in the case of reflexive binding (20a), the target DP1 bears a [CI-Num: PL] feature that can match the corresponding

⁶I must then explain why semantic agreement is possible in subject-verb agreement; that is, why can a T head agrees with the semantic number feature on the subject if T only bears SM number features. A possible solution is to hypothesize that all cases of collective nouns that lead to plural verbs are not rule-based but convention-based, similar to rule-based regular past tense versus convention-based irregular past tense (Pinker 1991). In other words, a collective noun must be matched by a singular verb, and the agreement can be adjusted due to other pragmatic or contextual information (Humphreys & Bock 2005). Another possibility is to capture this by activation spreading in memory retrieval (e.g., Reitter et al. 2011), assuming a model that allows activation to be spread to chunks that contain features not exactly matching but closely related to the retrieval cues.

the local government that ...	the tough detectives	themselves	retrieval cues
<u>ID: 1</u> Cat: DP SM-Num: SG CI-Num: SG (.99) <u>CI-Num: PL (.01)</u>	ID: 2 Cat: DP SM-Num: PL <u>CI-Num: PL</u>	ID: 3 Cat: DP SM-Num: PL CI-Num: PL	<u>C-commanded-by: 1</u> Local: 1 Cat: DP SM-Num: PL <u>CI-Num: PL</u>

the local government that ...	the candidates	were	retrieval cues
ID: 4 <u>SynP: TPSpec</u> Cat: DP SM-Num: SG CI-Num: SG (.99) CI-Num: PL (.01)	ID: 5 SynP: V-comp Cat: DP <u>SM-Num: PL</u> CI-Num: PL	ID: 6 Cat: T SM-Num: PL	<u>SynP: TPSpec</u> Local: 4 <u>SM-Num: PL</u>

retrieval cue. Therefore, no higher mismatch penalty associated with mismatches of semantically interpretable retrieval cues should apply here. Furthermore, given that the probability/weight of the [CI-Num: PL] feature on DP1 is very small, it will not add much associative activation boost to the target; therefore, the retrieval result could be quite similar to the case of subject–verb agreement (20b).

Also for comparison purposes, below I show in (21) the retrieval configurations for regular head noun conditions that previous studies have widely investigated.

(21) a. *Reflexive binding: regular head noun*

[_{DP5} The local lawyer that supported [_{DP6} the tough detectives]] embarrassed themselves with a recent allegation of serious corruption.

b. *Subject–verb agreement: regular head noun*

[_{DP7} The local lawyer that supported [_{DP8} the tough detectives]] were embarrassed with a recent allegation of serious corruption.

We see that the target DP5 mismatches the [CI-Num: PL] retrieval cue. According to

the local lawyer that ...	the tough detectives	themselves	retrieval cues
<div> <div>ID: 7</div> <div>Cat: DP</div> <div>SM-Num: SG</div> <div>CI-Num: SG</div> </div>	<div> <div>ID: 8</div> <div>Cat: DP</div> <div>SM-Num: PL</div> <div><u>CI-Num: PL</u></div> </div>	<div> <div>ID: 9</div> <div>Cat: DP</div> <div>SM-Num: PL</div> <div>CI-Num: PL</div> </div>	<div> <div><u>C-commanded-by: 7</u></div> <div>Local: 7</div> <div>Cat: DP</div> <div>SM-Num: PL</div> <div><u>CI-Num: PL</u></div> </div>

the local lawyer that ...	the tough detectives	were	retrieval cues
<div> <div>ID: 10</div> <div><u>SynP: TPSpec</u></div> <div>Cat: DP</div> <div>SM-Num: SG</div> <div>CI-Num: SG</div> </div>	<div> <div>ID: 11</div> <div>SynP: V-comp</div> <div>Cat: DP</div> <div><u>SM-Num: PL</u></div> <div>CI-Num: PL</div> </div>	<div> <div>ID: 12</div> <div>Cat: T</div> <div>SM-Num: PL</div> </div>	<div> <div><u>SynP: TPSpec</u></div> <div>Local: 10</div> <div><u>SM-Num: PL</u></div> </div>

the Asymmetry of Interpretability Hypothesis, this will bring a significant mismatch penalty to the retrieval of the target. Consequently, the activation level of the target is relatively low, which will further bring down the retrieval latency overall. This mismatch of semantic interpretable retrieval cues is not relevant for the subject–verb agreement in (21b), potentially explaining why facilitatory effects are only observed in subject–verb agreement.

(22) summarizes the prediction of the Asymmetry of Interpretability Hypothesis implemented in the cue-based retrieval theory.

(22) *A prediction of the Asymmetry of Interpretability Hypothesis*

Grammatical illusions/facilitatory effects are more likely to be detected in the collective head noun condition (20a) than in the regular noun condition (21a) in reflexive binding.

6.2 An overview of the experiments

Based on the Asymmetry of Interpretability Hypothesis in (10), the semantic interpretability of ϕ -feature is a crucial difference between subject–verb agreement and reflexive binding. According to the prediction of the Asymmetry of Interpretability Hypothesis in (22), if we replace the target head noun with a collective noun, then we are more likely to observe illusions of grammaticality or facilitatory effects.

6.2.1 Materials

The most crucial evidence for the Asymmetry of Interpretability Hypothesis, if there is, will be from reflexive binding. But we will also include subject–verb agreement, so that we can make a direct comparison with some previous studies which find a contrast between these two types of dependency (e.g., Dillon et al. 2013). Different from Dillon et al. (2013), the materials for reflexive binding and subject–verb agreement are tested separately in a between-subject design.

The most important test sentences are those introduced in the last section. I employ a 2 (head noun: collective, regular singular) \times 2 (distractor number: plural, singular) design. I will call the former factor HNoun, and the latter DNum. The particular names for each condition with their examples are shown below:

(23) *Reflexive binding conditions*

a. *collective head noun, plural distractor* [HC.DPL]

The local **government** that supported the tough **detectives** embarrassed **themselves** with a recent allegation of serious corruption.

b. *collective head noun, singular distractor* [HC.DSG]

The local **government** that supported the tough **detective** embarrassed **themselves** with a recent allegation of serious corruption.

c. *regular singular head noun, plural distractor* [HSG.DPL]

The local **lawyer** that supported the tough **detectives** embarrassed **themselves** with a recent allegation of serious corruption.

- d. *regular singular head noun, singular distractor* [HSG.DSG]

The local **lawyer** that supported the tough **detective** embarrassed **themselves** with a recent allegation of serious corruption.

The most important data that we are looking for are (i) the contrast between the “collective head noun, plural distractor” (HC.DPL) and the “collective head noun, singular distractor” (HC.DSG) conditions. (I also refer to these two conditions together as HC conditions.) If there are facilitatory effects in terms of reading time measures, then the reading time difference on the reflexive (and the spill-over position(s)) in the HC.DPL condition should be significantly shorter than that in the HC.DSG condition. (ii) Given that previous studies consistently find that no difference could be detected between the “regular singular head noun, plural distractor” (HSG.DPL) condition and the “regular singular head noun, singular distractor” (HSG.DSG) condition, if (i) exists, we should be able to observe an interaction between the two predictors: HNoun and DNum. That is, facilitatory effects can be measured in the following ways with reading time measures, where RT means reading times on the critical positions including the reflexive and spill-over position(s):

(24) *Measuring facilitatory effects*

- a. Facilitatory effects in the HC.DPL condition: $RT_{HC.DPL} - RT_{HC.DSG}$
- b. Facilitatory effects in the HSG.DPL condition: $RT_{HSG.DPL} - RT_{HSG.DSG}$

If we can replicate previous studies on the HSG conditions, we should find no difference on the reading time measures for (24b). By contrast, the Asymmetry of Interpretability Hypothesis predicts that facilitatory effects might be detected with a collective head noun, which means a negative number might be observed in (24a). Taken together, we have another prediction:

- (25) If facilitatory effects can be detected in HC conditions but not HSG conditions, we should expect an interaction between HNoun and DNum; namely, $(RT_{HC.DSG} - RT_{HC.DPL}) - (RT_{HSG.DSG} - RT_{HSG.DPL})$ is not equal to 0.

Similarly, with acceptability judgment data, we might be able to observe illusions of grammaticality with the HC.DPL condition but not with the HSG.DPL condition.

The following shows the test sentences for subject–verb agreement.

- (26) *Subject–verb agreement conditions*

- a. *collective head noun, plural distractor* [HC.DPL]

The local **government** that supported the tough **detectives were** embarrassed with a recent allegation of serious corruption.

- b. *collective head noun, singular distractor* [HC.DSG]

The local **government** that supported the tough **detective were** embarrassed with a recent allegation of serious corruption.

- c. *regular singular head noun, plural distractor* [HSG.DPL]

The local **lawyer** that supported the tough **detectives were** embarrassed with a recent allegation of serious corruption.

- d. *regular singular head noun, singular distractor* [HSG.DSG]

The local **lawyer** that supported the tough **detective were** embarrassed with a recent allegation of serious corruption.

Since the linking verb *were* bears only an SM-Num feature, no mismatch of semantic retrieval cues is involved. The predictions would be the same with previous studies: there will be facilitatory effects on the critical position (*were*) (and the spill-over position(s)) with (26a) compared to (26b), as well as on the critical positions in (26c) compared to that in (26d).

As for the structure of the text sentences, since previous studies have shown that spill-over positions may be important for the detection of facilitatory effects

(especially for the reflexive binding case), in the target sentences, the two words following the critical word, that is, *themselves* in reflexive binding case and *were* in subject–verb agreement case, are carefully controlled: they must be words without much content. The first word must be a preposition, and the second word is most frequently a determiner, and sometimes could be an adverb or a linking verb, but definitely cannot be a noun or a regular verb.

The animacy, frequency, orthographic neighborhood density, phonological neighborhood density, orthographic length, phonological length of the target head nouns are controlled against those of the distractors in an item basis. The frequency is determined both by Corpus of Contemporary American English (COCA) word frequency list and the frequency calculated based on CELEX database (Baayen et al. 1993). The data of orthographic neighborhood density, phonological neighborhood density, orthographic length, phonological length were taken from CLEARPOND (<http://clearpond.northwestern.edu/englishpond.php>) (Marian et al. 2012). Taking the test sentence in (26a) as an example: much effort has been made to balance the above factors of the target head noun *government* and the distractor head noun *lawyer*. The differences of these factors between the target head noun and the distractor head noun will be included as controlled factors in the statistical models.

An important reason to control these factors is to avoid a potential systematic influence from any of these factors on the base activation level of these lexical items.

I avoided using individual-level adjectives such as *happy*, *beautiful*, *angry*, *diligent* to modify the collective nouns, because these adjectives will force a plural member reading of the collective nouns. Similarly, I also avoided individual-level predicates for the relative clauses. The reason is the same as the one for the collective nouns modifier: these individual-level predicates such as *meet*, *visit* prefer the member reading of the collective nouns.

The 24 collective nouns used in the stimuli were carefully chosen. The collective nouns must agree with both singular linking verbs and singular reflexives no less than 80% of the time in the Google Books Ngram corpus. This is to ensure that at least according to the linguistic experience, the probability of the co-occurrence of the collective nouns with a plural linking verb or a plural reflexive is low, so that although the target sentences may not be ungrammatical, they are usually not very acceptable or at least not preferred.

109 collective nouns are firstly collected from previous studies (Bock et al. 2001, 2004, 2006; Kreiner et al. 2013; Levin 2001, 2006) and additional searches through Google. The 5-gram raw data where these collective nouns co-occur with singular or plural linking verbs and reflexives are extracted from Google Books Ngram through the Python client provided by the PhraseFinder API (<https://phrasefinder.io/>), with the language specified as American English. The main reason to use the Google Books Ngram corpus is that it is an extremely huge corpus (including about 155 billion words in American English). Numerous relevant instances can be easily found even for low frequent words. This is especially important for collective nouns, as many of the constructions do not occur frequently and occurrences of these constructions are difficult to find even in large corpora like COCA.

Nevertheless, the raw data include many irrelevant items, especially ones with prepositions, e.g., *...duties of the government are...* or *...members of the club are...*. These are not really instances of agreement between the collective nouns and the linking verbs. These data were subtracted by another extraction of data with the 119 collective nouns and all the prepositions that I can find ($N = 33$). Final cleaning of the data was conducted manually to exclude any other irrelevant items.

When choosing the 24 collective nouns out from the 109 collective nouns, I took into consideration also the results of previous cloze norming studies, to exclude collective nouns to which my corpus study and previous experimental studies give too

different results. In addition, relatively more frequent collective nouns were selected for the experiments.

All together the reflexive binding include 24 item sets and each item set includes a sentence from each of the four conditions. The 24 target item sets were distributed across 6 lists in a Latin Square design. For the reflexive binding experiments, the 24 target items were randomly mixed with 60 fillers: 12 fillers with singular animate reflexive binding (6 with *himself* and 6 with *herself*), 12 fillers with *themselves*, 12 fillers with *itself*, and 24 other fillers with different sentence structures but a similar length. The percentage of ungrammatical sentences in the fillers is 37%. For the subject–verb agreement, the 24 target items are randomly mixed with 36 fillers: 12 fillers with subject–verb agreement and 24 fillers with different sentence structures but a similar length. Ungrammatical sentences constitute 39% of the fillers.

To ensure that subjects pay attention to the task, half of all the sentences were followed by a comprehension question, targeting different parts of the sentences. Feedback was provided after subjects’ each response to the comprehension questions, indicating whether the response is correct or wrong.

6.2.2 Procedure

All the experiments were done online using Ibex Farm (<http://spellout.net/ibexfarm/>) with Amazon Mechanical Turk (<http://www.mturk.com>). Participants were recruited through Amazon Mechanical Turk. Workers who have signed up with the site through Amazon could search for the posted experiments and choose to participate. Participants were self-screened to be adult native speakers of American English. Only workers (i) whose IP address is located in the US, (ii) whose HIT approval rate is above or equal to 99%, (iii) who has more than 500 HITs approved before the current experiment, and (iv) who had never participated in any of the experiments presented in this chapter: they could take only one of the experiments.

Only one experiment was posted at a time.

Before participants started the experiment, they were asked to read the consent form. When workers have decided to participate, they will click on a link to the survey hosted on Ibex Farm. The tasks began by asking some questions about subject information, including age, gender, state of residence and language experience.

This study includes two types of tasks, the acceptability judgment task and the self-paced reading task. Two tasks had the same Mechanical Turk advertisement for subject recruitment, but these two tasks were done separately, and each subject did only one of them.

Each participant was paid 3.5 US dollars for the acceptability judgment tasks and 3.75 US dollars for the self-paced reading tasks (or self-paced reading + acceptability judgment task as in Experiment 6).

The judgment task normally took about 30 minutes or less, and the reading task took a little more than 30 minutes. The actual time varies, however, depending on whether the participants pay full attention to the task, as the subjects was doing the experiments unsupervised. Participants have to submit their responses within 2 hours once they start an experiment.

The details of each experiment will be given below when the experiments are introduced.

6.3 Experiment 1: judgment task with reflexive binding

The goal of this experiment is to examine one of the predictions of the Asymmetry of Interpretability Hypothesis: do illusions of grammaticality emerge with collective head nouns in reflexive binding.

6.3.1 Participants

Seventy-two subjects participated in the study. Seven subjects were excluded from data analysis because their accuracy rate for the comprehension questions are lower than 75%. The rest sixty-five subjects have a mean age of 36, ranging in age from 20 to 61. Thirty-six of them are female.

6.3.2 Stimuli

Examples of the target sentences of this experiment are the reflexive binding materials as listed in (23).

6.3.3 Procedure

The participants were presented the instruction and seven practice items for them to get familiarized with the judgment task. In the instruction, the participants were told that if they think that the sentence sounds acceptable and possible in English, then they should give it a high rating (6 or 7). If they think that the sentence does not sound like an acceptable and possible sentence in English, they should give it a low rating (1 or 2). Otherwise, if they think that some sentences may not sound like totally impossible sentences, but are also not completely acceptable, they could give those a more intermediate rating (3-5). In the practice and the test sessions, the participants were asked to silently read sentences and rate the acceptability by clicking on the scale number right below a test sentence. Other details about the procedure can be found in the Procedure section, Section 6.2.2, in the overview of the experiments.

The data were collected by two batches in Amazon Mechanical Turk in two different times. In order to capture possible differences caused by data collection time, this factor will be included in the statistical model as a control factor.

6.3.4 Results

The amount of each type of responses (from 1 to 7) are presented below in Figure 6.1.

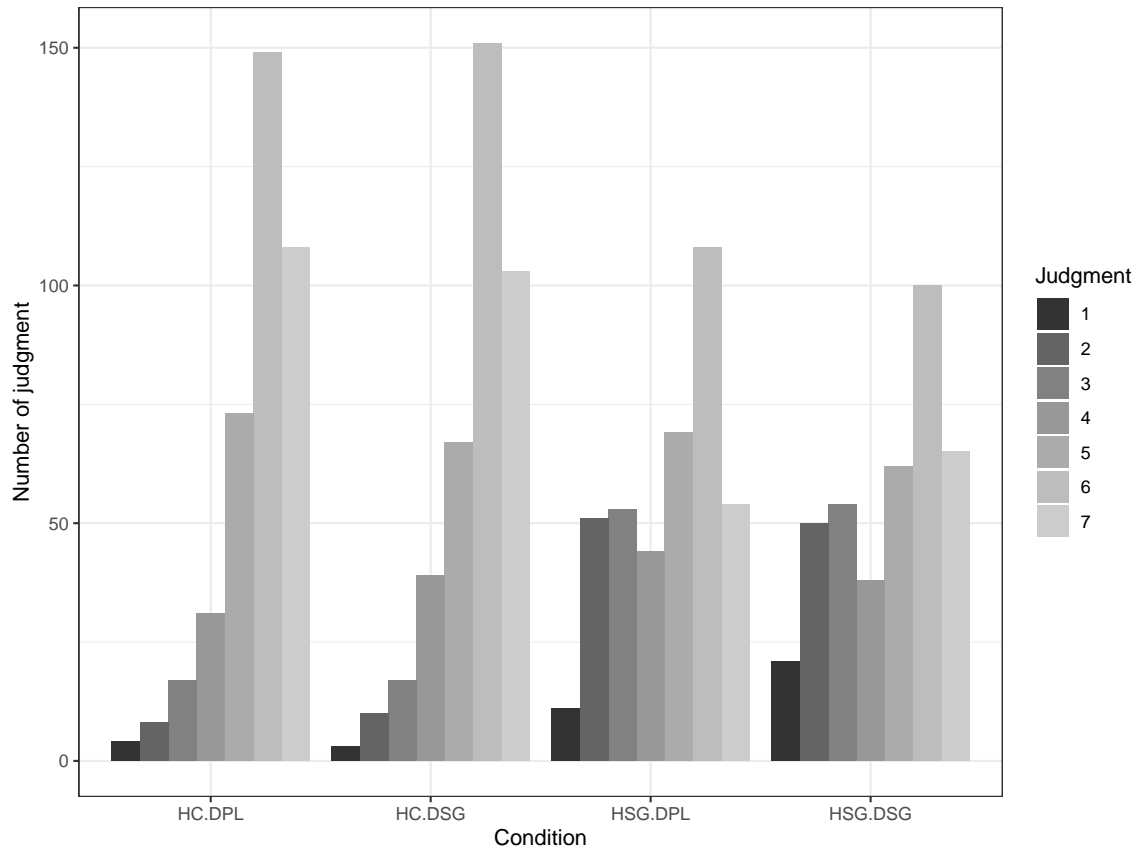


Figure 6.1: The 7 point Likert scale data from Experiment 1

The mean acceptance rates and reaction times are shown in Table 6.1. The table indicates that the acceptance rate difference between the HC conditions is negligible ($=0.03$). The difference between the HSG conditions is similar ($=0.09$). The HC.DSG is more acceptable than the HSG.DSG condition, and similarly the HC.DPL than the HSG.DPL, which are basically grammaticality effects: because when the head noun is a collective noun, it bears a plural number feature, which matches the number feature on the reflexive. The reaction times are generally longer than 10 seconds and there is much variation.

Table 6.1: Mean acceptance rates and reaction times by condition for Experiment 1

Condition	Accept_Rate (sd)	Reaction_Time (sd)
HC.DPL	5.61 (1.27)	13127 (22894)
HC.DSG	5.58 (1.29)	11291 (11260)
HSG.DPL	4.72 (1.68)	12568 (21500)
HSG.DSG	4.63 (1.81)	12151 (23158)

In an ideal situation, the full model with all the factors that are relevant to the design of the experiment, as shown below in (27), will be reported in all the following experiments. However, there are situations where the full model does not converge, or leads to a singular fit. I will report the maximal model that successfully converges for each experiment. The full model includes the two predictors (head noun type HNoun and distractor number DNum) and their interaction, the control factors (the accuracy rate of the participants for the comprehension questions, the data collection time, and the differences of the frequency, orthographic neighborhood density, phonological neighborhood density, orthographic length, phonological length between the targets and the distractors' head nouns) and the random factors (random intercepts for both subject and item and by-subject random slopes for the main effects and their interaction). The differences of the frequency, orthographic neighborhood density, phonological neighborhood density, orthographic length and phonological length between the targets and the distractors' head nouns are calculated by subtracting these types of information of the distractors from those of the head nouns. All these differences, as well as the judgment data, were scaled before entering into the statistic models, except the accuracy rate (for an easier interpretation of the statistical tests results) and the data collection time. Data collection time was included only when the data were collected with multiple batches in Amazon Mechanical Turk.

(27) Judgment \sim

(predictors:) HNoun + DNum + HNoun:DNum +

(control factors:) Frequency + Phonological-Neighborhood-Density +
 Orthographic-Neighborhood-Density +
 Orthographic-Length + Phonological-Length +
 Accuracy-Rate + (Data-Collection-Time +)

(random intercepts:) (1|Subject) + (1|Item)

(random slopes:) (0+HNoun|Subject) + (0+DNum|Subject) +
 (0+HNoun:DNum|Subject)

Statistical tests were conducted with the `lmer` function in the *lme4* package (Bates et al. 2015) in R (R Core Team 2014). The full model leads to singular fit for the judgment data from Experiment 1. Therefore, all random slopes except the by-subject random slope for the interaction between HNoun and DNum were removed from the model.

Note that for all the experiments, the HC is encoded as the default for the factor HNoun, and DPL is the default for the factor DNum. The Satterthwaite approximation implemented in the *lmerTest* package (Kuznetsova et al. 2017) was used to compute p-values. In addition, to make the models easier to converge, I also set the optimizer method to “bobyqa” (bound optimization by quadratic approximation), and increased the maximum iteration to 100000.

The results of the statistical test can be found in Table 6.2. The results indicate that the head noun being a collective noun significantly improves the acceptability of the target sentences ($t = 9.967$, $p < .001$). This is a grammaticality effect because when the head noun is a collective noun, the target sentences are grammatical given that the semantic plural feature on the collective head noun can match the plural number feature on the reflexive. For the control factors, the difference between the frequency of the target head and that of the distractor head is a significant contributor to the variance ($t = 2.728$, $p = .007$). In addition, the difference between the phonological length of the target head and that of the distractor head is marginally

significant at $\alpha = .05$ ($t = 1.876$, $p = .064$). All other factors are not significant predictors, which include the interaction between HNoun and DNum, giving no clue about whether facilitatory effects exist within the two HC conditions and the two HSG conditions.

Table 6.2: LMM results for Experiment 1

	Estimate	Std. Error	df	t value	Pr(> t)	
(Intercept)	-1.635	1.305	61	-1.253	0.215	
HNoun	0.562	0.056	1265	9.967	<2e-16	***
DNum	0.032	0.052	1387	0.607	0.544	
HNoun:DNum	5e-04	0.088	211	0.006	0.995	
Frequency	0.079	0.029	153	2.728	0.007	**
Orth-Neigh-Density	-0.019	0.031	180	-0.612	0.541	
Phon-Neigh-Density	0.043	0.031	296	1.384	0.167	
Orth-Length	-0.075	0.055	143	-1.378	0.170	
Phon-Length	0.115	0.061	100	1.876	0.064	.
Accu-Rate	1.488	1.397	61	1.064	0.291	
Collect-Time	-0.021	0.163	61	-0.127	0.899	

Planned comparisons were conducted to test if the mean acceptance of the HC.DPL condition is different from that of the HC.DSG condition, and similarly if that of HSG.DPL is different from that of HSG.DSG, using the `testFactors` function in the R package *phia* (De Rosario-Martinez et al. 2015). None of these comparisons show a stable effect. The HC.DPL condition is more acceptable than the HSG.DPL condition ($\chi^2(1) = 57.650$, $p < .001$) and HC.DSG is more acceptable than HSG.DSG ($\chi^2(1) = 99.335$, $p < 0.001$), confirming the grammaticality effects we observed in the statistical model.

6.3.5 Discussion

The most important result of Experiment 1 is: There is no difference between the HC conditions. Recall that the Asymmetry of Interpretability Hypothesis predicts that such a difference could be detected. Therefore, we get null results and thus the

prediction of the Asymmetry of Interpretability Hypothesis remains unjustified. In other words, illusions of grammaticality were not observed with collective nouns as the head noun. It may be argued that the Asymmetry of Interpretability Hypothesis is thus falsified. However, we should be very careful about conclusions based on null results. We should not make this conclusion until we have filtered out other equally viable possibilities. Since a better picture will be obtained once we introduced all the experiments, we leave to the General Discussion section the discussion of reasons why we have null results in Experiment 1.

The results for the HSG conditions are consistent with previous studies which suggest that no illusions of grammaticality could be found with reflexive binding.

While exploring the data, I found that there was a big variance on the participants' reaction times. Some participants provided their judgment in a very short time period, while others spent much more time. I then split the data into two categories, depending on whether the data were produced by participants who responded quickly or slowly. The participants who responded with a mean reaction time of the target sentences higher than or equal to the median reaction time of all subjects (11335 ms) were categorized as slow participants ($N = 33$), and the rest as fast participants ($N = 32$). New LMMs as that in (27) but with the by-subject random slopes for HNoun and DNum removed (to make the model converge) were built for each set of participants.

The results from the fast participants indicate that HNoun is a significant predictor ($t = 7.79$, $p < .001$), similar to the results for all participants in Table 6.2. In addition, the orthographic length is also a significant factor ($t = -1.999$, $p = .049$), and the phonological length approaches marginal significance ($t = 1.759$, $p = .083$). The results from the slow participants reveal a main effect of HNoun ($t = 6.64$, $p < .001$), in addition to a main effect of frequency ($t = 3.079$, $p = .003$). Therefore, no important differences between the fast and slow participants were found.

6.4 Experiment 2: speeded judgment task with reflexive binding

This experiment is a replication study of Experiment 1 using a speeded judgment task. There are two crucial differences compared to Experiment 1: (i) when participants proceed to provide their judgment on a 7-point scale, the test sentences have disappeared; (ii) participants were asked to provide their answer within 4 seconds after the screen with the 7-point scale appears. After 4 seconds, the screen with the 7-point scale would disappear and participants' responses would not be recorded. Participants were then reminded that they needed to respond faster. This is to address a potential concern regarding Experiment 1; namely, in Experiment 1, participants were able to look at the test sentences when they provide the judgment and they had sufficient time to (re-)analyze the sentences.

6.4.1 Participants

One hundred and forty-four participants were recruited from Amazon Mechanical Turk. One participant did not finish the experiment. Another fifteen participants had an accuracy rate below 75% and are thus excluded from data analysis. For the rest 128 participants, 53 were female. The mean age of the 128 participants is 38 (ranging from 21 to 70).

6.4.2 Stimuli

The stimuli for this experiment are the same with those for Experiment 1.

6.4.3 Procedure

After a participant finish reading a target sentence/filler, s/he needs to press the spacebar to continue to the next screen. The participant will be instructed to press the

spacebar again to proceed to the screen with a 7-point scale. S/he will click the box with the scale number s/he chooses. The participant will need to press the spacebar again. If the sentence is not followed by a comprehension question, the participant would see an instruction asking them to press the spacebar to continue to read the next sentence. If, on the other hand, the sentence is followed by a comprehension question, they would see the comprehension question instead. They would need to answer the question (without a time limitation) by clicking either the “Yes” or “No” box. Feedback regarding the correctness of their answer (“Good.” or “Wrong.”) will be provided, before they move to the next sentence.

6.4.4 Results and discussion

Figure 6.2 shows the original 7-point scale data. NA indicates trials where no data was collected because participants did not respond within 4 seconds. With a simple visual inspection, the HC conditions are more acceptable than the HSG conditions.

This impression is confirmed by the mean ratings of the target sentences by condition, as shown in Table 6.3.

Table 6.3: Mean acceptance rates and reaction times by condition for Experiment 2

Condition	Accept-Rate (sd)	Reaction-Time (sd)
HC.DPL	5.48 (1.36)	1374 (634)
HC.DSG	5.44 (1.37)	1315 (615)
HSG.DPL	4.75 (1.75)	1349 (623)
HSG.DSG	4.64 (1.74)	1349 (620)

To test if the differences between these four conditions reach statistical significance, an LMM model is built, with the scaled judgment responses as the dependent variable. The predictors, control factors and random factors of the statistic model are shown below in (28). The by-subject random slope for HNoun was excluded from the full model in Experiment 1 otherwise the model cannot converge. Different from Ex-

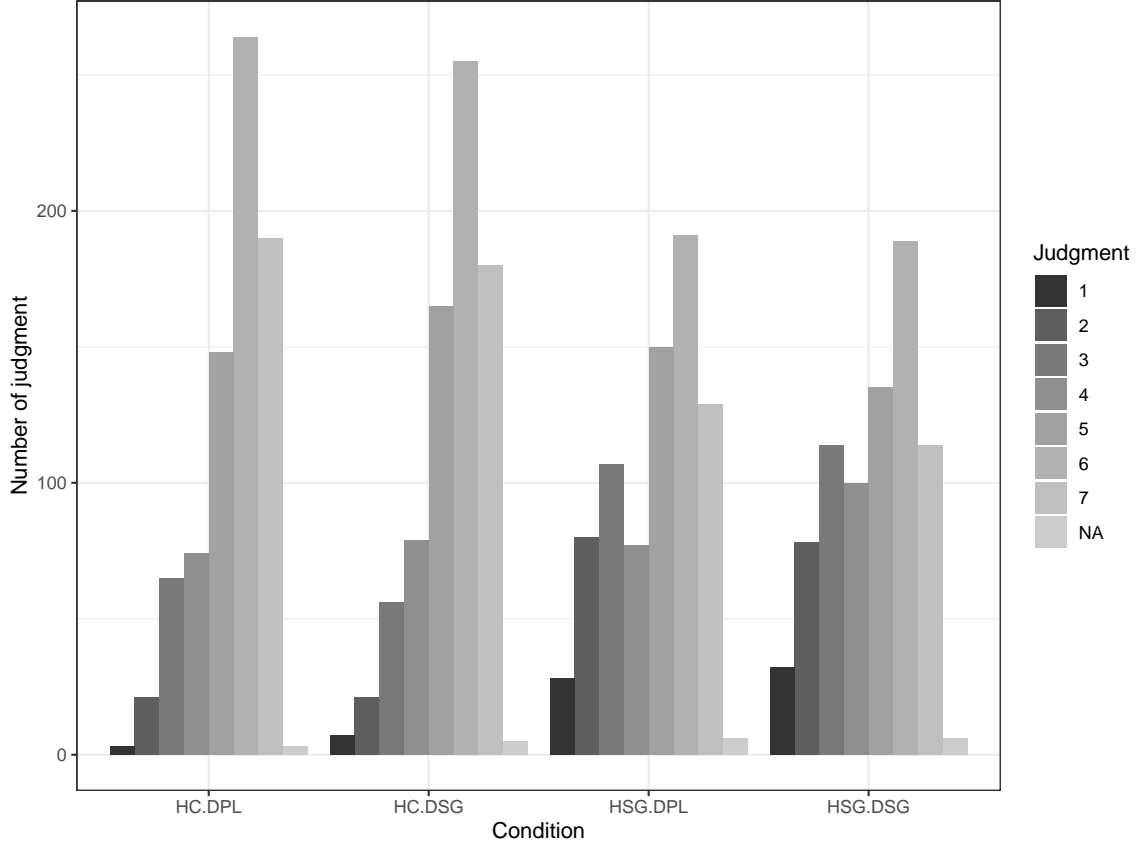


Figure 6.2: The 7 point Likert scale data from Experiment 2

periment 1, I include the three-way interaction between HNoun, DNum and accuracy rate in the model, to compare with the results from the speeded judgment task on agreement in Experiment 3. For Experiment 1, including this three-way interaction does not change the modeling results.

(28) Judgment \sim

(predictors:) HNoun + DNum + HNoun:DNum +
 (control factors:) Frequency + Phonological-Neighborhood-Density +
 Orthographic-Neighborhood-Density +
 Orthographic-Length + Phonological-Length +
 Accuracy-Rate + HNoun:DNum:Accuracy-Rate +
 (random factors:) (1|Subject) + (1|Item) + (0 + DNum|Subject) +

$$(0 + \text{HNoun:DNum}|\text{Subject})$$

The results from the model are presented in Table 6.4. The HC conditions are more acceptable than the HSG conditions, as indicated by the main effect of HNoun ($t = 11.302$, $p < .001$). There is also an interaction effect between HNoun and DNum ($t = -3.158$, $p = .002$), and another between accuracy rate, HNoun and DNum ($t = 3.12$, $p = .002$). In addition, the control factors frequency ($t = 3.568$, $p < .001$), phonological neighborhood density ($t = 3.3$, $p = .001$), and accuracy rate ($t = -2.106$, $p = .037$) are significant contributors to the variance.

Table 6.4: LMM results for Experiment 2

	Estimate	Std. Error	df	t value	Pr(> t)	
(Intercept)	1.299	0.746	131	1.742	0.084	.
HNoun	0.485	0.043	2531	11.302	<2e-16	***
DNum	0.058	0.040	378	1.441	0.150	
Frequency	0.089	0.025	575	3.568	3e-4	***
Orth-Neigh-Density	0.002	0.026	683	0.088	0.930	
Phon-Neigh-Density	0.085	0.026	1061	3.300	0.001	***
Orth-Length	-0.046	0.047	635	-0.975	0.330	
Phon-Length	0.033	0.055	384	0.605	0.545	
Accu-Rate	-1.780	0.845	129	-2.106	0.037	*
HNoun:DNum	-1.796	0.569	110	-3.158	0.002	**
HNoun:DNum:Accu-Rate	2.010	0.644	109	3.120	0.002	**

In order to pinpoint which comparison(s) between the conditions has led to the interaction between HNoun and DNum, planned comparisons are conducted using the `testFactors` function in R package *phia*, with HNoun and DNum combined as a single condition that has four levels.⁷ The results of the planned comparisons are shown in Table 6.5. We see that no difference is found between the two HC conditions and the two HSG conditions, indicating that no illusions of grammaticality occur in the HC.DPL as well as the HSG.DPL condition. The higher ratings in HC.DPL than

⁷The reason to present the model with the two predictors separated is to explore whether the interaction term is a significant predictor. In the model with the two predictors combined, such interaction information is no longer available.

that in HSG.DPL, and in HC.DSG than in HSG.DSG, are grammaticality effects, since the HC conditions are generally more acceptable than the HSG conditions.

Table 6.5: Planned comparisons between conditions in Experiment 2

Comparison	Adj. Mean ⁸	Std. Error	df	Chisq	Pr(>Chisq)	
HC.DPL – HC.DSG	0.026	0.044	1	0.351	0.553	
HSG.DPL – HSG.DSG	0.058	0.039	1	2.207	0.137	
HC.DPL – HSG.DPL	0.450	0.048	1	89.805	2e-16	***
HC.DSG – HSG.DSG	0.482	0.043	1	127.185	2e-16	***

Therefore, the results are essentially consistent with Experiment 1 in that the HC conditions are judged as more acceptable than the HSG conditions. In addition, no illusions of grammaticality were detected in both experiments.

Finally, given that the three-way interaction between HNoun:DNum and accuracy rate is also a significant predictor, I present below in Figure 6.3 the three-way interaction based on the model’s prediction of the judgment values, plotted with the *sjPlot* package (Lüdtke et al. 2019). The figure demonstrates clearly that the grammatical illusions in the HSG.DPL condition are more possible to occur with participants who have a lower accuracy rate, whereas such illusions disappear with participants who are very accurate in answering the comprehension questions. This result is interesting in that it may explain why illusions of grammaticality were not usually observed in the HSG.DPL condition in reflexive binding. I will turn back to this result when I compare the results of Experiment 3 to those of Experiment 2.

The acceptance rates of the HC conditions, however, do not interact with accuracy rates, as the acceptance rates of both HC conditions are consistently very high. The absence of grammatical illusions in the HC.DPL condition is therefore likely due to a ceiling effect. Since both the HC conditions are rated highly acceptable, the differences between them can be hardly induced.

⁸The adjusted means (Adj. Mean) are computed by subtracting the mean reading time in the second condition from that in the first condition under a comparison, with covariates controlled.

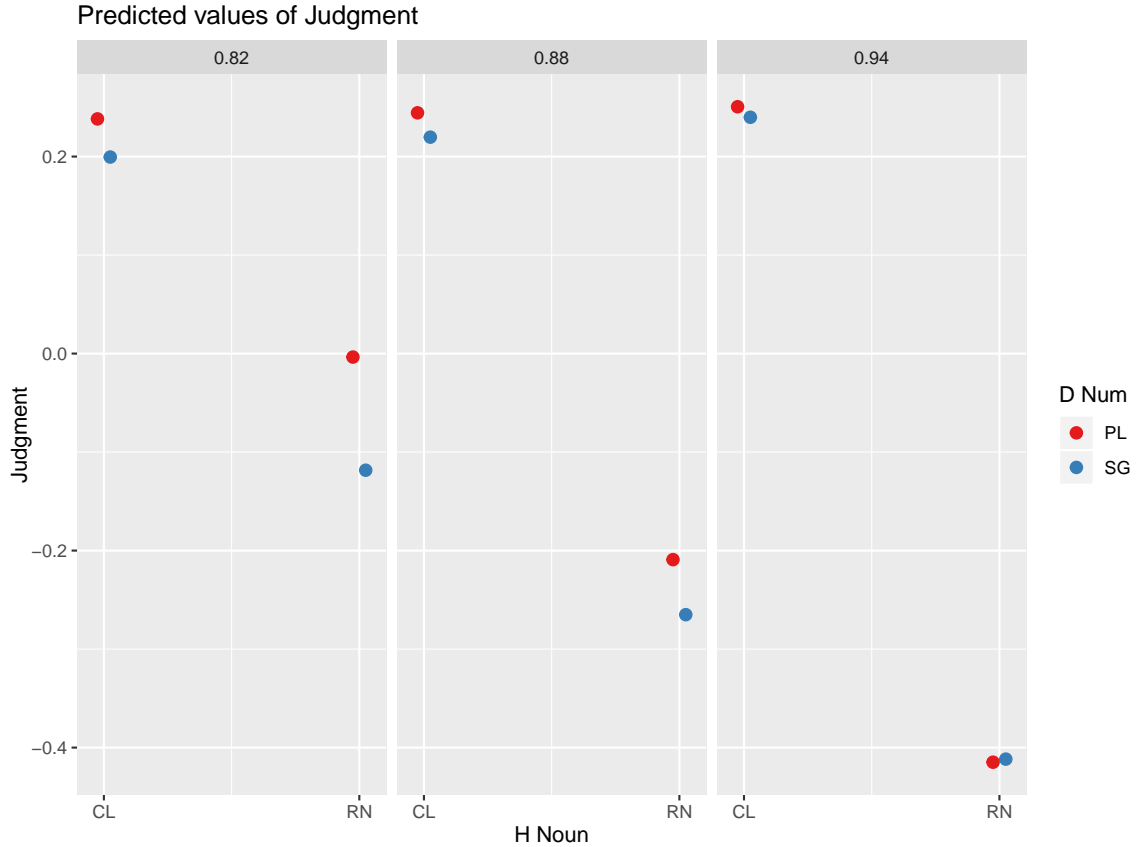


Figure 6.3: Three-way interaction HNum:DNum:Accu-Rate in Experiment 2

CL = collective noun head, RN = regular noun head, PL = plural distractor head, SG = singular distractor head

6.5 Experiment 3: judgment task with subject–verb agreement

The main goal of this experiment is to test if illusions of grammaticality occur in subject–verb agreement. The results from the HSG conditions (HSG.DPL vs. HSG.DSG) is to address the question whether illusions of grammaticality happen in subject–verb agreement in regular cases where the target head nouns are not collective nouns. The results from the HC conditions (HC.DPL vs. HC.DSG) are used as a baseline for the results from reflexive binding in Experiment 2.

Importantly, to the best of my knowledge, no studies have formally reported find-

ings regarding grammatical illusions with grammaticality or acceptability judgment tasks, except an offline study by Dillon et al. (2013) and another by Steven Langsford (p.c.). Langsford did a large-scale acceptability judgment study (as well as a grammaticality judgment task) online with test sentences constructed based on materials from Bock & Miller (1991), on participants recruited by Amazon Mechanical Turk. Following Bock & Miller (1991), Langsford tested different types of constructions, including simple sentences such as “the key to the cabinet are on the table.” and more complex ones such as “the receptionist who greeted the distinguished visitors were fluent in Spanish.” The results indicate that subject–verb agreement constructions do lead to grammatical illusions. Dillon et al.’s (2013) study is more relevant to us, as they tested the acceptability of sentences similar to the ones that I employ as target sentences in this chapter, and they tested both reflexive binding and subject–verb agreement, on 12 participants by asking them to judge the acceptability of the sentences with a 7-point scale. Dillon et al. report that no evidence for grammatical illusions caused by the distractor’s number is found in either the reflexive binding or the subject–verb agreement constructions.

Therefore, previous studies do not agree on whether grammatical illusions occur in offline acceptability judgment studies. Of course, it must be noted that Dillon et al. (2013) have a small sample size and thus may be lack of power to detect the effects.

6.5.1 Participants

One hundred and forty-four participants finished the experiment. Twelve participants were excluded from data analysis because their accuracy rate for the comprehension questions were lower than 75%. The rest 132 participants include 53 females and 79 males, ranging in age from 19 to 71, with a mean age of 37.

6.5.2 Stimuli

The target sentences are as described in the overview section (Section 6.2) for subject–verb agreement.

6.5.3 Procedure

In this experiment, the participants were asked to judge the sentences they read on a 7-point scale (1 being bad, and 7 being good). Similar to Experiment 2, participants read a sentence and the sentence disappears when they proceed to judge the acceptability on the 7-point Likert scale. In addition, participants are required to provide their judgment within 4 seconds, and after 4 seconds, the screen with the 7-point scale disappears and participants are reminded that they need to respond faster. Finally, feedback to participants' answers to the comprehension questions is provided after both their correct or wrong answers. Other details of the procedure can be found in the Procedure Section in Experiment 2.

6.5.4 Results and discussion

The amount of different types of responses are shown in Figure 6.4.

The results show clearly that when the acceptability ratings are smaller than 4; that is, when the participants judged the target sentences unacceptable, the amount of the responses increases monotonically from HC.DPL to HC.DSG, then to HSG.DPL, and finally to HSG.DSG, suggesting that the target sentences were considered less and less acceptable throughout these four conditions in order. The opposite pattern holds for the acceptability ratings bigger than 4, confirming the same point.

The mean acceptance rates (and the corresponding standard deviations) and the mean reaction times (and the corresponding standard deviations) by condition can be found in Table 6.6. The results confirm our observation based on the raw results presented in Figure 6.4.

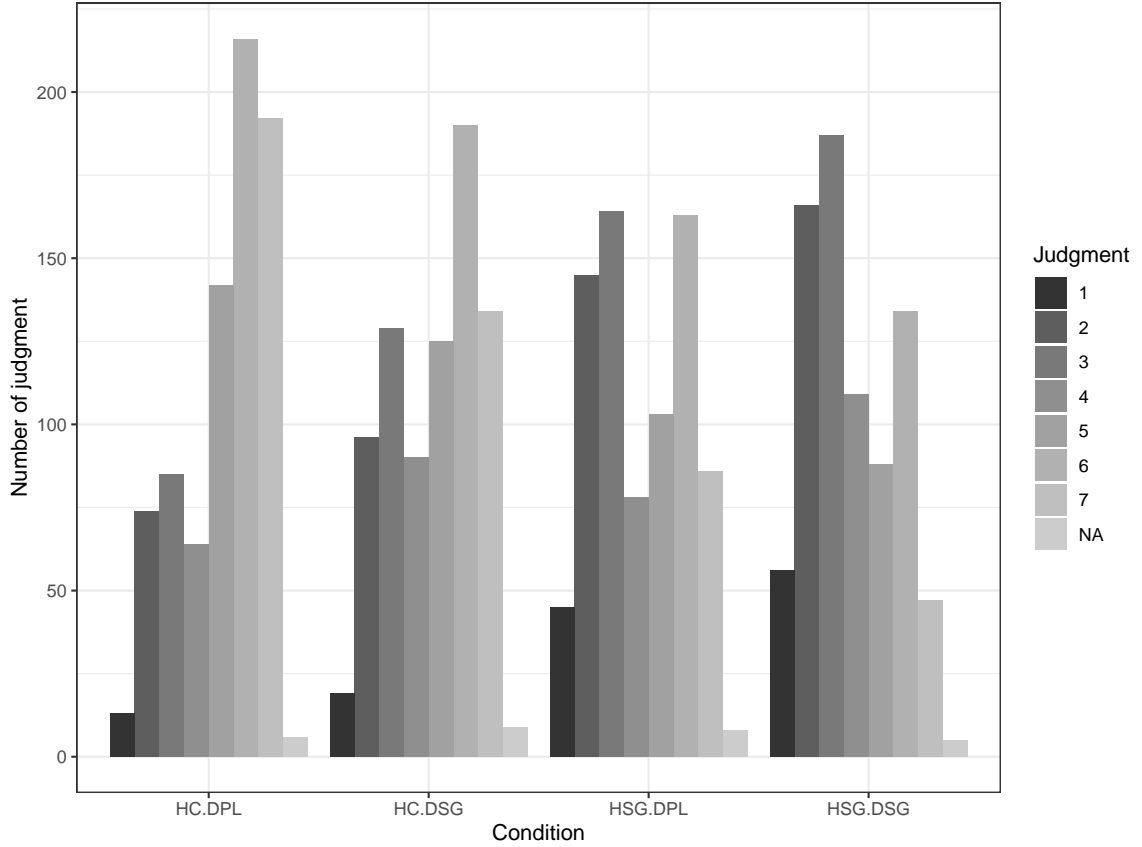


Figure 6.4: The 7 point Likert scale data from Experiment 3

Now let us see if statistical tests are consistent with the observation. As shown in (29), the model is the same to the one for Experiment 2, except that the by-subject random slope for DNum was excluded (otherwise the model cannot converge). I included the three-way interaction between HNoun, DNum and accuracy rate mainly because it is found that without such a three-way interaction, the accuracy rate contributes a significant amount of variance to the dependent variable. It is therefore interesting to examine whether accuracy rate affects or interacts with the emergence of illusions of grammaticality.

(29) Judgment \sim

(predictors:) HNoun + DNum + HNoun:DNum +

(control factors:) Frequency + Phonological-Neighborhood-Density +

Table 6.6: Mean acceptance rates and reaction times by condition for Experiment 3

Condition	Accept_Rate	Reaction_Time
HC.DPL	5.12 (1.69)	1290 (582)
HC.DSG	4.68 (1.76)	1333 (635)
HSG.DPL	4.12 (1.85)	1323 (636)
HSG.DSG	3.76 (1.73)	1283 (606)

Orthographic-Neighborhood-Density +
Orthographic-Length + Phonological-Length +
Accuracy-Rate + HNoun:DNum:Accuracy-Rate +
(random factors:) (1|Subject) + (1|Item) + (0 + HNoun:DNum|Subject)

The statistical model was built using the `lmer` function in the *lme4* package. The results are reported in Table 6.7. Importantly, there are main effects for both HNoun ($t = 13.062$, $p < .001$) and DNum ($t = 5.884$, $p < .001$), showing that both grammaticality effects and illusions of grammaticality occur. The accuracy rate is a significant negative contributor to the variance ($t = -3.111$, $p = .002$); that is, the more accurate the participants were, the lower score the participants gave to the target sentences. Interestingly, there is both an interaction between HNoun and DNum ($t = -2.394$, $p = .018$) and a three-way interaction between HNoun, DNum and accuracy rate.

Follow-up planned comparisons further indicate that all comparisons between these conditions are statistically significant. Most crucially, there are illusions of grammaticality in the HC.DPL condition compared to the HC.DSG condition; the same holds for the HSG.DPL condition, compared to the HSG.DSG condition. These results are all consistent with the predictions of the Asymmetry of Interpretability Hypothesis.

Given that there is a three-way interaction between HNoun, DNum and accuracy rate, a plot of the interaction is provided in Figure 6.5, using the *sjPlot* package. Fig-

Table 6.7: LMM results for Experiment 3

	Estimate	Std. Error	df	t value	Pr(> t)	
(Intercept)	2.599	0.954	133	2.724	0.007	**
HNoun	0.499	0.038	2468	13.062	<2e-16	***
DNum	0.205	0.035	2844	5.884	4e-09	***
Frequency	0.063	0.021	332	3.012	0.003	**
Orth-Neigh-Density	0.013	0.022	377	0.557	0.578	
Phon-Neigh-Density	0.004	0.022	660	0.170	0.865	
Orth-Length	0.017	0.040	327	0.436	0.663	
Phon-Length	-0.047	0.046	209	-1.032	0.303	
Accu-Rate	-3.309	1.064	133	-3.111	0.002	**
HNoun:DNum	-1.404	0.587	133	-2.394	0.018	*
HNoun:DNum:Accu-Rate	1.609	0.653	131	2.466	0.015	*

Table 6.8: Planned comparisons for Experiment 3

Comparison	Adj. Mean	Std. Error	df	Chisq	Pr(>Chisq)	
HC.DPL – HC.DSG	0.240	0.035	1	46.137	1e-11	***
HSG.DPL – HSG.DSG	0.204	0.036	1	31.224	2e-08	***
HC.DPL – HSG.DPL	0.532	0.048	1	124.453	<2e-16	***
HC.DSG – HSG.DSG	0.496	0.050	1	98.617	<2e-16	***

ure 6.5 shows that the three-way interaction is mainly driven by the HC conditions: if the target head noun is a collective noun, the facilitatory effects in the HC.DPL condition, compared to the HC.DSG condition, are smaller when the accuracy rate is low than when the accuracy rate is high. By contrast, facilitatory effects in the HSG.DPL condition, compared to the HSG.DSG condition, are more constant across participants with different accuracy rates. Importantly, this pattern is the opposite to what have been observed in Experiment 2, where the facilitatory effects in the HC.DPL condition are relatively unaffected by accuracy rate but those in the HSG.DPL condition are influenced, but in a different direction: when the accuracy rate is higher, facilitatory effects become smaller and even disappear.

Another interesting observation is that for both the HC conditions and HSG con-

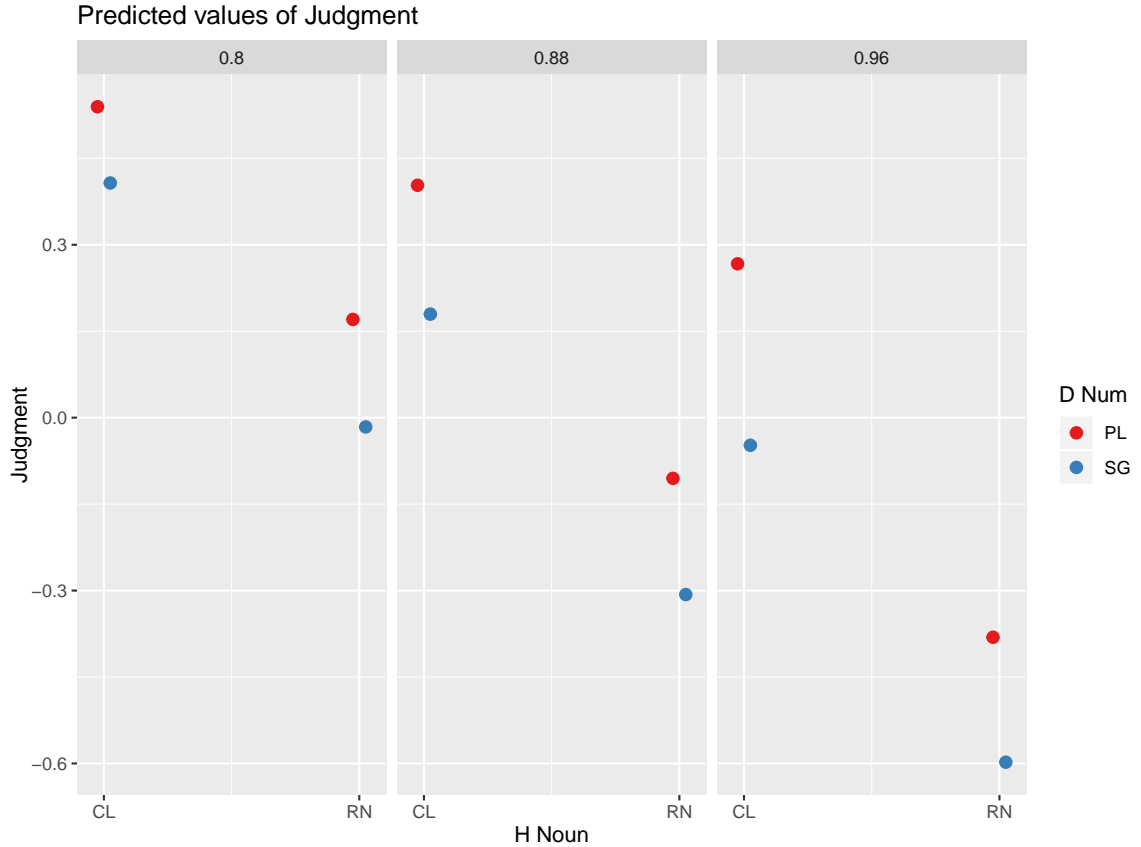


Figure 6.5: three-way interaction HNoun:DNum:Accu-Rate in Experiment 3

CL = collective noun head, RN = regular noun head, PL = plural distractor head, SG = singular distractor head

ditions, participants with a higher accuracy rate judged the sentences more unacceptable. That is, all the target sentences become more unacceptable if the participants read more carefully. This is different from what we have seen in reflexive binding (shown in Figure 6.3), where only the target sentences in the HSG conditions are degraded when the accuracy rate is higher. The HC conditions are in fact judged more acceptable (numerically) when the accuracy rate is higher. This is consistent with a linguistic theory that considers the reflexive binding target sentences in the HC conditions grammatical (although the structures rarely occur in the corpus), and the corresponding target sentences in the HSG conditions ungrammatical; by contrast, such a theory considers the subject–verb target sentences in both the HC conditions

and the HSG conditions ungrammatical.

Why should the above three-way interaction have a completely different pattern in the subject–verb agreement case (Experiment 3) compared to the reflexive binding case (Experiment 2)? We can break this question into two sub-questions: (i) Why do the facilitatory effects in the HSG.DPL condition disappear when the accuracy rate is very high in reflexive binding, but such effects persist under the same situation in subject–verb agreement? (ii) Why do the facilitatory effects in the HC.DPL condition in subject–verb agreement become much more prominent when the accuracy rate is higher, whereas such effects in the same condition in reflexive binding are consistently not detected.

Although better answers to these two sub-questions could only be provided when results from other experiments are discussed, I will discuss some conjectures below. For the first question, the results give us a clue about a condition where grammatical illusions are more likely to be observed in the HSG.DPL condition in reflexive binding: When participants do not read carefully enough (thus these participants have lower accuracy rates). This is potentially consistent with the Asymmetry of Interpretability Hypothesis. The Asymmetry of Interpretability Hypothesis predicts that the mismatch of semantic retrieval cue in reflexive binding will lead to a higher probability of “ungrammatical” judgement. The more careful the participants are, the more possible that they detect the mismatch of the semantic retrieval cue. Once the mismatch of the semantic retrieval cue is detected, the sentence will be judged ungrammatical no matter the distractor match or mismatch the retrieval cue (when the target is retrieved). In this case, they will give almost equally low ratings to the test sentences under the target mismatch, distractor match condition (e.g., the HSG.DPL condition) and the target mismatch, distractor mismatch condition (e.g., the HSG.DSG condition), thus grammatical illusions are concealed. Consequently, facilitatory effects are predicted to be less possible to be detected in the HSG.DPL

condition under reflexive binding. Given that there is no such mismatch of semantic retrieval cue in subject–verb agreement, we correctly predict that accuracy rate does not interact with facilitatory effects in the HSG.DPL condition in subject–verb agreement.

Regarding the HC.DPL condition in reflexive binding, the three-way interaction presented in Figure 6.3 for Experiment 2 suggests that it could be because that the sentences with a collective target head noun are judged acceptable/grammatical and thus illusions of grammaticality cannot be detected. That is, once a sentence is judged acceptable, no facilitatory effects should be detected for that sentence. Previous studies observe a relevant asymmetry: facilitatory effects were observed for ungrammatical sentences only; grammatical sentences do not exhibit such effects (e.g., Dillon et al. 2013; Pearlmutter et al. 1999; Wagers et al. 2009). This asymmetry is not hard to understand. If a sentence is grammatical/acceptable itself, the activation level of the target would be quite high, and therefore the misretrieval of the distractor does not add much advantage to the processing of the sentence. In other words, no illusions of grammaticality should be expected for grammatical sentences, because, by definition, illusions of grammaticality are a result of mis-perceiving something ungrammatical as grammatical. In subject–verb agreement, on the other hand, collective target head nouns do not grammatically license the plural agreeing T, especially in American English. Therefore, when participants read carefully, they would judge the HC.HSG condition more unacceptable. Careful reading does not make the HC.DPL condition as much unacceptable, probably because illusions of grammaticality as a consequence of the general memory retrieval process cannot be eliminated or significantly reduced with the involvement of more attention. This explains why facilitatory effects become more prominent in the HC.DPL condition with participants who have a higher accuracy rate in subject–verb agreement.

6.6 Experiment 4: self-paced reading with reflexive binding

This experiment is to investigate whether facilitatory effects could be detected in reflexive binding with reading time measures. The results will be compared to those from the self-paced reading experiment with subject–verb agreement in Experiment 5.

6.6.1 Participants

One hundred and twenty-seven participants finished the experiment. Ten of them are not included in the statistical analysis because their accurate rate to the comprehension questions is lower than 75%. Additional 11 subjects were identified as click-through participants (see the Data Preprocessing section below). The rest 106 participants include 47 females and 59 males, with a mean age of 38, ranging in age from 22 to 65.

6.6.2 Stimuli

The target sentences of this experiment are materials for reflexive binding as listed in (23), in the Overview section 6.2.

For all stimuli, I group NP level lexical items (and some other special cases; see below) together in the presentation of the stimuli. The main reason to have this type of grouping is not to create a traditional phrase-by-phrase self-paced reading task. In other words, the motivation is not to group the sentences into “natural phrase chunks.” Instead, the main motivation is to dis-encourage and identify the click-through without reading behavior. In the preprocessing of the data, I will resort to this design to filter out subjects who are suspicious of clicking through the sentences without reading. Another reason to only combine NP level lexical items together is to keep more fine-grained results, especially for the critical regions and the spill-over positions.

Specifically, I presented the NP [Adj N], [[Adv Adj] N] or [Possessor NP] together in this experiment. I also group [Adv Adj] together (e.g., *exceptionally talented*), as well as conjunctives (just one instance *repairing and maintaining*). Everything else was individually presented. (30) gives an example of how the target sentences are presented by regions.

(30) The | local government | that | supported | the | tough detectives |
 1 2 3 4 5 6
 embarrassed | themselves | with | a | recent allegation | of | serious
 7 8 9 10 11 12 13
 corruption.

Specifically, I do not combine determiners with the corresponding NPs, due to two reasons: (i) I want to obtain fine-grained data for at least two regions after the critical word, and one of these positions are usually a determiner; (ii) in terms of syntactic analysis, the determiner *the* in the relative clause in the subject position and in the [*the* NP PP] construction should be analyzed as a head that merges with the rest of the structure in the last derivational step, according to the DP hypothesis (see e.g., Abney 1987). That is, in the stimuli, [*the* NP PP] is usually not a [[*the* NP] PP] structure, but a [*the* [NP PP]] structure. Similarly, [*the* N *that* RC], where RC represents a relative clause, is not a [[*the* N] *that* RC] structure, but a [*the* [N *that* RC]] structure. If I combine *the* with the noun head together first, this violates the syntax we assume for DPs.

6.6.3 Procedure

A modified phrase-by-phrase self-paced reading task is employed (see the introduction in the Stimuli section). The experiment showed one region of the sentence at a time. The presentation of the stimuli started with two dashes in the center of the screen, in Times New Roman, with a font size of 22 pixels. The participants needed to press the spacebar to proceed. This will cause the current dashes to disappear,

and will reveal the next region of the sentence. When they press the spacebar again, that first part disappears to reveal the next part of the sentence. All the phrases were presented in the center of the screen. The participants were asked to continue hitting the spacebar all the way through to the end of the sentence at a normal pace.

The data were collected in two different times by two batches in Amazon Mechanical Turk. Data collection time is included as a control factor in the statistical modeling.

Other details about the procedure can be found in the general Procedure section 6.2.2.

6.6.4 Results and discussion

6.6.4.1 Data preprocessing

In addition to removing the 10 subjects who have an accuracy rate lower than 75% for the comprehension questions, I excluded reading times that are smaller than 150 ms or higher than 3000 ms. These data are usually produced by participants who click through sentence regions without reading (could be either intentionally or by accident) or by those who are distracted during reading.

Another way to identify data produced by click-through without reading participants is by comparing the reading times on the Region 2 and 6 versus those of 1, 3, 4, and 5. This is because in Region 2 and 6 I have combined the Adj and N together in the presentation. A participant who reads the target sentences in a normal pace would in general need to spend more time on Region 2 and 6 than Region 1, 3, 4 and 5. This is the case for almost all the participants. Another reason that the mean reading time on Region 2 and 6 is expected to be longer than that on Region 1, 3, 4, and 5 is because these latter regions include two determiners and one complementizer. These function words are usually read fast or even skipped in reading.

In practice, I excluded the (the so-called click-through) participants whose mean

reading time on Region 2 and 6 (combined) is not longer than the mean reading time on Region 1, 3, 4, 5 (combined). Eleven participants were identified as click-through participants.

Finally, if the reading time on a specific region from a specific subject is more than 1.5 standard deviations above the mean reading time of all subjects on that region, that point of data is excluded. 6.8% of the data (out from the data of 106 participants) was excluded by this method.

6.6.4.2 Reading times

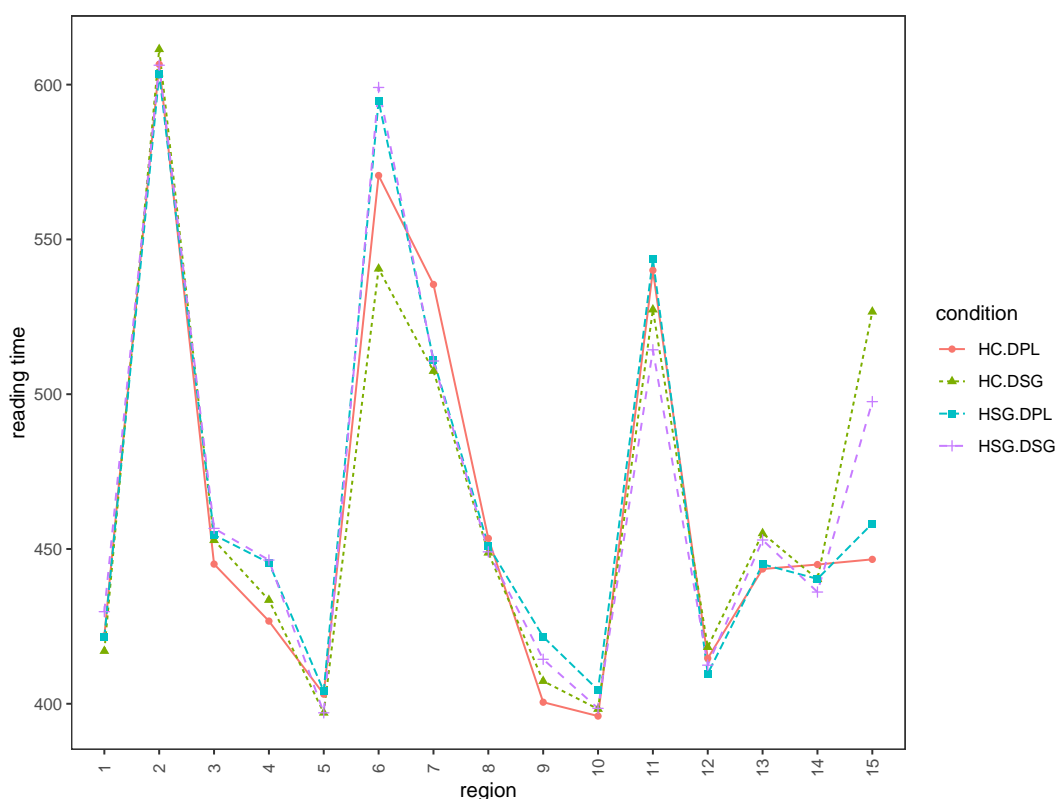


Figure 6.6: Reading times by region for Experiment 4

Region 8 is the critical region, as that is where the reflexive is. The mean reading times on Region 8 and the following two spill-over regions are shown in Table 6.9.

Since the effects due to the reflexives does not seem to be carried over to Region 10, below I will build statistical models for the reading times on Region 8 and 9 only.

Table 6.9: Mean reading times of the critical and spill-over positions in Experiment 4

Condition	Mean reading time (ms)		
	Region 8	Region 9	Region 10
HC.DPL	453 (138) ¹	400 (102)	396 (95)
HC.DSG	449 (132)	407 (103)	398 (100)
HSG.DPL	451 (141)	422 (117)	404 (100)
HSG.DSG	449 (131)	414 (110)	399 (101)

¹ The standard deviations are in parentheses.

6.6.4.3 Statistical tests

Region 8

Let us start from the statistical models that take the reading times on the reflexive in Region 8 as the dependent variable. According to the design of the experiment, I built a linear mixed effects model (LMM) using the `lmer` function in the *lme4* package in R. I attached package *lmerTest* to compute the degree of freedom values and p values are. The full model includes the two predictors (head noun type and distractor number) and their interaction, as well as the control factors (the data collection time, the participants' accuracy rate for comprehension questions, and the differences of the frequency, orthographic neighborhood density, phonological neighborhood density, orthographic length, phonological length between the targets' and the distractors' head nouns) and the random factors (random intercepts for both subject and item). Again, the differences of the frequency, orthographic neighborhood density, phonological neighborhood density, orthographic length and phonological length between the targets' and the distractors' head nouns are all scaled before entering into the statistical models. The maximal model that converges is specified as below in (31). The dependent variable RT is log transformed with the base set to 2 (this is also applied to other self-paced reading experiments reported in this chapter).

$$(31) \quad RT \sim$$

(predictors:) HNoun + DNum + HNoun:DNum +
 (control factors:) Frequency + Phonological-Neighborhood-Density +
 Orthographic-Neighborhood-Density +
 Orthographic-Length + Accuracy-Rate +
 Accuracy-Rate:HNoun:DNum +
 Data-Collection-Time +
 (random factors:) (1|Subject) + (1|Item)

The results of this model are shown below in Table 6.10.

Table 6.10: LMM results for Experiment 4 (Region 8)

	Estimate	Std. Error	df	t value	Pr(> t)	
(Intercept)	8.389	0.474	104	17.685	<2e-16	***
HNoun	-0.015	0.018	1664	-0.798	0.425	
DNum	-0.001	0.017	2252	-0.078	0.938	
Frequency	0.012	0.007	45	1.580	0.121	
Orth-Neigh-Density	-0.021	0.008	61	-2.670	0.010	**
Phon-Neigh-Density	0.006	0.008	71	0.733	0.466	
Orth-Length	-0.018	0.007	44	-2.573	0.014	*
Accu-Rate	0.321	0.529	104	0.607	0.545	
Collect-Time	0.062	0.070	100	0.892	0.375	
HNoun:DNum	0.268	0.205	2246	1.307	0.191	
HNoun:DNum:Accu-Rate	-0.291	0.232	2243	-1.252	0.210	

Importantly, the results show that none of the predictors are significant. There is no main effect for HNoun and DNum, neither for the interaction of them. Planned comparisons between conditions reveal no significant effects.

The results indicate that differences between the targets' and distractors' head nouns in orthographic neighborhood density and orthographic length have a negative influence on the reading times on the reflexive (Region 8), with $t = -2.67$, $p = .01$, and $t = -2.573$, $p = .014$, respectively. That is, when the orthographic neighborhood density or the phonological length of the target head nouns are bigger than those of the distractor head nouns, the reading times on the reflexive are reduced. Although an

consistent account for these effects needs to be developed, the current results, together with results for the control factors from the previous experiments on reflexive binding (Experiment 1 and 2), suggest that the distractors were not completely ignored in the reading of the reflexive, even though we do not have strong reading time evidence.

Region 9

Recall that Region 9 is the first spillover region after the reflexive. I report below in Table 6.11 the testing results from the maximal LMM model as specified in (31) for Region 8.

Table 6.11: LMM results for Experiment 4 (Region 9)

	Estimate	Std. Error	df	t value	Pr(> t)	
(Intercept)	2.513	0.130	103	19.362	<2e-16	***
HNoun	-0.009	0.005	1668	-1.831	0.067	.
DNum	0.006	0.005	2252	1.340	0.180	
Frequency	0.003	0.002	43	1.377	0.176	
Orth-Neigh-Density	-0.004	0.002	56	-1.856	0.069	.
Phon-Neigh-Density	0.001	0.002	60	0.278	0.782	
Orth-Length	-0.002	0.002	40	-0.990	0.328	
Accu-Rate	0.067	0.145	104	0.463	0.645	
Collect-Time	0.022	0.019	100	1.149	0.253	
HNoun:DNum	0.033	0.054	2239	0.616	0.538	
HNoun:DNum:Accu-Rate	-0.053	0.062	2236	-0.863	0.388	

As previously mentioned, the most important results that are relevant to our predictions are the interaction between the two predictors HNoun and DNum. The results show that there is no such interaction. This is a clue that there is no processing difference between the HC conditions and the HSG conditions.

In order to do planned comparisons, I build a new model that combines the two predictors, HNoun and DNum, as one predictor with four levels. The modeling results are not changed but now we can make contrasts to compare different levels of the combined predictor, using the testFactors function in the R package *phia*.

The contrast between the HC.DPL and HC.DSG is most interesting to us. As

shown in Table 6.12, this contrast is not significant. Similarly, the comparison between HSG.DPL and HSG.DSG is also not significant. The comparisons between HC.DPL and HSG.DPL, HC.DSG and HSG.DSG, on the other hand, are significant or marginally significant, with $\chi^2(1) = 21.471$, $p < .001$ and $\chi^2(1) = 3.29$, $p < .07$, respectively, suggesting that the interaction effect we observed in the LMM model in Table 6.11 is mainly caused by these last two contrasts, which reflect grammaticality effects.

Table 6.12: Planned comparisons for Region 9 in Experiment 4

Comparison	Adj. Mean	Std. Error	df	Chisq	Pr(>Chisq)	
HC.DPL – HC.DSG	-0.007	0.005	1	2.377	0.123	
HSG.DPL – HSG.DSG	0.006	0.005	1	1.815	0.178	
HC.DPL – HSG.DPL	-0.022	0.005	1	21.471	4e-06	***
HC.DSG – HSG.DSG	-0.009	0.005	1	3.290	0.070	.

With a LMM model built upon the sum of the reading times for Region 8 and 9, the main results remain the unchanged, except that now the frequency difference also becomes a significant factor ($t = 2.1$, $p = .04$), suggesting that the distractor is not ignored in the processing of the reflexive and the following spill-over position, although we do not observe reading time differences.

Therefore, it could be concluded from the results of Experiment 4 that in both the HC conditions and the HSG conditions, facilitatory effects were not observed.

6.7 Experiment 5: self-paced reading with subject–verb agreement

This experiment is to investigate whether facilitatory effects can be observed in subject–verb agreement with reading time measures, compared to the results from Experiment 4.

6.7.1 Participants

One hundred and forty-five participants finished the experiment. Six participants were excluded from the analysis because their accuracy rate for the comprehension questions were lower than 75% percent. Additional 8 participants were identified as click-through participants (using the method introduced in data preprocessing section 6.6.4.1 for Experiment 4. The rest 131 participants include 65 females and 66 males, with an age range from 18 to 70 (mean age: 37).

6.7.2 Stimuli

The target sentences are as in (26), identical to the stimuli for Experiment 3. The sentences were presented as a self-paced reading task as described in Experiment 4. An example of the segmentation of the target sentences is shown below:

- (32) The | local government | that | supported | the | tough detectives | were |
1 2 3 4 5 6 7
embarrassed | with | a | recent allegation | of | serious corruption.
8 9 10 11 12 13

The critical position the linking verb *were* in Region 7.

6.7.3 Procedure

The same to the procedure for Experiment 4.

6.7.4 Results and discussion

Reading times in a specific region that are more than 1.5 standard deviations above the mean reading time of that region across all participants were removed. This applied to 6.7% of the data from the 131 participants. The mean reading times by region for all participants are shown in Figure 6.7.

Recall that Region 7 is the critical region because it is the linking verb *were*, and Region 8 and 9 might be considered spill-over positions. The mean reading times

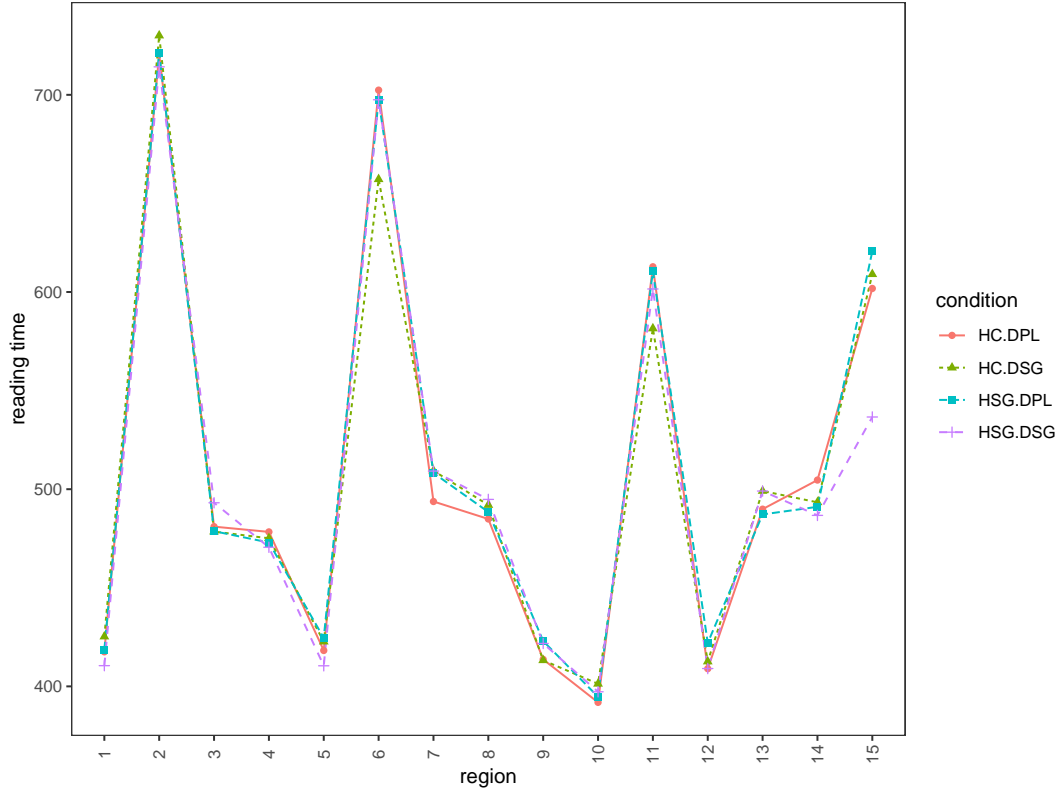


Figure 6.7: Reading times by region for Experiment 5

(and corresponding sds) on Region 7, 8 and 9 by condition can be found in Table 6.13.

Therefore, the average reading time on Region 7 in the HC.DPL condition is 15 ms shorter than that in the HC.DSG condition; however, the reading time on the same region in the HSG.DPL condition is only 1 ms shorter than that in the HSG.DSG condition. As for the first spill-over position, Region 8, the preposition right after the linking verb, the reading time in the HC.DPL condition is 7 ms shorter than that in the HC.DSG condition; and that the HSG.DPL condition is also 7 ms shorter than that in the HSG.DSG condition. Nothing interesting seems to happen on Region 9.

Now let us test, starting from Region 7, whether the differences between critical conditions are statistically significant.

Region 7

The full model as specified in (31) above for reflexive binding converges. The re-

Table 6.13: Mean reading times of important regions in Experiment 5

Condition	Mean reading time (ms)		
	Region 7	Region 8	Region 9
HC.DPL	494 (147) ¹	485 (152)	414 (102)
HC.DSG	509 (165)	492 (152)	413 (99)
HSG.DPL	508 (156)	488 (158)	423 (105)
HSG.DSG	509 (168)	495 (150)	422 (107)

¹ The standard deviations are in parentheses.

sults, as shown in Table 6.14, indicate that no main effect is detected for the two predictors. However, interestingly, an interaction is found between HNoun and DNum. We would need to figure out what has led to such an interaction.

Table 6.14: LMM results for Experiment 5 (Region 7)

	Estimate	Std. Error	df	t value	Pr(> t)	
(Intercept)	9.210	0.370	134	24.882	<2e-16	***
HNoun	0.025	0.019	1881	1.261	0.208	
DNum	0.017	0.018	2796	0.949	0.343	
Frequency	-5e-04	0.009	76	-0.062	0.951	
Orth-Neigh-Density	0.010	0.010	112	1.020	0.310	
Phon-Neigh-Density	-0.012	0.010	151	-1.259	0.210	
Orth-Length	0.040	0.017	82	2.349	0.021	*
Phon-Length	-0.060	0.018	62	-3.247	0.002	**
Accu-Rate	-0.335	0.419	134	-0.799	0.426	
HNoun:DNum	-0.153	0.212	2794	-0.722	0.470	
HNoun:DNum:AccuRate	0.110	0.238	2794	0.460	0.646	

Planned comparisons were conducted using the `testFactors` function in the R package *phia*. We see facilitatory effects in HC.DPL compared to HC.DSG. That is, the reading time on Region 7 in the HC.DPL condition is significantly shorter than that in the HC.DSG condition ($\chi^2(1) = 4.637$, $p = .031$). No facilitatory effects, on the other hand, are observed for the HSG.DPL condition, compared to the HSG.DSG condition.

Table 6.15: Planned comparisons for Region 7 in Experiment 5

Comparison	Adj. Mean	Std. Error	df	Chisq	Pr(>Chisq)	
HC.DPL – HC.DSG	-0.039	0.018	1	4.637	0.031	*
HSG.DPL – HSG.DSG	0.017	0.018	1	0.883	0.347	
HC.DPL – HSG.DPL	-0.032	0.019	1	2.711	0.0996	.
HC.DSG – HSG.DSG	0.025	0.019	1	1.589	0.207	

Region 8

For Region 8, the same LMM model as for Region 7 converges. There is a marginal main effect of DNum ($t = -1.933$, $p = .053$), signaling (weak) facilitatory effects across the HC.DPL and HSG.DPL conditions. The modeling results are shown in Table 6.16.

Table 6.16: LMM results for Experiment 5 (Region 8)

	Estimate	Std. Error	df	t value	Pr(> t)	
(Intercept)	8.909	0.421	132	21.177	<2e-16	***
HNoun	-0.015	0.018	1952	-0.789	0.430	
DNum	-0.033	0.017	2811	-1.933	0.053	.
Frequency	0.004	0.009	71	0.449	0.655	
Orth-Neigh-Density	-0.004	0.009	99	-0.485	0.629	
Phon-Neigh-Density	0.198	0.009	135	0.212	0.832	
Orth-Length	0.023	0.016	75	1.479	0.143	
Phon-Length	-0.013	0.017	57	-0.733	0.467	
Accu-Rate	-0.002	0.476	132	-0.005	0.996	
HNoun:DNum	-0.067	0.202	2810	-0.334	0.738	
HNoun:DNum:Accu-Rate	0.077	0.227	2810	0.340	0.734	

As shown in Table 6.17, planned comparisons further indicate that there are marginal facilitatory effects in both the HC.DPL and HSG.DPL conditions, with $\chi^2(1) = 3.599$, $p = .058$, and $\chi^2(1) = 3.753$, $p = .053$, respectively. All other comparisons do not reach statistical significance.

With the reading times on Region 7 and 8 combined and analyzed together in the statistical modeling, nothing except the intercept is a significant factor.

Therefore, we observed facilitatory effects in the HC.DPL and HSG.DPL condi-

Table 6.17: Planned comparisons for Region 8 in Experiment 5

Comparison	Adj. Mean	Std. Error	df	Chisq	Pr(>Chisq)
HC.DPL – HC.DSG	-0.033	0.017	1	3.599	0.058 .
HSG.DPL – HSG.DSG	-0.034	0.017	1	3.753	0.053 .
HC.DPL – HSG.DPL	-0.014	0.018	1	0.584	0.445
HC.DSG – HSG.DSG	-0.015	0.018	1	0.626	0.429

tions, compared to the HC.DSG and HSG.DSG conditions, in the spill-over position. The overall pattern in the reading task on subject–verb agreement is generally consistent with the predictions of the cue-based retrieval theory.

6.8 Experiment 6: a replication study for reflexive binding

In Experiment 4 and 5, we found facilitatory effects in subject–verb agreement but not in reflexive binding, and this is true for both the HC.DPL and HSG.DPL conditions. By contrast, Experiment 2 reveals a three-way interaction between HNoun, DNum and accuracy rate; specifically, facilitatory effects in the HSG.DPL is only present among the participants who had a lower accuracy rate. Given that the results do not completely converge, below I want to present results from a replication study using a slightly different experimental design: a combination of the self-paced reading and the acceptability judgment task.

6.8.1 Participants

One hundred and forty-four participants were recruited from Amazon Mechanical Turk. One participant did not finish the experiment, and another thirteen participants had an accuracy rate lower than 75% for the comprehension questions. Another two participants were identified as click-through participants. The rest 124 participants range in age from 19 to 67 (mean age: 37), including 58 females and 66 males.

6.8.2 Stimuli

The target sentences and the fillers are the same as those in Experiment 1, 2 and 4. An example of the target sentence is copied as in (33). The sentences were displayed word-by-word, different from Experiment 4 and 5.

(33) The | local | government | that | supported | the | tough | detectives |
1 2 3 4 5 6 7 8
embarrassed | themselves | with | a | recent | allegation | of | serious |
9 10 11 12 13 14 15 16
corruption.
17

6.8.3 Procedure

This experiment is a combination of a self-paced reading task and an acceptability judgment task (forced-choice options: “Good”–“Bad”). The self-paced reading task is different from that of Experiment 4 and 5 in that (i) the sentences were presented word-by-word using the moving window paradigm; the participants were asked to press the spacebar to reveal each word, and the words that have been read or have not yet be read are replaced/masked by slashes.

After the participants finished reading the sentences, they were asked to judge the acceptability of the sentences they just read. There were only two choices, “Bad” and “Good.” Half of the sentences have comprehension questions right after the judgment task.

6.8.4 Results and discussion

This experiment gives us two types of data: reading time data and acceptability judgment data. I will explore these two sets of data in turn in the following two subsections.

6.8.4.1 Reading time data

The self-paced reading task provides reading time data on each word of the sentences. As in previous self-paced reading experiments, reading times below 150 ms or higher than 3000 ms were excluded. In addition, reading times that are 1.5 standard deviations above the mean reading time of all subjects in the same region under the same condition were excluded from further statistical analysis. 0.9% of the data was excluded by this method.

Since in this experiment, the traditional word-by-word moving window self-paced reading task was adopted, I could not use the reading time differences between regions to identify click-through participants. Instead, I set a different criterion: if a participant has reading times on 10 out of the first 12 words that are consistently more than 200 ms lower than the mean of the reading times of all words for all 124 subjects, that participant is identified as a click-through participant. This is a quite arbitrary filtering method, and it is very conservative. It filters out data from two subjects, and the rest of data are not very problematic based on a simple visual inspection of the reading time plot for all subjects.

The overall pattern of the reading time data can be found in Figure 6.8.

The reading times on the reflexive, Region 10, and the two (potential) spill-over positions are shown as in Table 6.18. We see that the reading time on Region 10 under the HC.DPL condition is 20 ms shorter than that in HC.DSG, and that under the HSG.DPL is 19 ms longer than that in the HSG.DSG, a pattern we have observed in Experiment 4. In Region 11, the difference between the HC.DPL condition and the HC.DSG condition remains but becomes much smaller, i.e., the mean reading time under the HC.DPL condition is 5 ms shorter than that under the HC.DSG condition; whereas the direction of the difference between the HSG.DPL and HSG.DSG flips around: now the reading time under the HSG.DSG is 22 ms longer than that under the HSG.DPL. The difference between the HC conditions disappears in Region 12,

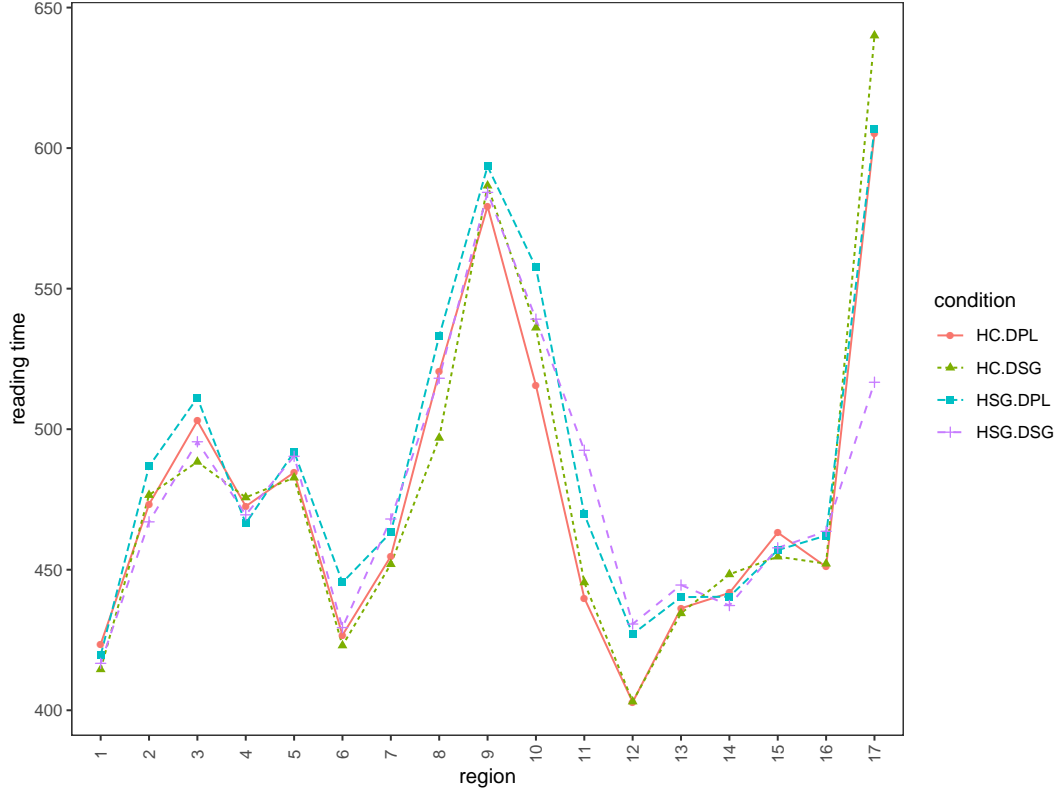


Figure 6.8: Reading times by region for Experiment 6

but the HSG.DPL is still 4 ms shorter than HSG.DSG. Over all the three regions, it is obvious that the grammatical conditions, namely, the HC conditions, are processed faster than the ungrammatical HSG conditions.

Region 10

Let us start from Region 10 and build statistical models to test if the above differences are real or simply noises. The maximal model, identical to the ones for Experiment 4 in (31) except that data collection time is excluded because the data were collected with one single batch in Amazon Mechanical Turk, converges and the modeling results are shown as in Table 6.19.

Besides the intercept, only two factors approach statistical significance at $\alpha = .05$: the difference between the frequency and phonological length of the target head nouns and those of the distractor head noun, with $t = 2.026$, $p = .047$ and $t = -2.089$, $p = .042$, respectively. Note that there is no main effect for HNoun, suggesting that

Table 6.18: Mean reading times of important regions in Experiment 6

Condition	Mean reading time (ms)		
	Region 10	Region 11	Region 12
HC.DPL	516 (264) ¹	440 (198)	403 (183)
HC.DSG	536 (321)	445 (210)	403 (175)
HSG.DPL	558 (353)	470 (238)	427 (222)
HSG.DSG	539 (313)	492 (296)	431 (234)

¹ The standard deviations are in parentheses.

although grammaticality effects are observable with the visual inspection of the raw data, such effects are not stable and do not reach statistical significance. In addition, there is no main effect for the DNum condition, indicating that no facilitatory effects are consistently observed across the HC and HSG conditions.

Table 6.19: LMM results for Experiment 6 (Region 10)

	Estimate	Std. Error	df	t value	Pr(> t)	
(Intercept)	8.193	0.586	125	13.980	<2e-16	***
HNoun	-0.011	0.026	2089	-0.424	0.672	
DNum	0.035	0.025	2802	1.400	0.162	
Frequency	0.024	0.012	66	2.026	0.047	*
Orth-Neigh-Density	-0.011	0.013	80	-0.894	0.374	
Phon-Neigh-Density	0.001	0.132	113	0.112	0.911	
Orth-Length	0.022	0.022	63	1.022	0.311	
Phon-Length	-0.050	0.024	49	-2.089	0.042	*
Accu-Rate	0.821	0.670	125	1.227	0.222	
HNoun:DNum	0.172	0.289	2800	0.595	0.552	
HNoun:DNum:Accu-Rate	-0.257	0.328	2800	-0.784	0.433	

Since a correlation matrix associated with the model shows that the accuracy rate is highly correlated with the intercept, I excluded it from the model and re-ran the model; the modeling results are almost identical to the results presented in Table 6.19.

Planned comparisons were made to test if facilitatory effects occur in the HC.DPL

and HSG.DPL condition. The results are shown in Table 6.20.

Table 6.20: Planned comparisons for Region 10 in Experiment 6

Comparison	Adj. Mean	Std. Error	df	Chisq	Pr(>Chisq)
HC.DPL – HC.DSG	-0.018	0.025	1	0.519	0.471
HSG.DPL – HSG.DSG	0.035	0.025	1	1.959	0.162
HC.DPL – HSG.DPL	-0.063	0.026	1	5.803	0.016 *
HC.DSG – HSG.DSG	-0.011	0.026	1	0.170	0.680

The planned comparisons clearly reveal that the reading times in the HC.DPL condition are significantly shorter than those in the HSG.DPL conditions. The question is whether this difference can be ascribed to grammaticality effects completely. I think the answer is probably negative. As shown by the comparison between the HC.DSG and HSG.DSG conditions, there is no such difference. If grammaticality differences between the HC.DPL condition and the HSG.DPL condition is the sole factor that has led to the reading time difference between the HC.DPL and HSG.DPL conditions, we should expect the same to occur with the HC.DSG and HSG.DSG conditions, contrary to fact.

In light of the qualitative observation from the raw reading time data, I believe the reading time difference between HC.DPL and HSG.DPL is a result of two (hidden) factors combined together: (i) a very weak facilitatory effect associated with HC.DPL; and (ii) a very weak inhibitory effect associated with HSG.DPL. If these two factors are tested separately, both of them are too weak to be detected; however, if these two factors are combined together, then the facilitatory effect is made more prominent due to the additional reading time contribution from the inhibitory effect to the opposite direction. This explanation is consistent with previous studies which observe very weak, statistically insignificant inhibitory effects with regard to regular reflexive binding (corresponding to the HSG.DPL condition in the reflexive binding experiments in this chapter). This explanation is the only account that

can reconcile the following otherwise puzzling patterns observed in this experiment and Experiment 4: (i) the mean reading times under the HC.DPL condition are numerically consistently shorter than those under the HC.DSG condition—however, this difference does not reach statistical significance; (ii) the mean reading times under the HSG.DPL condition are numerically consistently longer than those under the HSG.DSG condition—the differences are again not statistically significant; (iii) in both Experiment 4 and 6, the reading times in HC.DPL are significantly shorter than those in HSG.DPL; whereas grammaticality effects may account for part of this difference in Experiment 4, but not in Experiment 6, as the grammaticality effects (if there is any, although not statistically significant in Experiment 6) are not strong enough to derive such a difference between HC.DPL and HSG.DPL.

What the results tell us then is that facilitatory effects seem to be too weak to be detected in the reading time data on Region 10, although the effects are very likely real as reflected in the consistently observed reading time difference on the reflexive under HC.DPL and HSG.DPL.

Region 11

The maximal model that converges is as the same to that for Region 10 but with the three-way interaction HNoun:DNum:Accu-Rate removed. The results can be found in Table 6.21.

For Region 11, interestingly, there is a main effect for HNoun ($t = -4.602$, $p < .001$), signaling grammaticality effects: the reflexive was read faster in the HC conditions than that in the HSG conditions. There is a marginal main effect of DNum ($t = -1.692$, $p = .091$), and this could be potentially interpreted as facilitatory effects across the HC and HSG conditions. The latter is important to us and planned comparisons were conducted to further explore the facilitatory effects in the HC.DPL and HSG.DPL conditions. The results of the planned comparison are shown in Table 6.22.

Table 6.21: LMM results for Experiment 6 (Region 11)

	Estimate	Std. Error	df	t value	Pr(> t)	
(Intercept)	7.731	0.500	122	15.470	<2e-16	***
HNoun	-0.103	0.022	2131	-4.602	4e-06	***
DNum	-0.036	0.022	2820	-1.692	0.091	.
Frequency	0.012	0.009	49	1.341	0.186	
Orth-Neigh-Density	-0.009	0.010	57	-0.877	0.384	
Phon-Neigh-Density	-0.011	0.011	77	-1.044	0.300	
Orth-Length	-0.004	0.017	46	-0.250	0.804	
Phon-Length	-0.026	0.018	38	-1.408	0.167	
Accu-Rate	1.209	0.571	122	2.117	0.036	*
HNoun:DNum	0.029	0.030	2811	0.965	0.335	

Table 6.22: Planned comparisons for Region 11 in Experiment 6

Comparison	Adj. Mean	Std. Error	df	Chisq	Pr(>Chisq)	
HC.DPL – HC.DSG	-0.007	0.022	1	0.108	0.742	
HSG.DPL – HSG.DSG	-0.036	0.022	1	2.864	0.091	.
HC.DPL – HSG.DPL	-0.074	0.023	1	10.642	0.001	**
HC.DSG – HSG.DSG	-0.103	0.022	1	21.174	4e-06	***

All the comparisons except the one between HC conditions indicated a (marginal) significant difference. The adjusted mean reading time under HSG.DPL is shorter than that under HSG.DSG ($\chi^2(1) = 2.864$, $p = .091$), although this difference does not reach statistical significance at $\alpha = .05$. This is a facilitatory effect associated with regular reflexive binding. Again, grammaticality effect are observed: like the results for Region 10, the reading time in HC.DPL is significantly shorter than that in HSG.DPL ($\chi^2(1) = 10.642$, $p = .001$); the reading time on HC.DSG is also significantly shorter than that in HSG.DSG ($\chi^2(1) = 21.174$, $p < .001$).

Given that the grammaticality effects are very prominent with regard to Region 11, we should say that the difference between HC.DPL and HSG.DPL is mostly a reflection of the grammaticality effects. In addition, unstable facilitatory effects with regard to regular reflexive binding are detected in Region 11. Recall that such effects

were not present in Region 10. Taking the results from Region 10 and 11 together, it may be tentatively hypothesized that facilitatory effects with regular reflexive binding may occur in a later stage, although these effects are quite unstable.

The accuracy rate has a significant contribution to the reading times on Region 11. The more accurate a participant was, the more time the participant would spend on Region 11 ($t = 2.117$, $p = .036$). This could be another aspect of the grammaticality effects, as the more accurate the participants were, the more possible that they would detect the ungrammaticality.

6.8.4.2 Acceptability judgment data

Now let us see if we could replicate the results in Experiment 1 and 2 for the acceptability judgment data. Note that an important difference between the current acceptability judgment task and that in Experiment 1 is that participants cannot see the test sentences when they make the judgment in Experiment 6; whereas in Experiment 1 participants can look at the test sentences while they provide their acceptability judgments. This design is similar to that in Experiment 2. The judgment tasks in Experiment 2 and 6 are closer to the setting up assumed by the cue-based retrieval model, as the acceptability judgment would need to be made based upon the memory retrieval results, rather than direct visual input (e.g., by taking a look at the sentences directly). Another important distinction is that in Experiment 1 and 2, the tasks use a 7-point Likert scale, whereas in Experiment 6, the judgment is a “Good”–“Bad” forced choice task.

Encoding the answer “Good” as 1, and “Bad” as 0, a logistic generalized linear mixed effects model (GLMM) is built, with the maximal model specified in previous experiments (e.g., (31) for Experiment 4), using the *blmer* function in the *blme* package (Dorie 2015). The results are shown in Tabel 6.23.

The results from the statistical reveals a main effect of HNoun ($z = 9.866$, $p <$

Table 6.23: GLMM results for the judgment data in Experiment 6

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	1.676	2.050	0.817	0.414	
HNoun	1.579	0.160	9.866	<2e-16	***
DNum	0.207	0.128	1.621	0.105	
Frequency	0.279	0.071	3.915	9e-05	***
Orth-Neigh-Density	-0.009	0.077	-0.123	0.902	
Phon-Neigh-Density	0.070	0.084	0.836	0.403	
Orth-Length	-0.127	0.128	-0.998	0.318	
Phon-Length	0.161	0.138	1.170	0.242	
Accu-Rate	-0.785	2.339	-0.336	0.737	
HNoun:DNum	-5.498	2.123	-2.590	0.010	**
HNoun:DNum:Accu-Rate	6.427	2.446	2.628	0.009	**

.001), indicating that HC conditions are judged more acceptable than HSG conditions. This is, again, a grammaticality effect. The results also show an interaction between HNoun and DNum ($z = -2.59$, $p = .01$) and a three-way interaction between HNoun, DNum and accuracy rate ($z = 2.628$, $p = .009$). These interaction terms are potentially interesting as they may reveal facilitatory effects in some of the conditions but not others. In addition, the frequency difference between the target head noun and that of the distractor matters to the acceptability of the sentences ($z = 3.915$, $p < .001$), suggesting that both the target and the distractor are relevant for the judgment.

Let us take a look at the planned comparisons to see if grammatical illusions were detected. Table 6.24 shows the results of the planned comparisons. It is confirmed that grammaticality effects are prominent here, as both the HC conditions are more acceptable than the corresponding HSG conditions. There is also an unstable effect of grammatical illusions in the HSG.DPL condition compared to the HSG.DSG condition ($\chi^2(1) = 2.706$, $p = .01$).

Generally consistent with Experiment 1 and 2, we did not find grammatical illusions in the reflexive binding for judgment task in the HC.DPL condition. However,

Table 6.24: Planned comparisons for the judgment data in Experiment 6

Comparison	Adj. Mean	Std. Error	df	Chisq	Pr(>Chisq)	
HC.DPL – HC.DSG	0.567	0.183	1	2.206	0.137	
HSG.DPL – HSG.DSG	0.553	0.128	1	2.706	0.010	.
HC.DPL – HSG.DPL	0.848	0.175	1	96.256	<2e-16	***
HC.DSG – HSG.DSG	0.840	0.165	1	100.591	<2e-16	***

an unstable contrast between HSG.DPL and HSG.DSG is detected in the current experiment, signaling (weak) effects of illusions of grammaticality.

Since there is a three-way interaction between HNoun, DNum and accuracy rate, and further comparison tests show that this is driven by the interaction between the HSG conditions and accuracy rate. As shown in Figure 6.9, for participants who had a lower accuracy rate, the HSG.DPL condition is judged more acceptable than the HSG.DSG condition. These is an effect of grammaticality illusions. Such a pattern also holds for the HC.DPL condition compared to the HC.DSG condition. However, such differences between the DPL and DSG conditions disappear or is significantly reduced among the participants who have a higher accuracy rate.

An interesting, important difference between the HSG and HC conditions is worthy stressing. When the accuracy rate is higher, we find that the predicted percentage of the “Good/Acceptable” judgment of the test sentences becomes smaller for the HSG condition but bigger for the HC conditions. The same pattern has been observed in Experiment 2. This difference between the HC and HSG condition in terms of their interaction with accuracy rate may suggest that the grammaticality of the HC and HSG differ: The reflexive binding target sentences under the HC condition are grammatical whereas those under the HSG condition are ungrammatical. Consequently, more careful participants will judge the test sentences in the HSG conditions unacceptable, but they will judge those in the HC conditions acceptable.

Given that there was no time limitation imposed on the judgment responses,

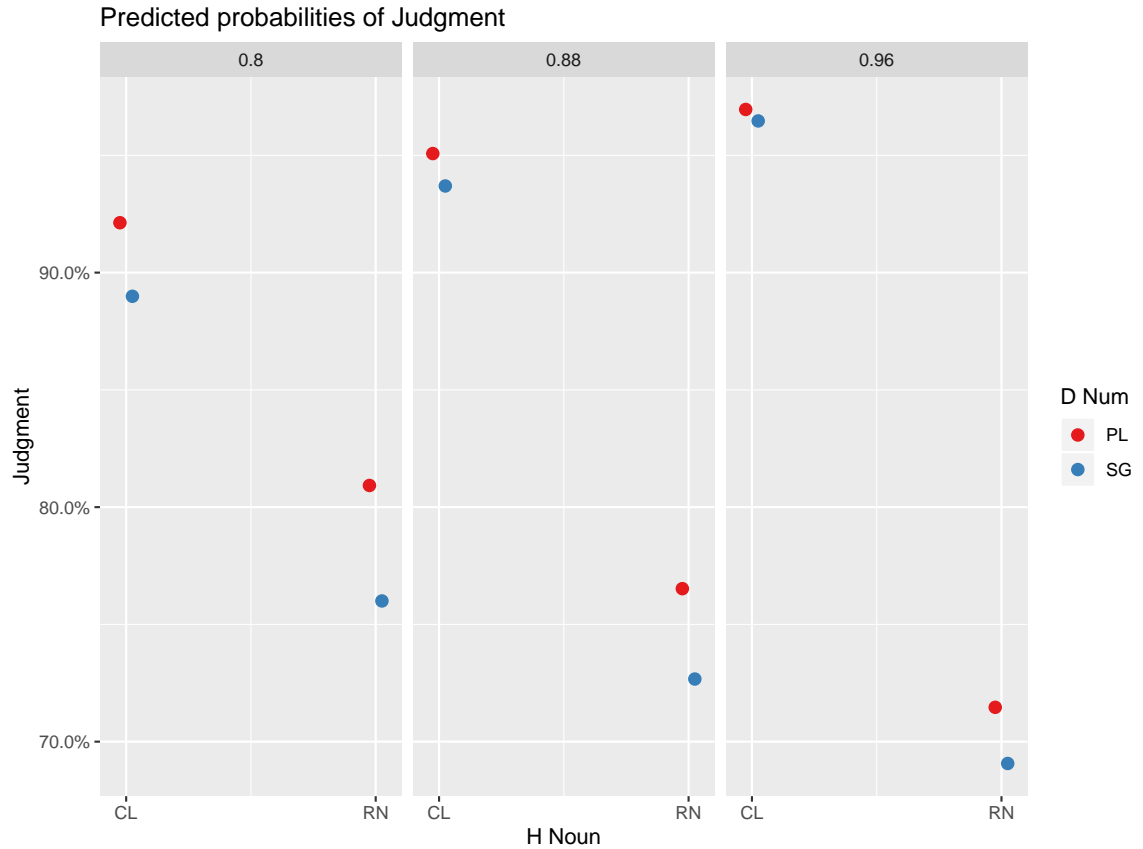


Figure 6.9: three-way interaction HNoun:DNum:Accu-Rate in Experiment 3

CL = collective noun head, RN = regular noun head, PL = plural distractor head, SG = singular distractor head

I explored the data a bit more by splitting the data into two sets, depending on whether participants provided their judgment relatively quickly or slowly, as I did for Experiment 1. The participants who had a mean reaction time (over all target items on the judgment task) that is smaller than the median mean reaction time of all participants are categorized as the fast participant group ($N = 63$) and the rest of the participants as the slow participant group ($N = 63$). I built the same GLMM models for the data from each of the groups.

The modeling results for the slow group do not differ much from the results of the overall model shown previously in Table 6.23, except that the intercept is also marginally significant. Given that there is still an effect of the interaction between

HNoun and DNum ($z = -3.324$, $p = 8e-4$), post-hoc comparisons are conducted to investigate the source of this interaction. However, the post-doc comparisons reveal no facilitatory effects in both the HC.DPL and HSG.DPL conditions.

For the fast participant group, by contrast, the results also indicate a marginal main effect of DNum ($z = 1.712$, $p = .087$). Results of the post-hoc comparisons, reported in Table 6.25, suggest that marginal facilitatory effects were detected in the HC.DPL condition, compared to the HC.DSG condition ($\chi^2(1) = 2.799$, $p = .094$) and in the HSG.DPL condition, compared to the HSG.DSG condition ($\chi^2(1) = 2.997$, $p = .083$).

Table 6.25: Planned comparisons for the judgment data from the fast group in Experiment 6

Comparison	Adj. Mean	Std. Error	df	Chisq	Pr(>Chisq)	
HC.DPL – HC.DSG	0.601	0.245	1	2.799	0.094	.
HSG.DPL – HSG.DSG	0.578	0.182	1	2.997	0.083	.
HC.DPL – HSG.DPL	0.805	0.237	1	36.070	1e-09	***
HC.DSG – HSG.DSG	0.790	0.220	1	36.286	2e-09	***

An immediate question is why (unstable) facilitatory effects were observed for both the HC.DPL and the HSG.DPL conditions only for the fast group. I leave this question to the General Discussion section, as I have to take the overall results from all the experiments (on reflexive binding) into consideration in order to address this question.

6.9 General discussion

The critical results from the six experiments are summarized as in Table 6.26, where SPR stands for self-paced reading tasks, and $SPR_{\text{refl/verb}}$ and SPR_{spill} signify the (critical) reflexive/linking verb region and the first spill-over region in SPR tasks. The acceptability judgment task with subject–verb agreement clearly revealed that

facilitatory effects were detected in the HC.DPL condition as well as the HSG.DPL condition. For reflexive binding, it is confirmed across Experiment 1, 2 and 6 that illusions of grammaticality/facilitatory effects were generally not observed with both the HC.DPL and HSG.DPL conditions. The only exception is Experiment 6, where marginal facilitatory effects were found in the HSG.DPL condition but not in the HC.DPL condition.

Table 6.26: Summary of the main findings of the experiments

Experiment	Facilitatory effects		Sig.Control
	HC.DPL	HSG.DPL	
Exp1:Bind_Judge	no	no	Frequency
Exp2:Bind_Judge	no	no	Phon-Neigh-Density
Exp4:Bind_SPR _{refl}	no	no	Orth-Length Orth-Neigh-Density
Exp4:Bind_SPR _{spill}	no	no	
Exp6:Bind_SPR _{refl}	no	no	Frequency Phon-Length
Exp6:Bind_SPR _{spill}	no	marginal ¹	
Exp6:Bind_Judge	slow: no	slow: no	Frequency
	fast: marginal ²	fast: marginal ³	Frequency
	combined: no	combined: marginal ⁴	Frequency
Exp3:Agree_Judge	yes	yes	Frequency
Exp5:Agree_SPR _{verb}	yes	no	Orth-Length Phon-Length
Exp5:Agree_SPR _{spill}	marginal ⁵	marginal ⁶	

¹ This effect is statistically marginal with $\chi^2(1) = 2.864$, $p = .091$.

² This effect is statistically marginal with $\chi^2(1) = 2.799$, $p = .094$.

³ This effect is statistically marginal with $\chi^2(1) = 2.997$, $p = .083$.

⁴ This effect is statistically marginal with $\chi^2(1) = 2.706$, $p = .010$.

⁵ This effect is statistically marginal with $\chi^2(1) = 3.599$, $p = .058$.

⁶ This effect is statistically marginal with $\chi^2(1) = 3.753$, $p = .053$.

For the self-paced reading experiment with subject–verb agreement, we found that facilitatory effects were clearly identified in the HC.DPL condition, and marginal

facilitatory effects were found in the HSG.DPL condition. For reflexive binding, generally no facilitatory effects were found. An exception is again Experiment 6, where we found marginal facilitatory effects in the HSG.DPL condition in the spill-over position.

It may be concluded that the results do not go along well with the predictions of my implementation of the Asymmetry of Interpretability Hypothesis under the cue-based retrieval model. This is because facilitatory effects were rarely observed for both the judgment and self-paced reading tasks with reflexive binding, especially under the HC.DPL condition. Facilitatory effects in the HC.DPL and the HSG.DPL condition were detected in the fast group in Experiment 6, but the effects were not always stable or were usually statistically marginal. By contrast, with the judgment task, facilitatory effects were found in both HC.DPL and HSG.DPL for subject–verb agreement.

Let us now address some important questions regarding the results. For reflexive binding, the most crucial question is why facilitatory effects generally did not occur in the HC.DPL condition, contrary to the prediction of the Asymmetry of Interpretability Hypothesis. The results from the three-way interaction between HNoun, DNoun and subject accuracy rate in reflexive binding does provide evidence for the Asymmetry of Interpretability Hypothesis. Such an interaction was explored in both Experiment 2 and 6. In Experiment 2, we found that, for the HC conditions, the predicted judgment are constantly very high (see Figure 6.3). In that case, facilitatory effects in the HC.DPL condition, compared to the HC.DSG condition, were absent. That is, given that both conditions are very acceptable, there are no facilitatory effects or illusions of grammaticality. By contrast, since the target sentences under the HSG.DPL condition were rated lower than the HC conditions, there is a higher possibility to observe facilitatory effects with this condition.

Above we suggested that facilitatory effects/illusions of grammaticality do not

occur with very acceptable sentences. However, this cannot explain the whole picture observed in the three-way interaction in Experiment 2 and 6. This is because in both Experiment 2 and 6, under the HSG.DPL condition, illusions of grammaticality disappeared or reduced for the participants whose accuracy rate was very high (above 90%). Figure 6.3 reveals a correlation between accuracy rate and the judgment scores given to the target sentences under the HSG conditions: when the accuracy rate is higher, the judgment scores are predicted to be lower. That is, the participants who had the highest accuracy rates are predicted to give the lowest judgment scores. When the accuracy rates are low, the facilitatory effects in the HSG.DPL condition are most prominent, and these effects are getting smaller for participants with a higher accuracy rate. The effects eventually disappeared for very accurate participants. At that point, both HSG conditions are rated very low. The interaction results therefore suggest that when the HSG conditions are both rated very low, facilitatory effects are less possible to be observed.

The observation on the three-way interactions with regard to the HSG conditions can be summarized in (34):

- (34) Facilitatory effects in the HSG.DPL condition in reflexive binding were less possible to occur with participants who had a higher accuracy rate.

As shown by the three-way interactions in Experiment 2 and 6, accuracy rate is predicted to be correlated with judgment scores: The more accurate a participant is, the lower judgment scores s/he would provide to the target sentences. In light of (34), I re-analyzed all the results from the judgment tasks by dividing each data set into two subsets: data from participants who provided a lower mean rating over all target sentences (the low-rating group), compared to the median ratings of all subjects, and data from participants who provided a higher mean rating (the high-rating group). The results are as shown in Table 6.27. Experiment 1 is an exception where no facilitatory effects were found even in the low-rating group. For all other reflexive

binding experiments, facilitatory effects were detected in the HSG.DPL condition only in the low-rating group, not in the high-rating group. Importantly, splitting into two groups based on the rating scores does not affect the emergence of facilitatory effects in subject–verb agreement.

Table 6.27: Facilitatory effects for the low-rating and high-rating groups in judgment experiments

Experiment	Group	Facilitatory effects		Accept-Rate
		HC.DPL	HSG.DPL	
Exp1:Bind_Judge	low	no	no	4.323
	high	no	no	5.986
Exp2:Bind_Judge	low	no	yes ¹	4.331
	high	no	no	5.819
Exp6:Bind_Judge	low	no	yes ²	63.2%
	high	no	no	94.1%
Exp3:Agree_Judge	low	yes ³	yes ⁴	3.415
	high	yes ⁵	yes ⁵	5.422

¹ $\chi^2(1) = 6.218$, $p = .013$.

² $\chi^2(1) = 4.044$, $p = .044$.

³ $\chi^2(1) = 28.271$, $p = 1.054\text{e-}07$.

⁴ $\chi^2(1) = 3.587$, $p = .058$.

⁵ $\chi^2(1) = 12.457$, $p = .0004$.

⁶ $\chi^2(1) = 31.766$, $p = 1.739\text{e-}08$.

The splitting of the participants into two sets can potentially separate participants who rated the HSG sentences such as (35) ungrammatical from those who rated them as grammatical. The reason that sentences such as (35) could be rated as grammatical is because *themselves* could be optionally interpreted as a singular, gender underspecified reflexive. According to this interpretation, *themselves* does not have a semantic number feature, and therefore, under the singular interpretation of *themselves*, there is no mismatch of semantic retrieval cues.

- (35) [DP₅ The local lawyer that supported [DP₆ the tough detectives]] embarrassed themselves with a recent allegation of serious corruption.

This singular usage of *themselves* is becoming popular among the younger generation. With these two types of participants separated, we could potentially detect grammatical illusions with the group of participants who did not accept sentences such as (35) as grammatical. These participants would, of course, give a lower rate to the HSG target sentences, consistent with the results in Table 6.27.

Why then splitting the data according to the median acceptance rate of all subjects did not help us detect facilitatory effects in the HC.DPL conditions, even in the low-rating groups? I think this could be explained by the grammaticality effects observed in all of these experiments. That is, the target sentences (or the critical words) in the HC conditions were judged higher or read faster than those in the HSG conditions. This difference between the HC and HSG conditions was observed consistently across experiments, which we referred to as grammaticality effects. This means that the target sentences under the HC.DPL and HC.DSG conditions were treated as grammatical sentences. For example, the target sentence in (36) might have been considered as a grammatical sentence and received a higher score in the acceptability judgment tasks.

(36) *Reflexive binding: collective head noun*

[_{DP1} The local government that supported [_{DP2} the tough detective(s)]] embarrassed themselves with a recent allegation of serious corruption.

This reminds us a well-established observation from previous studies, namely, the interaction between grammaticality and facilitatory effects. Many studies, including Pearlmutter et al. (1999), Wagers et al. (2009) and Dillon et al. (2013), find that facilitatory effects occur only in ungrammatical conditions, not in grammatical conditions. In terms of the experiments here, since the target sentences in the HC conditions in reflexive binding experiments were treated as grammatical sentences, they are rated significantly higher and read significantly faster than those in the HSG conditions across all experiments. Facilitatory effects are therefore not detected in

the grammatical conditions, as previous studies have shown.

The above account about why facilitatory effects were absent in the HC conditions in general leads us to another question: Why facilitatory effects occur in the subject–verb agreement constructions? The answer seems to be consistent with my assumption regarding the nature of the ϕ -features on the verb/T head: They are SM ϕ -features, not CI features. Therefore, for subject–verb agreement, only SM ϕ -features serve as retrieval cues. According to the Accurate Matching Hypothesis proposed in (18), such SM ϕ -retrieval cues can only match against corresponding SM features. Consequently, the following sentence in (37) should have been evaluated like an ungrammatical sentence in the experiments.⁹ The lower ratings (especially those from participants who had a higher accuracy rate) obtained in the subject–verb agreement experiments could then be explained.

(37) *Subject–verb agreement: collective head noun*

[_{DP3} The local government that supported [_{DP4} the tough detective(s)]] were embarrassed with a recent allegation of serious corruption.

Such explanation also accounts for the phenomenon that facilitatory effects in subject–verb agreement are not affected by the splitting of data into two groups based on the rating scores. This is because target mismatch in subject–verb agreement sentences are uniformly judged as ungrammatical (although individual differences exist on the actual scores participants gave to ungrammatical sentences). This is different from reflexive binding, where the participants include people from two knowledge systems: one group of the participants allows the singular use of *themselves*, and the other group does not.

Taken the results from the three-way interactions and those from the low-rating

⁹The facilitatory effects in the HC conditions seem to be more stably observed across both Experiment 3 and 5 with regard to subject–verb agreement; whereas only Experiment 3 exhibits facilitatory effects in the HSG.DPL condition. I do not have an explanation for this difference and I am suspicious about whether this is just noise, given that previous studies have consistently detected facilitatory effects in the HSG.DPL condition.

versus high-rating group, I conclude two situations that facilitatory effects are less likely to occur: (i) when the test sentences are judged grammatical/very acceptable, and (ii) when the sentence is judged very unacceptable. I believe that it is because of the mixture of these two factors that facilitatory effects/illusions of grammaticality were generally not observed in the reflexive binding experiments reported in this chapter.

What do the results say about the Asymmetry of Interpretability Hypothesis? I have suggested that the results are confounded and do not shed much light on this hypothesis because the target sentences were judged grammatical, at least for a subset of participants. Therefore, the target sentences used in the series of experiments in this chapter are not the most appropriate materials for us to test the Asymmetry of Interpretability Hypothesis. If the reflexive *themselves* can be treated either as a semantically plural, or as a semantically singular, gender underspecified reflexive, then the corresponding CI-Num retrieval cue would not be [CI-Num: PL]; instead, the retrieval cue should be [CI-Num: PL/SG], as *themselves* now is ambiguous. Consequently, the null results we obtained in the experiments cannot be used as counter-evidence against the Asymmetry of Interpretability Hypothesis. On the contrary, the findings with respect to the three-way interaction between HNoun, DNum and accuracy rate offers a (probably weak) argument for this hypothesis.

If the current interpretation of the experimental results are on the right track, the previously obtained null results on reflexive binding with regard to facilitatory effects also cannot be employed as evidence toward the syntactic filter approach or the weighting approach. This is because the null results seem to have nothing to do with the c-command requirement in reflexive binding; instead, it is still an issue related to the ϕ -features on reflexives, as the critical question here is whether *themselves* can be used as a semantically singular reflexive.

Another set of results from current experiments that disfavors the syntactic filter

approach is the significant control factors listed in Table 6.27. Recall that these factors are scaled differences between the target head noun and the distractor head noun. For instance, frequency difference is equal to the scaled difference between the frequency of the target head nouns and the distractor head nouns: $\text{Freq_Diff} = \text{Scale} (\text{Freq} (\text{target head noun}) - \text{Freq} (\text{distractor head noun}))$. In other words, the control factors reflect influence from both the distractor and the target head nouns. These factors being significant contributors to the variance of the dependent variable indicates that the distractor cannot be completely ignored in the processing of the reflexive. This argues against the syntactic filter approach as the syntactic filter approach predicts that the non-commanding distractor does not interfere with the target.

Finally, the last question I want to address is that whether the Asymmetry of Interpretability Hypothesis is no longer motivated if data from reflexive binding with *themselves* is excluded as a relevant empirical basis. The answer is negative. Note that the Asymmetry of Interpretability Hypothesis is built upon a large set of studies that are included in Jäger et al.'s (2017) meta-analysis, including studies that found inhibitory effects with retrieval cues for which the number feature is not important, e.g., Sturt (2003) (Experiment 1 and 2, gender feature), Cunnings & Felser (2013) (Experiment 2 for participants with low memory capacity, gender feature), Cunnings & Sturt (2014) (Experiment 1, gender feature), Jäger et al. (2015) (Experiment 1, animacy feature), and Parker & Phillips (2017) (Experiment 1 and 3 with 1-feature mismatch, gender feature). All these experiments do not use problematic *themselves* as the critical retrieval trigger, instead use other CI-interpretable features such as gender and animacy as the critical retrieval cues. Given that the Asymmetry of Interpretability Hypothesis applies to all CI-interpretable features, the hypothesis may explain why facilitatory effects were absent and why sometimes inhibitory effects were present in these studies.

6.10 Future direction

Feature studies would need to avoid using *themselves* as the critical retrieval trigger, since it is considered grammatical to refer to a singular antecedent by at least a subset of participants. To circumvent this problem, future studies could use gender feature to test the Asymmetry of Interpretability Hypothesis. The idea is similar to the experiment design proposed in this chapter. There are nouns that bear ambiguous gender features, to some extent similar to collective nouns which bear ambiguous number features. For instance, *librarian* is stereotypically feminine. However, it can also refer to a male antecedent. The stereotypical feature should thus be treated in a probabilistic sense. Suppose in language use *librarian* to refer to a female more than 90% of the time, the gender feature of *librarian* can be represented as in (38a), in contrast to the words *king* and *chairwoman*, which have only a masculine (38b) or a feminine gender (38c).

- (38) a. *librarian*: [CI-Gender: FEM (.90)/MAS (.10)]
b. *king*: [CI-Gender: MAS]
c. *chairwoman*: [CI-Gender: FEM]

Now we can use this difference to test the Asymmetry of Interpretability Hypothesis, with test sentences such as the following:

- (39) a. The librarian that supported the king embarrassed himself ...
b. The chairwoman that supported the king embarrassed himself ...

The Asymmetry of Interpretability Hypothesis predicts that facilitatory effects should be less likely observed with (39b), because (39b) involves a mismatch of the semantic gender retrieval cue. Facilitatory effects are more likely to be observed in (39a) compared to (39b), if participants considered (39a) unacceptable or rate it as lowly acceptable. On the other hand, if participants rate (39a) as highly acceptable, then

facilitatory effects should not be observed in (39a) either, due to the same reason why no facilitatory effects were not detected in the HC conditions in the experiments in this chapter. The difference between (39a) and (39b) was previously unnoticed, to the best of my knowledge. Studies such as Sturt (2003) use both types of target head nouns in their experiments, and this could be why facilitatory effects were not consistently observed.

Due to the limitations of time and many other practical issues (funding status, job, etc.), I am forced to leave the mentioned study to the future.

CHAPTER VII

General Conclusions

In this dissertation, I considered a linking theory between (i) some important aspects of the syntax and semantics of subject–verb agreement and reflexive binding and (ii) the processing of these two constructions with respect to illusions of grammaticality/facilitatory effects. This dissertation is motivated by a processing difference between these two constructions: Illusions of grammaticality/facilitatory effects were consistently observed in subject–verb agreement, but to a much less extent in reflexive binding. In order to account for such processing difference, this dissertation investigated the syntactic and semantic representations of these two constructions, aiming to find a crucial representational difference that may be relevant to the processing difference.

This dissertation made contributions to the syntactic and semantic analysis of subject–verb agreement and reflexive binding. In Chapter II, I argued that various types of agreement, including subject–verb agreement, can receive a unified syntactic analysis. Based on a formalization of the independent, freely available third-factor principle, i.e., minimal search, I argue that some intriguing agreement phenomena observed in different languages, e.g., (local) upward agree, multiple agree, cyclic agree and long-distance upward agree, can be better analyzed as instances of minimal search-based agree (MS-agree). MS-agree sheds light on the current debate on

the directionality of agree, by arguing that agree, with minimal search as its search algorithm, always searches downward. Attested cases of upward agree are in fact downward search from the next built phase above the head that bears unvalued features to be valued by agree.

My formalization of minimal search has several theoretical consequences. Labeling was for the first time formally defined, based on minimal search (MS-labeling). By comparing MS-agree and MS-labeling, it becomes clear that they cannot be completely reduced to minimal search, although the search algorithm that both of them employ is minimal search. In addition, I argue that $\langle \phi, \phi \rangle$ cannot be returned as a label for a finite TP set by MS-labeling. Instead, MS-labeling returns the D head of the SpecTP and the T head, rendering the label $\langle D, T \rangle$. This is a desirable result because the ϕ -features on T are SM features and should not be present at CI as part of the label for TP. Furthermore, MS-labeling provides an explanation for the invisibility of lower copies and solves several problems the Invisibility of Lower Copy Hypothesis faces.

In Chapter III I made the claim that reflexive binding can also be reduced to MS-agree. The main problem I addressed is the phenomenon that, in double object or dative constructions, either the subject or the first object can be the antecedent of the reflexive inside the second object. This optionality is not predicted by both the MS-agree analysis and Hicks's (2009) upward agree analysis. The MS-agree analysis predicts that only the subject can be the antecedent of the reflexive, since minimal search will search down from the vP phase and will find the subject before reaching the first object. By contrast, the upward agree analysis predicts that the first object is the only possible antecedent because upward search will return the first object before reaching the subject. Following Pesetsky (1995), I assume that the first and second object in double object and dative constructions are dominated by a PP. I further assume that PPs are optionally phases. If PP is a phase, then minimal search will

look down from the PP (since the PP is the next built phase above the reflexive) and find the first object. If PP is not a phase, minimal search will look down from the *v*P instead and return the *v*P-internal subject. The optionality phenomenon therefore receives an explanation. The MS-agree analysis for reflexive binding makes accurate predictions regarding the distribution of reflexives in general.

Taking the MS-agree analysis for the syntax of subject–verb agreement and reflexive binding together, I concluded that the syntactic dependency in these two constructions do not differ from each other in an essential way. Therefore, to account for the processing differences between these two constructions, any theory that is built upon syntactic differences between these two constructions (e.g., the weighting approach and the syntax filter approach) loses its theoretical basis. We should thus look for other representational differences between these two constructions to account for the processing difference.

Chapter IV serves this purpose by pointing out that the ϕ -features involved in subject–verb agreement and reflexive binding are different. Specifically, the ϕ -features on verbs/T are semantically uninterpretable, because they are phonological features. However, the ϕ -features on reflexives are semantically interpretable. I built the argument in the context of an important debate in the field of semantics: Whether the ϕ -features on bound variables (including bound reflexives) are transmitted from their binder at the phonological interface. This discussion is relevant to the overall project because if the ϕ -features on reflexives are transmitted from their binders at the phonological interface, then the ϕ -features on the bound reflexives are phonological features, similar to those on verbs. Consequently, we cannot explain the processing differences between subject–verb agreement and reflexive binding using the difference of semantic interpretability between these two constructions.

I argue that collective nominal binders in English and Chinese suggest clearly that the ϕ -features (the number feature specifically) on bound reflexives and pronouns

are semantically interpretable. If they are semantically uninterpretable and are in fact phonological features that are transmitted from their binders, then we expect that when the binder is a collective nominal, only the phonological ϕ -features of the collective nominal can be transmitted to the reflexive. Given that collective nominals bear only a singular phonological number feature, the variable they bound must also be singular. However, we found that in both English and Chinese, the bound variable could be either singular or plural. Crucially, when the bound variable is singular, the collective nominal binder must be interpreted as a singular group; whereas when the bound variable is plural, the collective nominal binder must be interpreted as plural group members. Since collective nouns are semantically ambiguous between a singular group interpretation and a plural member interpretation, the following explanation seems quite straightforward: The ϕ -features on the bound variable are semantically interpretable, and these semantic ϕ -features on the bound variable match with either the singular or the plural semantic interpretations of the collective nominal binder. Chinese data further confirms that when the ϕ -features are underspecified on the bound variable, the plural member reading on a collective nominal binder is not accessible. That is, the default semantic interpretation of a collective nominal is the singular group reading. The plural member reading is only elicited when the bound variable is plural, which imposes a “member of” operation on the semantics of collective nominal binders. This further supports the argument that the plural number feature on bound variables in Chinese is semantically interpretable.

Another argument was built upon bound variables with collective nominal binders in Arabic and Russian, languages that are morphologically rich. I showed that in Arabic, both the number and gender features are semantically interpretable, and they must be consistent with their reference. In Russian, possessive pronouns and reflexives bound by collective nominal binders bear a phonological gender feature, besides its semantic number features. The gender feature is considered phonological

because it does not interact with the interpretation of the bound variables. I provided evidence suggesting that the semantic number features on bound variables must be consistent with their reference; however, the phonological gender feature is acquired through a local agreement relation with the bound variables' syntactic complement. The binding relation with the collective nominal binder thus has nothing to do with the acquisition of the phonological gender feature on the bound variable in Russian, contrary to the prediction of the morphosyntactic approach.

I then related the interpretability difference between subject–verb agreement and reflexive binding to the processing difference between these two constructions, proposing an account for the processing difference based on the interpretability difference under the cue-based retrieval theory. However, there is another issue that we must deal with first. The problem is that long-distance syntactic dependencies such as reflexive binding cannot be implemented in cue-based retrieval theory in the first place, because the c-command and locality requirements in reflexive binding are non-local syntactic relations. Non-local syntactic relations are not features that can be directly usable in the cue-based retrieval theory. To encode non-local syntactic relations as features, in Chapter V I proposed the derivational approach of c-commanded-by relations, built upon Epstein et al.'s (1998) derivational approach to c-command relations. The leading idea is to encode the c-command feature on the c-commandees rather than the c-commanders. Following Epstein et al.'s (1998) derivational approach, I reduced the c-commanded-by relations to two local relations: co-membership and containment. I then propose two operations, Pass and Inherit, to compute the [C-commanded-by] feature on c-commandees: Pass is to add the chunk ID of a member to its co-member's [C-commanded-by] feature, dealing with co-membership. Inherit is to copy the value of the [C-commanded-by] feature from a mother chunk/node to its child members, dealing with the containment relation. I further simplified the computation by restricting the Pass operation to only possible antecedents in vari-

ous constructions where c-command is claimed to be important. Finally, I extended the current encoding method to compute the locality feature ([Local]). After the [C-commanded-by] and [Local] features are successfully computed, the content of these features can serve as retrieval cues to access the target binder/antecedent.

Finally, in Chapter VI, I proposed the Asymmetry of Interpretability Hypothesis to explain why the processing difference with respect to the illusions of grammaticality/facilitatory effects should exist between subject–verb agreement and reflexive binding. The Asymmetry of Interpretability Hypothesis assumes that mismatch against a semantically interpretable cue is costly, and is more likely to be detected by the parser. This hypothesis, coupled with the semantic analysis of the ϕ -features involved in subject–verb agreement and reflexive binding, correctly predicts that illusions of grammaticality/facilitatory effects should be less likely to be observed with reflexive binding, compared to subject–verb agreement, as reflexive binding involves a mismatch of semantic retrieval cues. I tested the Asymmetry of Interpretability Hypothesis with results from both judgment and self-paced reading experiments with collective nouns as the target head nouns. The Asymmetry of Interpretability Hypothesis predicts that there should not be a mismatch of a semantic retrieval cue with the plural reflexive “themselves,” different from when the target head noun is a regular singular noun. This is because collective nouns are ambiguous in their semantic number features: They can be either singular or plural semantically. The results indicated that, even with collective target head nouns, no stable illusions of grammaticality or facilitatory effects were found in reflexive binding, in contrast to subject–verb agreement, where such effects were detected. Further exploration of the judgment data revealed that there was a correlation between participants’ judgment scores and the presence of illusions of grammaticality: Illusions of grammaticality occurred only among the participants who gave a lower rating to all the target sentences. This means only when participants rated the target sentences as relatively unacceptable,

illusions of grammaticality appeared. This correlation was then interpreted in light of a previous finding: Illusions of grammaticality/facilitatory effects were observed only for ungrammatical sentences. The results from the self-paced reading experiments with subject–verb agreement confirmed the above conclusion. Participants whose accuracy rate was lower tended to give a higher rating to the target sentences, and for these participants, facilitatory effects tended to be less likely observed for both of the conditions where the target head noun is either a singular collective noun or a singular regular noun. Similar phenomenon was observed for participants whose accurate rate was higher, as they tended to judge all the target sentences less acceptable, thus concealing the differences between the distractor match condition and the distractor mismatch condition, no matter whether the target head noun is a collective noun or a regular noun.

The complexity of the ϕ -features on the reflexive “themselves” suggests an important reason why illusions of grammaticality/facilitatory effects were not observed with reflexive binding even when the head noun is a collective nominal. This is because “themselves” can be used as a singular, gender neutral reflexive. Dillon et al. (2013) did an offline acceptability judgment study which revealed that “themselves” bound by a singular antecedent was not acceptable to their participants. But Dillon et al.’s study has a small sample size (12 participants). The results from the acceptability judgment experiments in Chapter VI, with much bigger sample sizes, cast doubt on Dillon et al.’s conclusion: The target sentences in reflexive binding were generally rated higher than those in subject–verb agreement. Therefore, I concluded that reflexive binding with “themselves” are not proper materials to test the Asymmetry of Interpretability Hypothesis.

I proposed that gender feature mismatch may provide a more appropriate testing group for the Asymmetry of Interpretability Hypothesis. I leave this to future studies due to time and other limitations.

BIBLIOGRAPHY

- Abels, Klaus. 2003. Successive cyclicity, anti-locality, and adposition stranding: University of Connecticut Dissertation.
- Abney, Steven P. 1987. The English noun phrase in its sentential aspect: Massachusetts Institute of Technology Dissertation.
- Adger, David. 2003. Core syntax: A minimalist approach, vol. 20. Oxford: Oxford University Press.
- Adger, David. 2010. A minimalist theory of feature structure. In Anna Kibort & Greville Corbett (eds.), Features, 185–218. Oxford: Oxford University Press.
- Adger, David & Peter Svenonius. 2011. Features in minimalist syntax. In Cedric Boeckx (ed.), The Oxford handbook of linguistic minimalism, 27–51. Oxford: Oxford University Press.
- Alcocer, Pedro & Colin Phillips. 2012. Using relational syntactic constraints in content-addressable memory architectures for sentence parsing. Unpublished. http://www.colinphillips.net/wp-content/uploads/2014/08/alcocer_phillips2012_v2.pdf.
- Anderson, John R. 2005. Human symbol manipulation within an integrated cognitive architecture. Cognitive Science 29(3). 313–341. doi:http://dx.doi.org/10.1207/s15516709cog0000_22.
- Anderson, John R. & Christian J. Lebiere. 1998. The atomic components of thought. New York, NY: Psychology Press.
- Andrews, Caroline, Anthony Yacovone, Shayne Sloggett & Brian Dillon. 2016. Reflexives: We don't see the attraction. Poster at Architectures and Mechanisms for Language Processing (AMLaP 22), Bilbao, Spain.
- Antonenko, Andrei. 2012. Feature-based binding and phase theory: Stony Brook University Dissertation.
- Arregi, Karlos & Andrew Nevins. 2012. Morphotactics: Basque auxiliaries and the structure of spellout, vol. 86. Springer.
- Baayen, Harald, Richard Piepenbrock & Hedderick van Rijn. 1993. The CELEX lexical database. Linguistic Data Consortium. Philadelphia, PA .

- Badecker, William & Kathleen Straub. 2002. The processing role of structural constraints on interpretation of pronouns and anaphors. Journal of Experimental Psychology: Learning, Memory, and Cognition 28(4). 748–769. doi:10.1037/0278-7393.28.4.748.
- Baker, Mark C. 2008. The syntax of agreement and concord, vol. 115. Cambridge: Cambridge University Press.
- Barker, Chris. 2012. Quantificational binding does not require c-command. Linguistic Inquiry 43(4). 614–633. doi:https://doi.org/10.1162/ling_a_00108.
- Barker, Jason, Janet Nicol & Merrill Garrett. 2001. Semantic factors in the production of number agreement. Journal of Psycholinguistic Research 30(1). 91–114. doi:10.1023/a:1005208308278.
- Barss, Andrew & Howard Lasnik. 1986. A note on anaphora and double objects. Linguistic Inquiry 17(2). 347–354. <http://www.jstor.org/stable/4178491>.
- Bassi, Itai & Nicholas Longenbaugh. 2017. Features on bound pronouns: An argument for a semantic approach. In Proceedings of the 48th Annual Meeting of the North East Linguistic Society (NELS 48), .
- Bassi, Itai & Nicholas Longenbaugh. to appear. Against tanglewood by focus movement: A reply to erlewine and kotek (2018). Linguistic Inquiry 1–28. doi:https://doi.org/10.1162/ling_a_00348.
- Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. Journal of Statistical Software .
- Beck, Sigrid & Kyle Johnson. 2004. Double objects again. Linguistic Inquiry 35(1). 97–123. doi:https://doi.org/10.1162/002438904322793356.
- Béjar, Susana. 2003. Phi-syntax: A theory of agreement: University of Toronto Dissertation.
- Béjar, Susana & Milan Rezac. 2003. Person licensing and the derivation of PCC effects. In Ana Teresa Pérez-Leroux & Yves Roberge (eds.), Romance linguistics: Theory and acquisition: Selected papers from the 32nd Linguistic Symposium on Romance Languages (LSRL), vol. 4 Amsterdam Studies in the Theory and History of Linguistic Science, 49–62. Amsterdam/Philadelphia: John Benjamins Publishing.
- Béjar, Susana & Milan Rezac. 2009. Cyclic agree. Linguistic Inquiry 40(1). 35–73. https://muse.jhu.edu/journals/linguistic_inquiry/v040/40.1.bejar.html.
- Bhatt, Rajesh & Martin Walkow. 2013. Locating agreement in grammar: An argument from agreement in conjunctions. Natural Language & Linguistic Theory 31(4). 951–1013. doi:https://doi.org/10.1007/s11049-013-9203-y.

- Bjorkman, Bronwyn M. & Hedde Zeijlstra. to appear. Checking up on (ϕ -)agree. *Linguistic Inquiry* 10. 1–60. doi:https://doi.org/10.1162/ling_a_00319.
- Bobaljik, Jonathan David. 2008. Where's phi. In Daniel Harbour, David Adger & Susana Béjar (eds.), *Phi-theory: Phi-features across modules and interfaces* Oxford Studies in Theoretical Linguistics, 295–328. Oxford, UK: Oxford University Press.
- Bock, Kathryn, Sally Butterfield, Anne Cutler, J. Cooper Cutting, Kathleen M. Eberhard & Karin R. Humphreys. 2006. Number agreement in British and American English: Disagreeing to agree collectively. *Language* 82(1). 64–113. <https://muse.jhu.edu/article/194941>.
- Bock, Kathryn, Kathleen M. Eberhard & J. Cooper Cutting. 2004. Producing number agreement: How pronouns equal verbs. *Journal of Memory and Language* 51(2). 251–278. doi:<http://dx.doi.org/10.1016/j.jml.2004.04.005>. <http://www.sciencedirect.com/science/article/pii/S0749596X04000476>.
- Bock, Kathryn, Kathleen M. Eberhard, J. Cooper Cutting, Antje S. Meyer & Herbert Schriefers. 2001. Some attractions of verb agreement. *Cognitive Psychology* 43(2). 83–128. doi:<http://doi.org/10.1006/cogp.2001.0753>. <http://www.sciencedirect.com/science/article/pii/S0010028501907534>.
- Bock, Kathryn & Carol A. Miller. 1991. Broken agreement. *Cognitive Psychology* 23(1). 45–93. doi:[http://dx.doi.org/10.1016/0010-0285\(91\)90003-7](http://dx.doi.org/10.1016/0010-0285(91)90003-7). <http://www.sciencedirect.com/science/article/pii/0010028591900037>.
- Boeckx, Cedric. 2008. *Aspects of the syntax of agreement*. New York, NY: Routledge.
- Bošković, Željko. 2014. Now I'm a phase, now I'm not a phase: On the variability of phases with extraction and ellipsis. *Linguistic Inquiry* 45(1). 27–89. <https://muse.jhu.edu/article/535449>.
- Bošković, Željko & Jairo Nunes. 2007. The copy theory of movement: A review from PF. In Norbert Corver & Jairo Nunes (eds.), *The copy theory of movement*, vol. 107, 13–74. Amsterdam/Philadelphia: John Benjamins Publishing.
- Branigan, Phil & Marguerite MacKenzie. 2002. Altruism, \bar{A} -movement, and object agreement in Innu-aimûn. *Linguistic Inquiry* 33(3). 385–407. doi:<https://doi.org/10.1162/002438902760168545>.
- Bruening, Benjamin. 2001. *Syntax at the edge: Cross-clausal phenomena and the syntax of Passamaquoddy*: Massachusetts Institute of Technology Dissertation.
- Bruening, Benjamin. 2014. Precede-and-command revisited. *Language* 90(2). 342–388. doi:10.1353/lan.2014.0037.
- Carstens, Vicki. 2016. Delayed valuation: A reanalysis of goal features, “upward” complementizer agreement, and the mechanics of case. *Syntax* 19(1). 1–42. doi:<https://doi.org/10.1111/synt.12116>.

- Carstens, Vicki & Michael Diercks. 2013. Parameterizing case and activity: Hyper-raising in Bantu. In Seda Kan, Claire Moore-Cantwell & Robert Staubs (eds.), Proceedings of the 40th Annual Meeting of the North East Linguistic Society (NELS 40), 99–118. Graduate Linguistic Student Association (GLSA) Publications.
- Cecchetto, Carlo. 2001. Proper binding condition effects are phase impenetrability condition effects. In Min-Joo Kim (ed.), Proceedings of the 31th Annual Meeting of the North East Linguistic Society (NELS 31), vol. 31, 99–116. Graduate Linguistic Student Association (GLSA) Publications.
- Chametzky, Robert. 1996. A theory of phrase markers and the extended base. Albany: State University of New York Press.
- Charnavel, Isabelle. 2018a. Apparent exemption from condition A: a perspective-based theory. Manuscript. [Http://ling.auf.net/lingbuzz/002683](http://ling.auf.net/lingbuzz/002683).
- Charnavel, Isabelle. 2018b. Presupposition failure and intended pronominal reference: person is not so different from gender after all. Linguistic Inquiry 1–35. doi: 10.1162/ling_a_00307. https://doi.org/10.1162/ling_a_00307.
- Charnavel, Isabelle, James C.-T. Huang, Peter Cole & Gabriella Hermon. 2017. Long-distance anaphora: syntax and discourse. In Martin Everaert & Henk C. van Riemsdijk (eds.), The Wiley Blackwell Companion to Syntax, Hoboken, NJ: John Wiley & Sons, Inc. 2nd edn. doi:<http://dx.doi.org/10.1002/9781118358733.wbsyncom074>.
- Charnavel, Isabelle & Dominique Sportiche. 2016. Anaphor binding: What French inanimate anaphors show. Linguistic Inquiry 47(1). 35–87. doi:https://doi.org/10.1162/LING_a_00204.
- Chen, Zhong, Lena Jäger & Shravan Vasisht. 2012. How structure-sensitive is the parser? Evidence from Mandarin Chinese. In Britta Stalterfoht & Sam Featherston (eds.), Empirical approaches to linguistic theory: Studies in meaning and structure Studies in Generative Grammar, 43–62. Berlin: Mouton de Gruyter.
- Chierchia, Gennaro. 2006. Broaden your views: Implicatures of domain widening and the “logicality” of language. Linguistic Inquiry 37(4). 535–590. doi:<https://doi.org/10.1162/ling.2006.37.4.535>.
- Chomsky, Noam. 1981. Lectures on government and binding. Dordrecht, Netherlands: Foris.
- Chomsky, Noam. 1986. Knowledge of language: Its nature, origin, and use. New York: Praeger Publishers.
- Chomsky, Noam. 1995a. Bare phrase structure. In Héctor Campos & Paula Kempchinsky (eds.), Evolution and revolution in linguistic theory, 51–109. Washington, DC: Georgetown U Press.

- Chomsky, Noam. 1995b. The minimalist program. Cambridge, MA: MIT Press.
- Chomsky, Noam. 1999. Derivation by phase, vol. 18. Cambridge, MA: MIT Occasional Papers in Linguistics.
- Chomsky, Noam. 2000a. Minimalist inquiries: The framework. In Roger Martin, David Michaels & Juan Uriagereka (eds.), Step by step: Essays on minimalist syntax in honor of Howard Lasnik, 89–155. Cambridge, MA: MIT Press.
- Chomsky, Noam. 2000b. Noam Chomsky: The architecture of language. Oxford: Oxford University Press.
- Chomsky, Noam. 2001a. Beyond explanatory adequacy. In Michael Kenstowicz (ed.), MIT Occasional Papers in Linguistics 20, 1–52. Cambridge, MA: MIT Working Papers in Linguistics (MITWPL).
- Chomsky, Noam. 2001b. Derivation by phase. In Michael Kenstowicz (ed.), Ken Hale: A life in language, 1–52. Cambridge, MA: MIT Press.
- Chomsky, Noam. 2004. Beyond explanatory adequacy. In Adriana Belletti (ed.), Structures and beyond, vol. The Cartography of Syntactic Structures, 104–131. Oxford: Oxford University Press.
- Chomsky, Noam. 2005. Three factors in language design. Linguistic Inquiry 36(1). 1–22.
- Chomsky, Noam. 2008. On phases. In Robert Freidin, Carlos P. Otero & Maria Luisa Zubizarreta (eds.), Foundational issues in linguistic theory: Essays in honor of Jean-Roger Vergnaud, 133–166. Cambridge, MA: MIT Press.
- Chomsky, Noam. 2013. Problems of projection. Lingua 130. 33–49.
- Chomsky, Noam. 2015a. A discussion with Naoki Fukui and Mihoko Zushi (march 4, 2014). Sophia Linguistica: Working Papers in Linguistics (64). 69–97. <http://digital-archives.sophia.ac.jp/repository/view/repository/00000036046>.
- Chomsky, Noam. 2015b. Problems of projection: Extensions. In Elisa Di Domenico, Cornelia Hamann & Simona Matteini (eds.), Structures, strategies and beyond : Studies in honour of Adriana Belletti, vol. 223, book section 1, 3–16. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Chomsky, Noam. 2017. Generative linguistics in the 21st century: The evidence and the rhetoric. <https://www.facebook.com/theuniversityofreading/videos/1115026661937460/>.
- Chomsky, Noam, Ángel Gallego & Dennis Ott. 2017. Generative grammar and the faculty of language: insights, questions, and challenges. To appear in Catalan Journal of Linguistics <http://ling.auf.net/lingbuzz/003507>.

- Chomsky, Noam & Howard Lasnik. 1993. The theory of principles and parameters. In J. Jacobs, A. von Stechow, W. Sternefeld & T. Vennemann (eds.), Syntax: An international handbook of contemporary research, vol. 1, 506–569. Berlin, Germany: de Gruyter.
- Choy, Jungwon Janet & Cynthia K. Thompson. 2010. Binding in agrammatic aphasia: Processing to comprehension. Aphasiology 24(5). 551–579. doi:<http://dx.doi.org/10.1080/02687030802634025>.
- Cinque, Guglielmo. 2010. Mapping spatial pps: an introduction. In Guglielmo Cinque & Luigi Rizzi (eds.), Mapping spatial pps The Cartography of Syntactic Structures, 3–25. Oxford: Oxford University Press.
- Cinque, Guglielmo & Luigi Rizzi. 2010. Mapping spatial pps: The cartography of syntactic structures, vol. 6. Oxford: Oxford University Press.
- Citko, Barbara. 2014. Phase theory: An introduction. New York, NY: Cambridge University Press.
- Clackson, Kaili, Claudia Felser & Harald Clahsen. 2011. Children’s processing of reflexives and pronouns in English: Evidence from eye-movements during listening. Journal of Memory and Language 65(2). 128–144. doi:<http://dx.doi.org/10.1016/j.jml.2011.04.007>. <http://www.sciencedirect.com/science/article/pii/S0749596X11000465>.
- Clackson, Kaili & Vera Heyer. 2014. Reflexive anaphor resolution in spoken language comprehension: Structural constraints and beyond. Frontiers in Psychology 5.
- Clifton, Charles Jr., Lyn Frazier & Patricia Deevy. 1999. Feature manipulation in sentence comprehension. Rivista di Linguistica 11(1). 11–39.
- Cole, Peter, Gabriella Hermon & Li-May Sung. 1990. Principles and parameters of long-distance reflexives. Linguistic Inquiry 21(1). 1–22. doi:10.2307/4178657. <http://www.jstor.org/stable/4178657>.
- Cole, Peter & Li-May Sung. 1994. Head movement and long-distance reflexives. Linguistic Inquiry 25(3). 355–406. <http://www.jstor.org/stable/4178868>.
- Collins, Chris. 1994. Economy of derivation and the generalized proper binding condition. Linguistic Inquiry 25(1). 45–61. <http://www.jstor.org.proxy.lib.umich.edu/stable/4178848>.
- Collins, Chris & Erich Groat. 2018. Distinguishing copies and repetitions. Manuscript. <https://ling.auf.net/lingbuzz/003809/current.pdf>.
- Collins, Chris & Edward Stabler. 2016. A formalization of minimalist syntax. Syntax 19(1). 43–78. doi:<https://dx.doi.org/10.1111/synt.12117>.

- Cooper, Robin. 1979. The interpretation of pronouns. In Frank Heny & Helmut S. Schnelle (eds.), *Syntax and semantics: Selections from the 3rd Groningen Round Table*, vol. 10, 61–92. New York, NY: Academic Press.
- Cooper, Robin. 1983. *Quantification and syntactic theory*. Dordrecht, Holland: Reidel.
- Corbett, Greville G. 1979. The agreement hierarchy. *Journal of Linguistics* 15(2). 203–224. doi:10.1017/S0022226700016352. <https://www.cambridge.org/core/article/agreement-hierarchy/400E4AEDE60D4CD3CDB5AAF0FFD22675>.
- Corbett, Greville G. 2006. *Agreement*, vol. 109 Cambridge Textbooks in Linguistics. Cambridge, UK: Cambridge University Press.
- Cummings, Ian & Claudia Felser. 2013. The role of working memory in the processing of reflexives. *Language and Cognitive Processes* 28(1-2). 188–219. doi:http://dx.doi.org/10.1080/01690965.2010.548391.
- Cummings, Ian, Clare Patterson & Claudia Felser. 2015. Structural constraints on pronoun binding and coreference: Evidence from eye movements during reading. *Frontiers in Psychology* 6. 840. doi:10.3389/fpsyg.2015.00840. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4477054/>.
- Cummings, Ian & Patrick Sturt. 2014. Coargumenthood and the processing of reflexives. *Journal of Memory and Language* 75. 117–139. doi:http://dx.doi.org/10.1016/j.jml.2014.05.006. <http://www.sciencedirect.com/science/article/pii/S0749596X14000552>.
- Depraetere, Ilse. 2003. On verbal concord with collective nouns in British English. *English Language & Linguistics* 7(1). 85–127.
- Despić, Miloje. 2015. Phases, reflexives, and definiteness. *Syntax* 18(3). 201–234. doi:http://dx.doi.org/10.1111/synt.12031.
- Diercks, Michael. 2013. Indirect agree in Lubukusu complementizer agreement. *Natural Language & Linguistic Theory* 31(2). 357–407. doi:https://doi.org/10.1007/s11049-013-9187-7.
- Diercks, Michael James Kroening. 2010. *Agreement with subjects in Lubukusu*: Georgetown University Dissertation.
- den Dikken, Marcel & Eric Hoekstra. 1997. Parasitic participles. *Linguistics* 35(6). 1057–1089. doi:10.1515/ling.1997.35.6.1057. <https://www.degruyter.com/view/j/ling.1997.35.issue-6/ling.1997.35.6.1057/ling.1997.35.6.1057.xml>.
- Dillon, Brian, Wing Yee Chow, Matthew Wagers, Taomei Guo, Fengqin Liu & Colin Phillips. 2014. The structure-sensitivity of memory access: Evidence from Mandarin Chinese. *Frontiers in Psychology* 5. 1025. doi:https://doi.org/10.3389/fpsyg.2014.01025.

- Dillon, Brian, Alan Mishler, Shayne Sloggett & Colin Phillips. 2013. Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence. *Journal of Memory and Language* 69(2). 85–103. doi:<http://dx.doi.org/10.1016/j.jml.2013.04.003>. <http://www.sciencedirect.com/science/article/pii/S0749596X13000338>.
- Dorie, Vincent. 2015. *blme: Bayesian linear mixed-effects models* 1st edn. <https://github.com/vdorie/blme>.
- Drummond, Alex, Dave Kush & Norbert Hornstein. 2011. Minimalist construal: Two approaches to A and B. In Cedric Boeckx (ed.), *The Oxford Handbook of Linguistic Minimalism*, 396–426. Oxford: Oxford University Press.
- Van Dyke, Julie A. & Richard L. Lewis. 2003. Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language* 49(3). 285–316. doi:[http://dx.doi.org/10.1016/S0749-596X\(03\)00081-0](http://dx.doi.org/10.1016/S0749-596X(03)00081-0).
- Elbourne, Paul D. 2005. *Situations and individuals*, vol. 90. Cambridge, MA: MIT Press.
- Epstein, Samuel David. 1999. Un-principled syntax: The derivation of syntactic relations. In Samuel David Epstein & Norbert Hornstein (eds.), *Working minimalism*, 317–345. Cambridge, MA: MIT Press.
- Epstein, Samuel David. 2007. On i(nternalist) functional explanation in minimalism. *Linguistic Analysis* 33(1–2). 20–53.
- Epstein, Samuel David, Erich M. Groat, Ruriko Kawashima & Hisatsugu Kitahara. 1998. *A derivational approach to syntactic relations*. New York: Oxford University Press.
- Epstein, Samuel David, Hisatsugu Kitahara & T Daniel Seely. 2014. Labeling by minimal search: Implications for successive-cyclic a-movement and the conception of the postulate “phase”. *Linguistic Inquiry* 45(3). 463–481. https://muse.jhu.edu/journals/linguistic_inquiry/v045/45.3.epstein.html.
- Fiengo, Robert. 1977. On trace theory. *Linguistic Inquiry* 8(1). 35–61. <http://www.jstor.org.proxy.lib.umich.edu/stable/4177972>.
- Fukui, Naoki. 1986. *A theory of category projection and its applications*: Massachusetts Institute of Technology Dissertation.
- Giannakidou, Anastasia. 1999. Affective dependencies. *Linguistics and Philosophy* 22(4). 367–421.

- Giannakidou, Anastasia. 2011. Negative and positive polarity items. In Claudia Maienborn, Klaus von Stechow & Paul Portner (eds.), Semantics: An international handbook of natural language meaning, vol. 3, 1660–1712. Berlin: Mouton de Gruyter.
- Hasegawa, Hiroshi. 2005. Reflexive binding as agreement and its interactions with the phase system. In Noriko Imanishi (ed.), The world of linguistic research: A festschrift for Kinsuke Hasegawa on the occasion of his seventieth birthday, 53–69. Tokyo: Kaitakusha.
- Heim, Irene. 1990. E-type pronouns and donkey anaphora. Linguistics and Philosophy 13(2). 137–177. doi:<https://doi.org/10.1007/BF00630732>.
- Heim, Irene. 1997. Predicates or formulas? Evidence from ellipsis. In Proceedings of the 7th Semantics and Linguistic Theory Conference (SALT 7), 197–221.
- Heim, Irene. 2005. Features on bound pronouns. Unpublished manuscript.
- Heim, Irene. 2008. Features on bound pronouns. In Daniel Harbour, David Adger & Susana Béjar (eds.), Phi theory: Phi-features across modules and interfaces, 35–56. Oxford, United Kingdom: Oxford University Press.
- Heim, Irene & Angelika Kratzer. 1998. Semantics in generative grammar Blackwell textbooks in linguistics. Malden, MA: Blackwell.
- Heinat, Fredrik. 2006. Probes, pronouns, and binding in the minimalist program: Lund University Dissertation.
- Heinat, Fredrik. 2008. Probes, pronouns and binding in the minimalist program. VDM Verlag.
- Hestvik, Arild & William Philip. 2001. Syntactic vs. logophoric binding: Evidence from Norwegian child language. In Peter Cole, Gabriella Hermon & James C.-T. Huang (eds.), Long distance reflexives, vol. 33 Syntax and Semantics, 119–139. San Diego: Academic Press.
- Hewitt, Brian George. 1995. Georgian: A structural reference grammar, vol. 2. Amsterdam/Philadelphia: John Benjamins Publishing.
- Hicks, Glyn. 2009. The derivation of anaphoric relations, vol. 139 Linguistics Today. Amsterdam: John Benjamins Publishing.
- Hiraiwa, Ken. 2001. Multiple agree and the defective intervention constraint in Japanese. In MIT Working Papers in Linguistics, vol. 40, 67–80. Cambridge, MA: MITWPL.
- Hiraiwa, Ken. 2003. Movement and derivation: Eliminating the PBC. In Proceedings of the 26th Penn Linguistics Colloquium, vol. 9, 89–102. UPenn Working Papers in Linguistics.

- Hirschberg, Julia & Gregory Ward. 1991. Accent and bound anaphora. Cognitive Linguistics 2(2). 101–122.
- Hornstein, Norbert. 2001. Move!: A minimalist theory of construal. Oxford: Blackwell.
- Hornstein, Norbert. 2007. Pronouns in a minimalist setting. In Norbert Corver & Jairo Nunes (eds.), The copy theory of movement, vol. 107 *Linguistics Today*, 385–351. Amsterdam/Philadelphia: John Benjamins.
- Huang, James C.-T., Audrey Yen-hui Li & Yafei Li. 2009. The syntax of Chinese. Cambridge, UK: Cambridge University Press.
- Huang, James C.-T. & Luther C.-S. Liu. 2001. Logophoricity, attitudes, and *ziji* at the interface. In Peter Cole, Gabriella Hermon & James C.-T. Huang (eds.), Long-distance reflexives, vol. 33 *Syntax and Semantics*, 141–195. San Diego, CA: Academic Press.
- Huang, James C.-T. & Jane Chih-Chen Tang. 1991. The local nature of the long-distance reflexive in Chinese. In Jan Koster & Eric Reuland (eds.), Long distance anaphora, 263–282. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511627835.014.
- Humphreys, Karin R. & Kathryn Bock. 2005. Notional number agreement in English. Psychonomic Bulletin & Review 12(4). 689–695. doi:<https://doi.org/10.3758/BF03196759>.
- Hunter, Tim. 2018. Left-corner parsing of minimalist grammars. In Robert C. Berwick & Edward Stabler (eds.), Minimalist parsing, Oxford, UK: Oxford University Press.
- Israel, Michael. 2004. The pragmatics of polarity. In Laurence R. Horn & Gregory Ward (eds.), The handbook of pragmatics, 701–723. Oxford/Malden, MA: Blackwell.
- Jacobson, Pauline. 2012. Direct compositionality and 'uninterpretability': The case of (sometimes) 'uninterpretable' features on pronouns. Journal of Semantics 29(3). 305–343. doi:<http://dx.doi.org/10.1093/jos/ffs005>.
- Jäger, Lena A., Felix Engelmann & Shravan Vasishth. 2015. Retrieval interference in reflexive processing: experimental evidence from Mandarin, and computational modeling. Frontiers in Psychology 6. 617. doi:<https://doi.org/10.3389/fpsyg.2015.00617>.
- Jäger, Lena A., Felix Engelmann & Shravan Vasishth. 2017. Similarity-based interference in sentence comprehension: Literature review and Bayesian meta-analysis. Journal of Memory and Language 94. 316–339. doi:<http://dx.doi.org/10.1016/j.jml.2017.01.004>.

- Jäger, Lena A., Daniela Mertzen, Julie A. Van Dyke & Shravan Vasishth. 2019. Interference patterns in subject-verb agreement and reflexives revisited: A large-sample study. Manuscript.
- Kayne, Richard. 1989. Facets of Romance past participle agreement. In Dialect variation and the theory of grammar, Foris Publishers.
- Kayne, Richard S. 1984. Connectedness and binary branching, vol. 16 Studies in Generative Grammar. Dordrecht: Foris Publications.
- Kayne, Richard S. 1994. The antisymmetry of syntax. Cambridge, MA: MIT Press.
- Kayne, Richard S. 2002. Pronouns and their antecedents. In Samuel David Epstein & Daniel Seely (eds.), Derivation and explanation in the minimalist program, 133–166. Blackwell Publishers Ltd. doi:10.1002/9780470755662.
- Ke, Alan Hezao. 2016. Full phase transfer. In Penn Linguistics Conference (PLC) 40, University of Pennsylvania, PA.
- Ke, Alan Hezao. in preparation. Search downward: a direct agree analysis of lubukusu complementizer agreement. Manuscript.
- Ke, Alan Hezao & Acrisio Pires. 2018. Covert reflexive argument in inalienable relational nouns. In Patrick Farrell (ed.), Proceedings of the Linguistic Society of America, vol. 3, 1–15. doi:<https://doi.org/10.3765/plsa.v3i1.4334>.
- Ke, Alan Hezao & Acrisio Pires. accepted. Local versus long-distance bound implicit arguments of inalienable relational nouns in Chinese. Journal of Linguistics .
- Ke, Alan Hezao, Ya Zhao, Liqun Gao, Shuying Liu & Acrisio Pires. 2019. On the implicit anaphoric argument of relational nouns in Mandarin Chinese. Journal of Psycholinguistic Research doi:<https://doi.org/10.1007/s10936-019-09633-2>.
- Keine, Stefan. to appear. Selective opacity. Linguistic Inquiry .
- Keine, Stefan & Bhamati Dash. 2017. Agreement reversals and the cyclicity of agree. In Proceedings of the 48th Annual Meeting of the North East Linguistic Society (NELS 48), University of Iceland.
- Keine, Stefan & Bhamati Dash. 2018. The ups and downs of agreement. Manuscript. http://www-bcf.usc.edu/~keine/papers/Up_down.pdf.
- King, Joseph, Caroline Andrews & Matthew Wagers. 2012. Do reflexives always find a grammatical antecedent for themselves? In the 25th Annual CUNY Conference on Human Sentence Processing, The CUNY Graduate Center New York, NY.
- Kitagawa, Yoshihisa. 1986. Subjects in English and Japanese: University of Massachusetts, Amherst Dissertation.

- Kitahara, Hisatsugu. 1995. Target α : Deducing strict cyclicity from derivational economy. *Linguistic Inquiry* 26(1). 47–77. <http://www.jstor.org/stable/4178888>.
- Koopman, Hilda. 1992. On the absence of case chains in Bambara. *Natural Language & Linguistic Theory* 10(4). 555–594. doi:<https://doi.org/10.1007/BF00133330>.
- Koopman, Hilda. 2006. Agreement configurations. In Cedric Boeckx (ed.), *Agreement systems*, vol. 92, book section 6, 159–199. Amsterdam: John Benjamins Publishing.
- Koopman, Hilda & Dominique Sportiche. 1991. The position of subjects. *Lingua* 85(2). 211–258. doi:[https://doi.org/10.1016/0024-3841\(91\)90022-W](https://doi.org/10.1016/0024-3841(91)90022-W). <http://www.sciencedirect.com/science/article/pii/002438419190022W>.
- Kratzer, Angelika. 1991. The representation of focus. *Semantics: An international handbook of contemporary research* 825–834.
- Kratzer, Angelika. 1998. More structural analogies between pronouns and tenses. In Devon Strolovitch & Aaron Lawson (eds.), *Proceedings of 8th Semantics and Linguistic Theory Conference (SALT 8)*, 92–110.
- Kratzer, Angelika. 2009. Making a pronoun: Fake indexicals as windows into the properties of pronouns. *Linguistic Inquiry* 40(2). 187–237. <https://muse.jhu.edu/article/263446>.
- Kreiner, Hamutal, Simon Garrod & Patrick Sturt. 2013. Number agreement in sentence comprehension: The relationship between grammatical and conceptual factors. *Language and Cognitive Processes* 28(6). 829–874.
- Kush, Dave. 2013. Respecting relations: Memory access and antecedent retrieval in incremental sentence processing: University of Maryland Dissertation.
- Kush, Dave, Jeffrey Lidz & Colin Phillips. 2015. Relation-sensitive retrieval: Evidence from bound variable pronouns. *Journal of Memory and Language* 82. 18–40. doi:<http://dx.doi.org/10.1016/j.jml.2015.02.003>. <http://www.sciencedirect.com/science/article/pii/S0749596X15000224>.
- Kuznetsova, Alexandra, Per B Brockhoff & Rune Haubo Bojesen Christensen. 2017. lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software* 82(13).
- Ladusaw, William Allen. 1980. *Polarity sensitivity as inherent scope relations*. New York: Garland Pub.
- Larson, Richard K. 1988. On the double object construction. *Linguistic Inquiry* 19(3). 335–391. doi:[10.2307/25164901](https://doi.org/10.2307/25164901). <http://www.jstor.org.proxy.lib.umich.edu/stable/25164901>.
- Lasnik, Howard & Mamoru Saito. 1999. *Minimalist analysis*. Oxford: Blackwell Publishing.

- Lee-Schoenfeld, Vera. 2004. Binding by phase: (non-)complementarity in German. *Journal of Germanic Linguistics* 16(2). 111–171. doi:10.1017/S1470542704000376.
- Lee-Schoenfeld, Vera. 2008. Binding, phases, and locality. *Syntax* 11(3). 281–298. doi:<https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9612.2008.00118.x>.
- Levin, Magnus. 2001. *Agreement with collective nouns in English*, vol. 103 Lund Studies in English. Stockholm, Sweden: Almqvist & Wiksell International.
- Levin, Magnus. 2006. Collective nouns and language change. *English Language and Linguistics* 10(2). 321–343. doi:10.1017/S1360674306001948. <https://www.cambridge.org/core/article/collective-nouns-and-language-change/306A3ABB167F0F5898FFC72380FAC841>.
- Lewandowsky, Stephan & Bennet Bronson Murdock. 1989. Memory for serial order. *Psychological Review* 96(1). 25–57.
- Lewis, Richard L. & Shravan Vasishth. 2005. An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science* 29(3). 375–419. doi:http://dx.doi.org/10.1207/s15516709cog0000_25.
- Lewis, Richard L., Shravan Vasishth & Julie A. Van Dyke. 2006. Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences* 10(10). 447–454. doi:<http://dx.doi.org/10.1016/j.tics.2006.08.007>. <http://www.sciencedirect.com/science/article/pii/S1364661306002142>.
- Link, Godehard. 1983. The logical analysis of plurals and mass terms: A latticetheoretical approach. In Reiner Bäuerle, Christoph Schwarze & Arnim von Stechow (eds.), *Meaning, use and interpretation of language*, 303–323. Berlin, Germany: de Gruyter.
- Logačev, Pavel & Shravan Vasishth. 2015. A multiple-channel model of task-dependent ambiguity resolution in sentence comprehension. *Cognitive Science* 40(2). 266–298. doi:<https://doi.org/10.1111/cogs.12228>.
- Lüdecke, Daniel, Alexander Bartel & Carsten Schwemmer. 2019. *sjplot: Data visualization for statistics in social science*. <https://strengjacke.github.io/sjPlot/>.
- Marian, Viorica, James Bartolotti, Sarah Chabal & Anthony Shook. 2012. CLEAR-POND: Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities. *PLOS ONE* 7(8). e43230. doi:<https://doi.org/10.1371/journal.pone.0043230>.
- Matushansky, Ora. 2008. On the linguistic complexity of proper names. *Linguistics and Philosophy* 31(5). 573–627. doi:<https://doi.org/10.1007/s10988-008-9050-1>.
- May, Robert. 1985. *Logical form: Its structure and derivation*. Cambridge, MA: MIT Press.

- Mayr, Clemens. 2012. Focusing bound pronouns. *Natural Language Semantics* 20(3). 299–348. doi:<https://doi.org/10.1007/s11050-012-9083-4>.
- McElree, Brian. 1998. Attended and non-attended states in working memory: Accessing categorized structures. *Journal of Memory and Language* 38(2). 225–252.
- McElree, Brian. 2000. Sentence comprehension is mediated by content-addressable memory structures. *Journal of Psycholinguistic Research* 29(2). 111–123. doi:<http://dx.doi.org/10.1023/A:1005184709695>.
- McElree, Brian, Stephani Foraker & Lisbeth Dyer. 2003. Memory structures that subserve sentence comprehension. *Journal of Memory and Language* 48(1). 67–91. doi:[http://dx.doi.org/10.1016/S0749-596X\(02\)00515-6](http://dx.doi.org/10.1016/S0749-596X(02)00515-6). <http://www.sciencedirect.com/science/article/pii/S0749596X02005156>.
- Merchant, Jason. 2001. *The syntax of silence: Sluicing, islands, and the theory of ellipsis*. Oxford University Press on Demand.
- Moulton, Keir & Chung-hye Han. 2018. C-command vs. scope: An experimental assessment of bound-variable pronouns. *Language* 94(1). 191–219.
- Narita, Hiroki. 2012. Remarks on the nature of headedness and compositionality in bare phrase structure. In *Proceedings of Sophia University Linguistic Society* 26, 81–126.
- Nicol, Janet & David Swinney. 1989. The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research* 18(1). 5–19. doi:<https://doi.org/10.1007/BF01069043>.
- Nunes, Jairo Morais. 1995. *The copy theory of movement and linearization of chains in the minimalist program*: University of Maryland at College Park Dissertation.
- Parker, Dan, Sol Lago & Colin Phillips. 2015. Interference in the processing of adjunct control. *Frontiers in Psychology* 6. 1346. doi:10.3389/fpsyg.2015.01346. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4561755/>.
- Parker, Dan & Colin Phillips. 2017. Reflexive attraction in comprehension is selective. *Journal of Memory and Language* 94. 272–290. doi:<http://dx.doi.org/10.1016/j.jml.2017.01.002>. <http://www.sciencedirect.com/science/article/pii/S0749596X17300025>.
- Parker, Dan, Michael Shvartsman & Julie A Van Dyke. 2017. The cue-based retrieval theory of sentence comprehension: New findings and new challenges. In Linda Escobar, Vicenç Torrens & Teresa Parodi (eds.), *Language processing and disorders*, 121–144. Newcastle: Cambridge Scholars Publishing.
- Parsons, Terence. 1978. Pronouns as paraphrases. Unpublished.

- Partee, Barbara H. 2008. Lecture 9: Pronouns and reflexives: Syntax and semantics. Lecture notes. http://people.umass.edu/partee/RGGU_2008/RGGU089_2up.pdf.
- Patil, Umesh, Sandra Hanne, Frank Burchert, Ria De Bleser & Shravan Vasishth. 2016a. A computational evaluation of sentence processing deficits in aphasia. *Cognitive Science* 40(1). 5–50. doi:10.1111/cogs.12250.
- Patil, Umesh, Shravan Vasishth & Richard L. Lewis. 2016b. Retrieval interference in syntactic processing: The case of reflexive binding in English. *Frontiers in Psychology* 7(329). doi:10.3389/fpsyg.2016.00329. <http://journal.frontiersin.org/article/10.3389/fpsyg.2016.00329>.
- Pearlmutter, Neal J., Susan M. Garnsey & Kathryn Bock. 1999. Agreement processes in sentence comprehension. *Journal of Memory and Language* 41(3). 427–456. doi:<http://dx.doi.org/10.1006/jmla.1999.2653>. <http://www.sciencedirect.com/science/article/pii/S0749596X99926530>.
- Pesetsky, David Michael. 1995. *Zero syntax: Experiencers and cascades*. Cambridge, MA: MIT Press.
- Phillips, Colin, Matthew W. Wagers & Ellen F. Lau. 2011. Grammatical illusions and selective fallibility in real-time language comprehension. In *Experiments at the interfaces*, vol. 37 Syntax and Semantics, 147–180. Bingley, UK: Emerald Group Publishing Limited. doi:10.1108/S0092-4563(2011)0000037009.
- Pica, Pierre. 1987. On the nature of the reflexivization cycle. In Joyce McDonough & Bernadette Plunkett (eds.), *Proceedings of the 17th Annual Meeting of the North East Linguistic Society (NELS 17)*, 483–499.
- Pinker, Steven. 1991. Rules of language. *Science* 253(5019). 530–535.
- Polinsky, Maria & Eric Potsdam. 2001. Long-distance agreement and topic in Tsez. *Natural Language & Linguistic Theory* 19(3). 583–646. doi:<http://dx.doi.org/10.1023/A:1010757806504>.
- Preminger, Omer. 2009. Breaking agreements: Distinguishing agreement and clitic doubling by their failures. *Linguistic Inquiry* 40(4). 619–666. doi:<https://doi.org/10.1162/ling.2009.40.4.619>.
- Preminger, Omer. 2011. Asymmetries between person and number in syntax: a commentary on Baker’s SCOPA. *Natural Language & Linguistic Theory* 29(4). 917–937.
- Preminger, Omer. 2013. That’s not how you agree: A reply to Zeijlstra. *The Linguistic Review* 30(3). 491–500. doi:10.1515/tlr-2013-0015. <https://www.degruyter.com/view/j/tlir.2013.30.issue-3/tlr-2013-0015/tlr-2013-0015.xml>.

- Preminger, Omer. 2014. Agreement and its failures. Cambridge: The MIT Press.
<https://muse.jhu.edu/book/35120>.
- Preminger, Omer. 2018. The anaphor agreement effect: Further evidence against anaphora-as-agreement. Manuscript.
- Preminger, Omer & Maria Polinsky. 2015. Agreement and semantic concord: A spurious unification. Manuscript. *Lingbuzz*/002363.
- Quicoli, Carlos A. 2008. Anaphora by phase. *Syntax* 11(3). 299–329. doi: doi:10.1111/j.1467-9612.2008.00116.x. <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9612.2008.00116.x>.
- R Core Team. 2014. R: A language and environment for statistical computing. <http://www.R-project.org/>.
- Reinhart, Tanya & Eric Reuland. 1991. Anaphors and logophors: an argument structure perspective. In Jan Koster & Eric Reuland (eds.), Long distance anaphora, 283–321. Cambridge: Cambridge University Press.
- Reinhart, Tanya & Eric Reuland. 1993. Reflexivity. *Linguistic Inquiry* 24(4). 657–720. doi:10.2307/4178836. <http://www.jstor.org/stable/4178836>.
- Reitter, David, Frank Keller & Johanna D Moore. 2011. A computational cognitive model of syntactic priming. *Cognitive Science* 35(4). 587–637.
- Resnik, Philip. 1992. Left-corner parsing and psychological plausibility. In Proceedings of the 14th Conference on Computational Linguistics, 191–197. Nantes, France: Association for Computational Linguistics. doi:10.3115/992066.992098.
- Reuland, Eric. 2001. Primitives of binding. *Linguistic Inquiry* 32(3). 439–492. doi: <https://doi.org/10.1162/002438901750372522>.
- Reuland, Eric. 2006. Agreeing to bind. In Hans Broekhuis, Norbert Corver, Ursula Kleinhenz, Jan Koster & Riny Huijbregts (eds.), Organizing grammar : Linguistic studies in honor of Henk van Riemsdijk, 505–513. Berlin/Boston: De Gruyter.
- Reuland, Eric. 2011. Anaphora and language design Linguistic Inquiry Monographs. Cambridge: The MIT Press.
- Rezac, Milan. 2004. Elements of cyclic syntax: Agree and merge: University of Toronto Dissertation.
- Rizzi, Luigi. 1982. Issues in italian syntax. Dordrecht: Foris Publications.
- Rizzi, Luigi. 1990. On the anaphor-agreement effect. *Rivista di Linguistica* 2(1). 27–42.

- Rizzi, Luigi. 1997. The fine structure of the left periphery. In Liliane Haegeman (ed.), Elements of grammar: A handbook of generative syntax, 281–337. Dordrecht: Kluwer.
- Rizzi, Luigi. 2001. On the position “int(errogative)” in the left periphery of the clause. In Guglielmo Cinque & Giampaolo Salvi (eds.), Current studies in Italian syntax: Essays offered to Lorenzo Renzi, 287–296. Amsterdam: Elsevier.
- Rizzi, Luigi. 2016. Labeling, maximality and the head–phrase distinction. The Linguistic Review 33(1). 103–127.
- Rooth, Mats. 1985. Association with focus: University of Massachusetts at Amherst Dissertation.
- Rooth, Mats. 1992a. Ellipsis redundancy and reduction redundancy. In Steve Berman & Arild Hestvik (eds.), Proceedings of the Stuttgart ellipsis workshop, vol. 29 Arbeitspapiere des Sonderforschungsbereichs 340, Bericht Nr. 29, 1–26. Heidelberg: IBM Germany.
- Rooth, Mats. 1992b. A theory of focus interpretation. Natural Language Semantics 1(1). 75–116. doi:<http://dx.doi.org/10.1007/BF02342617>.
- De Rosario-Martinez, Helios, John Fox & R Core Team. 2015. Post-hoc interaction analysis: package ‘phia’ 0.2-0 edn.
- Ross, John Robert. 1967. Constraints on variables in syntax: MIT Dissertation.
- Rullmann, Hotze. 2004. First and second person pronouns as bound variables. Linguistic Inquiry 35(1). 159–168. <http://www.jstor.org/stable/4179267>.
- Safir, Ken. 2004. The syntax of anaphora. New York: Oxford University Press.
- Sag, Ivan A. 1976. Deletion and logical form: Massachusetts Institute of Technology Dissertation.
- Sauerland, Uli. 2000. The content of pronouns: Evidence from focus. In Proceedings of 10th Semantics and Linguistic Theory Conference (SALT 10), 167–184.
- Sauerland, Uli. 2003. A new semantics for number. In Robert B. Young & Yuping Zhou (eds.), Proceedings of 13th Semantics and Linguistic Theory Conference (SALT 13), 258–275.
- Sauerland, Uli. 2008. On the semantic markedness of ϕ -features. In Daniel Harbour, David Adger & Susana Béjar (eds.), Phi theory: Phi-features across modules and interfaces, 57–82. Oxford: Oxford University Press.
- Sauerland, Uli. 2013. Presuppositions and the alternative tier. In Proceedings of the 23rd Semantics and Linguistic Theory Conference (SALT 23), 156–173.

- Schlenker, Philippe. 2003. A plea for monsters. Linguistics and Philosophy 26(1). 29–120. doi:<https://doi.org/10.1023/A:1022225203544>.
- Schlenker, Philippe. 2005a. Minimize restrictors! (Notes on definite descriptions, Condition C and epithets). In Proceedings of Sinn und Bedeutung (SuB) 9, 385–416.
- Schlenker, Philippe. 2005b. Non-redundancy: Towards a semantic reinterpretation of binding theory. Natural Language Semantics 13(1). 1–92. doi:<https://doi.org/10.1007/s11050-004-2440-1>.
- Schütze, Carson T. 2001. On the nature of default case. Syntax 4(3). 205–238. doi:<https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-9612.00044>.
- Sells, Peter. 1987. Aspects of logophoricity. Linguistic Inquiry 18(3). 445–479. <http://www.jstor.org/stable/4178550>.
- Shen, Zheng & Peter W. Smith. 2019. Morphological and semantic agreement beyond hybrid nouns. In Proceedings of the 49th Annual Meeting of the North East Linguistic Society (NELS 49), .
- Shim, Jae-Young. 2018. $\langle \phi, \phi \rangle$ -less labeling. Language Research 54(1). 23–39. doi:10.30961/lr.2018.54.1.23. <http://hdl.handle.net/10371/139743>.
- Sigurðsson, Halldór Ármann. 1991. Icelandic case-marked pro and the licensing of lexical arguments. Natural Language & Linguistic Theory 9(2). 327–363. doi:<https://doi.org/10.1007/BF00134679>.
- Sigurðsson, Halldór Ármann. 1996. Icelandic finite verb agreement. In Working Papers in Scandinavian Syntax, vol. 57, 1–46.
- Slabakova, Roumyana. 2006. Is there a critical period for semantics? Second Language Research 22(3). 302–338. doi:<https://journals.sagepub.com/doi/abs/10.1191/0267658306sr270oa>.
- Slabakova, Roumyana. 2008. Meaning in the second language, vol. 34. Berlin: Walter de Gruyter.
- Slabakova, Roumyana. 2009. Features or parameters: which one makes second language acquisition easier, and more interesting to study? Second Language Research 25(2). 313–324. doi:<https://journals.sagepub.com/doi/abs/10.1177/0267658308100291>.
- Sloggett, Shayne & Brian Dillon. 2017. Animacy in reflexive processing. In the 30th Annual CUNY Conference on Human Sentence Processing, MIT, Cambridge, MA.
- Smith, Peter W. 2017. The syntax of semantic agreement in English. Journal of Linguistics 53(4). 823–863. doi:10.1017/S0022226716000360.

- Spathas, Giorgos. 2010. Focus on anaphora: Universiteit Utrecht Dissertation.
- Sportiche, Dominique. 1988. A theory of floating quantifiers and its corollaries for constituent structure. Linguistic Inquiry 19(3). 425–449.
- Stabler, Edward. 2013. Two models of minimalist, incremental syntactic analysis. Topics in Cognitive Science 5(3). 611–633. doi:<https://onlinelibrary.wiley.com/doi/abs/10.1111/tops.12031>.
- Stechow, Arnim von. 2003. Feature deletion under semantic binding: Tense, person, and mood under verbal quantifiers. In Proceedings of the 33rd Annual Meeting of the North East Linguistic Society (NELS 33), vol. 33, 379–404. Graduate Linguistic Student Association (GLSA) Publications.
- Sturt, Patrick. 2003. The time-course of the application of binding constraints in reference resolution. Journal of Memory and Language 48(3). 542–562. doi:[http://dx.doi.org/10.1016/S0749-596X\(02\)00536-3](http://dx.doi.org/10.1016/S0749-596X(02)00536-3).
- Sudo, Yasutada. 2012. On the semantics of phi features on pronouns: Massachusetts Institute of Technology Dissertation.
- Sudo, Yasutada. 2014. Dependent plural pronouns with Skolemized choice functions. Natural Language Semantics 22(3). 265–297. doi:<https://doi.org/10.1007/s11050-014-9107-3>.
- Svenonius, Peter. 2006. The emergence of axial parts. In Peter Svenonius & Marina Pantcheva (eds.), Nordlyd: Tromsø Working Papers in Linguistics (special issue on adpositions), vol. 33.1, 49–77. Tromsø: CASTL. <http://www.ub.uit.no/munin/nordlyd/>.
- Takahashi, Shoichi & Danny Fox. 2005. Maxelide and the re-binding problem. In Proceedings of the 15th Semantics and Linguistic Theory Conference (SALT 15), 223–240.
- Takita, Kensuke. 2009. The proper binding condition effect as a consequence of cyclic linearization. In Anisa Schardl, Martin Walkow & Muhammad Abdurrahman (eds.), Proceedings of the 38th Annual Meeting of the North East Linguistic Society (NELS 38), 427–440. Graduate Linguistic Student Association (GLSA) Publications.
- Takita, Kensuke. to appear. Labeling for linearization. The Linguistic Review .
- Tang, Chih-Chen Jane. 1989. Chinese reflexives. Natural Language & Linguistic Theory 7(1). 93–121. <http://www.jstor.org/stable/4047781>.
- Vasishth, Shravan, Sven Brüssow, Richard L. Lewis & Heiner Drenhaus. 2008. Processing polarity: How the ungrammatical intrudes on the grammatical. Cognitive Science 32(4). 685–712. doi:<http://dx.doi.org/10.1080/03640210802066865>.

- Wagers, Matthew W., Ellen F. Lau & Colin Phillips. 2009. Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language* 61(2). 206–237. doi:<http://dx.doi.org/10.1016/j.jml.2009.04.002>.
- Wiklund, Anna-Lena. 2001. Dressing up for vocabulary insertion: The parasitic supine. *Natural Language & Linguistic Theory* 19(1). 199–228. <http://www.jstor.org.proxy.lib.umich.edu/stable/4047916>.
- Williams, Edwin. 1977. Discourse and logical form. *Linguistic Inquiry* 8(1). 101–139. doi:<http://www.jstor.org/stable/4177974>.
- Wong, May Lai-Yin. 2013. Concord patterns with collective nouns in Hong Kong English. With illustrative material from the International Corpus of English (Hong Kong component). *Linguistik online* 37(1). doi:<http://dx.doi.org/10.13092/lo.37.514>.
- Woolford, Ellen. 1999. More on the anaphor agreement effect. *Linguistic Inquiry* 30(2). 257–287. <http://www.jstor.org/stable/4179061>.
- Wurmbrand, Susi. 2012a. Parasitic participles in Germanic: Evidence for the theory of verb clusters. *Taal en Tongval* 64(1). 129–156.
- Wurmbrand, Susi. 2012b. The syntax of valuation in auxiliary-participle constructions. In E. Jaehoon Choi, Alan Hogue, Jeffrey Punske, Deniz Tat, Jessamyn Schertz & Alex Trueman (eds.), *Coyote working papers: Proceedings of the 29th West Coast Conference on Formal Linguistics (WCCFL 29)*, 154–162. Cascadia Proceedings Project.
- Wurmbrand, Susi. 2017. Feature sharing or how I value my son. In Claire Halpert, Hadas Kotek & Coppe van Urk (eds.), *The Pesky Set: Papers for David Pesetsky*, 173–182. Cambridge, MA: MIT Working Papers in Linguistics (MITWPL).
- Xiang, Ming, Brian Dillon & Colin Phillips. 2009. Illusory licensing effects across dependency types: ERP evidence. *Brain and Language* 108(1). 40–55. doi:<http://dx.doi.org/10.1016/j.bandl.2008.10.002>.
- Xue, Ping, Carl Pollard & Ivan A. Sag. 1994. A new perspective on Chinese *ziji*. In Raul Aranovich, William Byrne, Susanne Preuss & Martha Senturia (eds.), *Proceedings of the 13th West Coast Conference on Formal Linguistics (WCCFL)*, 432–447. Stanford, CA: CSLI Publication.
- Zeijlstra, Hedde. 2004. *Sentential negation and negative concord*: University of Amsterdam Dissertation.
- Zeijlstra, Hedde. 2012. There is only one way to agree. *The Linguistic Review* 29(3). 491–539. doi:10.1515/tlr-2012-0017. <https://www.degruyter.com/view/j/tlr.2012.29.issue-3/tlr-2012-0017/tlr-2012-0017.xml>.

- Zwart, Jan-Wouter. 2002. Issues relating to a derivational theory of binding. In Samuel David Epstein & T. Daniel Seely (eds.), Derivation and explanation in the minimalist program, book section 11, 269–304. Oxford: Blackwell Publishers Ltd.
- Zyman, Erik. 2018. Quantifier float as stranding. Natural Language & Linguistic Theory 36(3). 991–1034. doi:<https://doi.org/10.1007/s11049-017-9393-9>.