A CROSS-LINGUISTIC ACCOUNT OF REFLEXIVITY USING SYNCHRONOUS TREE ADJOINING GRAMMAR

by

Dennis Ryan Storoshenko

BA, Simon Fraser University, April 2002

MA, Simon Fraser University, April 2004

THESIS SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN THE DEPARTMENT

OF

LINGUISTICS

SIMON FRASER UNIVERSITY

Summer 2010

APPROVAL

Name: Dennis Ryan Storoshenko

Degree: Doctor of Philosophy

Title of Thesis: A Cross-Linguistic Account of Reflexivity Using Synchronous Tree Adjoining Grammar

Examining Committee: Dr. Panayiotis Pappas, Associate Professor of Linguistics (Chair)

Dr. Chung-hye Han, Senior Supervisor
Associate Professor of Linguistics

Dr. Nancy Hedberg, Supervisor
Associate Professor of Linguistics

Dr. Rose-Marie Déchaine, Supervisor
Associate Professor of Linguistics
University of British Columbia

Dr. Réjean Canac-Marquis, Internal Examiner
Associate Professor of French

Dr. Martha McGinnis, External Examiner
Associate Professor of Linguistics,
University of Victoria

Date Approved: July 7, 2010

STATEMENT OF ETHICS APPROVAL

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

(a) Human research ethics approval from the Simon Fraser University Office of Research Ethics,

or

(b) Advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University;

or has conducted the research

(c) as a co-investigator, collaborator or research assistant in a research project approved in advance,

or

(d) as a member of a course approved in advance for minimal risk human research, by the Office of Research Ethics.

A copy of the approval letter has been filed at the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Simon Fraser University Library
Simon Fraser University
Burnaby, BC, Canada

Last update: Spring 2010

Abstract

This thesis examines the different ways in which reflexive relations are expressed across languages. Using the reflexive typology of Déchaine and Wiltschko (2002) as a starting point, case-studies of four languages are presented: English, Korean, Shona, and Plains Cree. These linguistic case-studies rely on a broad spectrum of data collection methods, including corpus research, psycholinguistic experimentation, and field-elicitation. Through these diverse methodologies, a sound empirical basis for the conclusions of the thesis is constructed, showing how the different methods can be combined to complement each other in the formulation of linguistic theory. Then, the data are treated in terms of Synchronous Tree Adjoining Grammar (STAG), showing that the STAG formalism is not only robust enough to handle a diverse selection of languages, but also showing how the various forms of reflexive expression can be formalised. Specifically, with English, reflexives are treated as functions which take predicates as arguments, establishing an explicit co-reference between the arguments of the predicate. In STAG, it emerges that the familiar c-command relationship between a reflexive pronoun and its antecedent is a consequence of the analysis, rather than a stipulated constraint. Korean and Shona reflexivity is expressed in terms of Bound Variable Anaphora; the STAG implementation for these languages shows how parametric variation between languages in terms of binding restrictions (local, anti- local, or unconstrained) can be expressed in terms of a constraint on the derivation of sentences containing bound variables. Plains Cree, while having the simplest syntax in that its reflexives are intransitives, emerges to be the biggest challenge for STAG, exposing the need for further work in defining the STAG formalism at a finer syntactic and semantic level than present implementations allow.

Acknowledgments

Firstly, thanks to Dr. Chung-hye Han for giving me my first copy of Reinhart and Reuland back during my MA and telling me this was something to look into. In the years since, she has been a continual source of sound scholarly guidance and support without which this work would never have been possible. The combination of diverse data collection methods with theoretical analysis in this thesis is a result of her tutelage. Dr. Nancy Hedberg, I believe, was the first person to suggest Grad School, so she is just as responsible for this work. Her unflinching encouragement of both teaching and studies, and her help in navigating officialdom are much appreciated. Thanks also to Dr. Rose-Marie Déchaine for my introduction to Bantu, and for pushing me to keep digging for the generalisations underlying the Shona data. The discovery and deduction of good old fashioned elicitation have made for some of the best "aha!" moments of the past five years.

My examiners, Dr. Réjean Canac-Marquis of SFU and Dr. Martha McGinnis of UVic have both contributed to the improvement of this thesis. Their questions and comments have led to a sharpening of the discussion presented here, and will definitely inform my work as I move forward with this research. Thanks also to Dr. Panayiotis Pappas, chair of my defense, for making the whole thing painless.

While I have stayed an SFU student, this thesis is partially a product of UBC, and the experimental work presented here has its roots in the 2007 LSA Summer Institute at Stanford. This work has benefitted from individuals at all three. In addition to those listed above, thanks go to Michael Barrie, Cliff Burgess, Henry Davis, Donna Gerdts, Lisa Matthewson, Dean Mellow, Murray Munro, Michael Rochemont, Hotze Rullmann, Jeff Runner, Maite Taboada, Eric Vatikiotis-Bateson, Martina Wiltschko, and Yue Wang. You have asked the right questions and given the right advice, helping me to keep on top of courses, research, conferences, teaching, and life in general.

Through the years of this work, there have been many fellow students come and gone at SFU who have been a source of encouragement, constructive criticism, and dinner parties. Sam Al-Khatib, Lorna Fadden, Ashleigh Gonzales, Emrah Görgülü, Morgan Mameni, Emma Mileva, David Potter, Anne Rimrott, Yasuko Sakurai, and Loreley Wiesemann: thank you. My adopted cohort at UBC have also been a huge source of support over the past few years. Thanks to my Shona research partner Heather Bliss, along with Elizabeth Ferch, Raphael Girard, Beth Rogers, Murray Schellenberg, Anita Szakay and Carmela Toews; you all helped make Totem Fields feel like a second home. A large part of the final analysis and first drafting of this thesis took place during a visiting term at the University of Edinburgh, so here's a shout-out also to all my DSB 1.15 officemates and a thanks for Monday afternoon coffee and biscuits overlooking Arthur's Seat.

Sincere thanks are owed to Calisto Mudzingwa and Stanley Murevesi, my exceptionally patient Shona language consultants, for sharing their language with me: *mazviita*. Lan Kim, Robert Ho-Sup Kim, and Seunghun Lee are also thanked for their judgements on Korean. The visual stimuli the Korean eyetracking study were produced Beverly Wu-Hannah, and the PDF-OCR-X software used in the Shona and Plains Cree text analysis was produced by her husband Steve (both longtime friends of mine), in response to my complaining about how much work the corpus analysis was turning out to be.

Outside of Linguistics, there has been a wealth of support and encouragement that kept me going through this process. My family have always been an unquestioning source of support, never failing to help when things have gotten tough. JP, Alex, and everyone else Clubhouse, thanks so much for your relentless back-patting and even more relentless sarcasm, you have really made a difference in these past few years. Mike, thanks for twenty years of unremitting amity, which we can definitely still afford. Finally, Dave, the best roommate this PhD student could ever have asked for, thanks for everything.

Lastly, this thesis would never have been written if not for some wonderful purveyors of coffee. Parminder at Renaissance Coffee, your cheerful greetings and uncanny ability to figure out my teaching schedule based on what I order have been a fixture of my time at SFU. Melriches in Vancouver and Black Medicine Coffee in Edinburgh practically became second offices where a lot of the work represented in this was done. Thank you all for supporting my habit, and tolerating my spread of papers and laptop without ever spilling on me.

Contents

Approval **ii**

Abstract **iii**

Acknowledgments **iv**

Contents **vi**

List of Tables **x**

List of Figures **xi**

List of Abbreviations **xii**

1 Setting the Stage **1**

1.1 Defining Reflexivity 1

1.2 Two Theories on the Distribution of Reflexives 2

1.3 A Typology of Reflexives 8

1.4 Modelling Reflexivity 19

1.5 Roadmap for This Thesis 22

2 One Face, Many Functions **24**

2.1 Background on English *self* Pronouns 25

2.1.1 Reflexive Pre-history: Before 1981 26

2.1.2 Picture It: Representational DPs 27

2.1.3 Long Distance? Not Exactly 34

2.2 Corpus Study . . . 37
2.2.1 Selecting and Analysing the Corpus . . . 38
2.2.2 Findings . . . 43
2.3 Emphatics and Manner Adjuncts . . . 44
2.3.1 Disambiguating Sentence Final *self* Pronouns . . . 46
2.3.2 External Argument Adjuncts . . . 53
2.3.3 Emphatic *Self* Pronouns . . . 61
2.4 Semantics of English *self* Pronouns . . . 68
2.4.1 Relating Arguments . . . 69
2.4.2 Emphatics and the EAA . . . 75
2.5 Loose Ends . . . 84

3 The Accidental Reflexive 88
3.1 Korean Reflexives based on *casin* . . . 88
3.1.1 *pronoun-casin* . . . 89
3.1.2 Monomorphemic *casin* . . . 93
3.1.3 *caki-casin* . . . 98
3.2 *Caki* . . . 102
3.2.1 The Pronominal Argument . . . 102
3.2.2 *Caki* as Anaphor . . . 105
3.2.3 Refuting the Pronoun Arguments . . . 116
3.2.4 Bound Variable *caki* . . . 122
3.3 *Caki* in Corpus . . . 131
3.3.1 Non C-Commanding Antecedents . . . 134
3.3.2 Antecedentless *Caki* . . . 136
3.4 Testing the Context-Sensitivity of *Caki* . . . 139
3.4.1 Methods and Materials . . . 140
3.4.2 Results . . . 144
3.4.3 Conclusions . . . 148
3.5 Summarising Korean . . . 149

4 Keeping it Local 153
4.1 Introducing Shona . . . 154

4.2 Shona Reflexives 162
4.2.1 The Literature on Shona and Beyond 163
4.2.2 Object Marker versus Detransitivisation 166
4.2.3 *Zvi* as a unique member of the set of Object Markers 170
4.3 Corpus Study 180
4.3.1 Building the Corpus 182
4.3.2 Corpus Results 182
4.4 Concluding Shona 189
4.4.1 Comparison with Romance Reflexives 189
4.4.2 The Wider Bantu Picture 192

5 Losing an Argument **194**
5.1 Plains Cree Reflexives 195
5.1.1 Co-argument Reflexivity 196
5.1.2 Dynamic Unaccusatives 199
5.1.3 Inanimate Reflexives 201
5.2 Emphasis in Plains Cree 202
5.2.1 *Wiya* as an Emphatic 202
5.2.2 Proximates as Emphatic? 204
5.3 Looking at Texts 205
5.3.1 Corpus Assembly 206
5.3.2 Corpus Findings 206
5.4 Summarising Plains Cree 208

6 Formalising the Forest **210**
6.1 STAG Basics 211
6.2 English *Self* Pronouns 217
6.2.1 Coargument Reflexives 218
6.2.2 ECM and Raising 223
6.2.3 Emphatic Uses 228
6.3 Parametrizing Bound Variables 230
6.4 Representational Challenges 242
6.4.1 Bantu Verbal Morphology meets STAG 242

6.4.2 Plains Cree Reflexives 244
6.5 Final Summary 246

7 Out of the Woods **248**
7.1 Looking Back 248
7.2 Looking Ahead 250

Bibliography **253**

List of Tables

1.1 Reinhart and Reuland (1993) Typology of Anaphors 10
1.2 Déchaine and Wiltschko 2002b Typology of Reflexives 18

2.1 English *Self* Pronouns 25
2.2 Distribution of *Self* Pronouns across WSJ and SWB Corpora 43
2.3 Emphatic *Self* Pronoun Distinctions 68
2.4 Semantic Forms and Restrictions on Emphatics 80

3.1 Distribution of *Caki*: First Summary 104
3.2 Distribution of *Caki*: Second Summary 116
3.3 Distribution of *Caki*: Third Summary 122
3.4 Distribution of *Caki*: Fourth Summary 130
3.5 Antecedents of accusative-marked forms in KOREA-1 Corpus (B.-M. Kang 2001) 131
3.6 Person Features of *caki* in Corpus 133
3.7 Antecedents of 3^{rd} Person *caki* in Corpus 133
3.8 *caki* with Non-C-Commanding Antecedents 134
3.9 *caki* with No Antecedent 136
3.10 Distribution of *Caki*: Final Summary 151

6.1 Parameter Settings for Bound Variables 241

List of Figures

2.1 Screen shots from the magnitude estimation task instructions 49
2.2 Participant view of magnitude estimation task experiment interface 50
2.3 Mean Ratio Scores for each Condition in Magnitude Estimation Task Experiment 51

3.1 Korean Eyetracking Experiment Sample Display 141
3.2 Proportion of Fixations over Time to Proper Name Referents 145
3.3 Proportions of Fixations in Image after *caki* for each Bias 146
3.4 Proportions of Fixations in all Conditions 147
3.5 Average Means for each Anahpor/Bias Combination 147

List of Abbreviations

1SG	First Person Singular
1PL	First Person Plural
2SG	Second Person Singular
2PL	Second Person Plural
3F	Third Person Female
3M	Third Person Male
3SG	Third Person Singular
3PL	Third Person Plural
ACC	Accusative Case
ADNOM	Adnominal Marker
ASP	Aspect
CJCT	Conjunct Marker
CL#	Noun Class Marker
COMP	Complementizer
CONJ	Conjunction
DAT	Dative
DECL	Declarative
DEF	Definite
DIR	Direct Word Order
EMPH	Emphatic
GEN	Genitive Case
HON	Honourific
INAN	Inanimate

I.TH	Inanimate Theme Marker
INT	Interrogative
INTR	Intransitive Marker
LCAL	Local Referent
NOM	Nominative Case
PART	Partitive
PL	Plural
POSS	Possessive
POSSD	Possessed Item
POSSR	Possessor
PRES	Present Tense
PRN	Pronoun
PROG	Progressive
PST	Past Tense
REFL	Reflexive
RELAT	Relativiser
RPT	Reportative
SE	Simplex Expression Anaphor
STAT	Stative
TNS	Tense
TOP	Topic
TRANS	Transitive Marker

Chapter 1

Setting the Stage
Introducing Reflexivity and Synchronous Tree Adjoining Grammar

A beginning is the time for taking the most delicate care that the balances are correct.
-Princess Irulan. *Dune* by Frank Herbert.

In this opening Chapter, I begin with a short definition of reflexivity, moving on into a brief overview of the Chomsky (1981) and Reinhart and Reuland (1993) principles governing the use of reflexives. Then, I present the Déchaine and Wiltschko (2002b) typology of reflexive expressions, outlining how this typology informs the selection of language case-studies through the rest of the thesis.

1.1 Defining Reflexivity

Reflexivity is the phenomenon which exists when two semantic argument positions within a predicate are occupied by the same entity. Expressed in lambda calculus, using a transitive predicate as an example, it is a situation where both argument positions of a two-argument predicate are filled by the same entity:

(1) $[\![\text{hit}]\!] = \lambda x \lambda y.\ \text{hit}(y,x) \rightarrow \text{hit}(a,a)$

The goal of this thesis then is to provide a model of what takes place at the arrow in (1); how does a transitive predicate with two distinct variables come to be in a form where both of those variables stand for the same individual?

With a ditransitive predicate, the issue is more complex, as there are four possible combinations:

(2) a. $[\![\text{introduce}]\!] = \lambda x \lambda y \lambda z.z$ introduce y to $x \rightarrow a$ introduce a to b

b. $[\![\text{introduce}]\!] = \lambda x \lambda y \lambda z.z$ introduce y to $x \rightarrow a$ introduce b to a

c. $[\![\text{introduce}]\!] = \lambda x \lambda y \lambda z.z$ introduce y to $x \rightarrow a$ introduce b to b

d. $[\![\text{introduce}]\!] = \lambda x \lambda y \lambda z.z$ introduce y to $x \rightarrow a$ introduce a to a

For the purposes of this thesis, I will consider all the possible permutations in (2) to be equally reflexive. Thus, I will not be limiting the discussion to cases where the subject of the sentence is one of the two arguments involved in the reflexive relationship.

1.2 Two Theories on the Distribution of Reflexives

However simple the semantic form of reflexivity may be, the syntactic means of expressing this meaning are diverse and much more complex. In English, this is most commonly done using a *self* pronoun:

(3) Jim_i hit himself_i.

While there is, in a sense, an underlyingly intransitive semantics here, there being only one entity involved in the action, the syntax remains transitive. The direct object of the clause is the *self* pronoun *himself*; syntactically, this is an argument, but it does not represent a new entity. Rather, *himself* is dependent upon its antecedent, *Jim*, to determine its reference. This is where the notion of binding comes into play when discussing reflexivity; the *self* pronoun is c-commanded by a co-indexed antecedent, and thus bound. The definitions of c-command and binding which will govern the rest of this thesis are presented in (4) and (5):

(4) DEFINITION OF C-COMMAND (Chomsky 1981):
Node α c-commands node β iff the branching node most immediately dominating α also dominates β and α does not contain β.

(5) DEFINITION OF BINDING (Chomsky 1981):
a is bound iff it is c-commanded by a co-indexed antecedent.

This interrelation between reflexivity and binding has led to reflexivity generally being considered a part of Binding Theory, and strongly connected to the Binding Conditions of Chomsky (1981):[1]

- Condition A: An anaphor must be bound in its governing category.
- Condition B: A pronoun must be free in its governing category.
- Condition C: An R-Expression must be free.

Of most relevance to the *self* pronouns is Condition A, as this covers reflexive and reciprocal expressions in English. As a result, Condition A is strongly associated with the concept of reflexivity, remaining a widely-used diagnostic, even though the word 'reflexive' does not appear in the rule itself. Instead, the term 'anaphor' is used, and somewhat loosely. I prefer to proceed with the broader definition advanced by Hirst (1981):

(6) DEFINITION OF ANAPHORA:
The device of making in discourse an **abbreviated** reference to an entity (or entities) in the expectation that the perceiver of the discourse will be able to disabbreviate the reference and thereby determine the identity of the entity.

From this definition, there is nothing in the term 'anaphor' which connects it explicitly to *self* pronouns, or even nominal categories for that matter. I will thus refer to cases where there is a relationship between arguments of a single predicate as *reflexivity*, and the associated lexical items as *reflexives* or, using English as a basis, *self* pronouns. The broader term *anaphor* will be applied to those items which can be bound by arguments of a

[1]Technically speaking, in the original 1981 rendition, Chomsky uses 'Principle' rather than 'Condition', but by the publication of Chomsky (1986), 'Condition' has been adopted, so I choose to stay in line with this more recent formulation.

different predicate, and are not necessarily reflexive. Crucially, what I will be describing as anaphors throughout this thesis should not be understood to have any implicit connection to Condition A of Chomsky's theory. Likewise, reflexivity is not connected by definition to Condition A.

Chomsky's use of *anaphor* is motivated by the fact that Condition A is not designed to account only for *self* pronouns, but also for reciprocals:

(7) [Stephen and Phil]$_i$ mocked *each other*$_i$.

Here, *each other* is argued to be bound by the conjoined subject *Stephen and Phil*, and the reciprocal meaning is obtained. Reflexives and reciprocals are treated together in Condition A, as they can both be captured by this local binding relationship in English. This thesis will have little to say on the subject of reciprocals, except to point out those cases where the conflation of reflexives and reciprocals is not supported in a given language, thus motivating their separation.

Getting back to reflexivity, what is important to recognise is that Chomsky's Condition A is a statement on the felicitous use of *self* pronouns; it says nothing specifically about reflexivity being necessarily tied to a binding relationship. However, discussions of reflexivity are often concentrated upon finding a way to fit the data into terms of a *self* pronoun being c-commanded by an antecedent within a given local domain.

A major exception to this generalisation is the work of Reinhart and Reuland (1993), wherein two new binding principles are proposed:[2]

- Principle A: A reflexive-marked syntactic predicate is reflexive.
- Principle B: A reflexive semantic predicate is reflexive-marked.

Noteworthy here is Reinhart and Reuland's recognition that reflexivity is expressed at a semantic level as well as a syntactic one.

In order to fully grasp the Principles put forward by Reinhart and Reuland, some terminological clarification is in order. First of all, reflexive-marking can be realised in one of two ways. This can be either via the inherent reflexivity of the predicate (at the lexical level), or by having a *self* pronoun as one of the arguments:

[2]To disambiguate, I will refer to these throughout as Principles, and Chomsky's ABC's as Conditions.

(8) a. Jim shaved.

b. Jim shaved a slice of meat off the roast.

c. Jim$_i$ shaved himself$_i$.

Inherent reflexivity is shown in (8a). Here, where there is no other object in the sentence, the predicate is understood as reflexive, meaning that Jim shaved himself. This inherent reflexivity can be cancelled with the introduction of a different object, as in (8b). Inherently reflexive predicates can also be felicitously used with a reflexive pronoun, shown in (8c), though for some speakers, there may be a sense of redundancy. However, not all predicates in English have this inherent reflexivity:

(9) a. Jim$_i$ hit himself$_i$.

b. Jim hit.

In (9a), one of the arguments of this predicate is a reflexive pronoun, making the predicate *hit* reflexive. The lack of inherent reflexivity is clearly shown in (9b) where no direct object is specified. While native speakers may be able to debate whether this sentence has a meaning of either expressing Jim's tendency toward violence, or has an altogether different meaning of Jim striking some surface as a sort of projectile, there is no room here to argue that Jim hit himself in (9b).

Syntactic predicates are defined as those which have an external argument, along with all other arguments receiving case and/or a θ-role from that predicate, whereas semantic predicates refer to the predicate and all its arguments at the relevant semantic level. By taking this approach, reflexivity is not necessarily connected to syntactic structure, but gives more emphasis to the underlying semantic form.

These definitions allow for a sketch of the operation of these Principles. First considering the case of (9a), Principle A will only apply if there is a syntactic predicate; because there is an external argument to the sentence, this example qualifies. Furthermore, the *self* pronoun contributes reflexive-marking. The Principle is satisfied by the fact that both arguments of the predicate are coindexed. Principle B operates from the semantic level, recognising a situation in which there are two co-indexed arguments of the same predicate, and checking for reflexive marking. For (8a), Principle A cannot apply, as there are not two arguments between which co-indexing can be checked. Principle B is still met

though, as the underlying semantics of this predicate are transitive and reflexive, and the reflexive-marking is provided at the lexical level.

The definition of reflexive-marking captures a broader range of data than just Condition A, which only makes reference to *self* pronouns. For example, Condition A would have nothing to say regarding the difference between (8a) and (9b); by including inherent reflexivity, Reinhart and Reuland are able to capture through Principle B the underlying semantic similarity between (8a) and (9a) which is lost on Condition A. However, Reinhart and Reuland are also more restricted in that they are only concerned with elements appearing in argument positions, whereas Chomsky's Conditions are not so limited.

Reinhart and Reuland's Principles also capture some other facts which elude Chomsky's conditions. This can be shown with the following pair of grammatical sentences, both of which violate Condition A:

(10) a. There were five tourists in the room apart from myself.
(Reinhart and Reuland 1993, ex 22a)

b. She gave both Brenda and myself a dirty look.
(Reinhart and Reuland 1993, ex 24a)

Dealing first with (10a), both Principles are easily satisfied: this may be a syntactic predicate, but it is not reflexive-marked, as there is no *self* pronoun in any of its argument positions. Similarly, this is not a reflexive predicate at the semantic level, so Principle B is silent. In (10b), the reasoning runs in exactly the same way, capitalising on the observation that the direct object of this predicate is actually a conjoined element (which just happens to contain a *self* pronoun), and therefore the predicate does not count as being either reflexive or reflexive marked.

In both sentences, a pronoun would be equally acceptable:

(11) a. There were five tourists in the room apart from me.

b. She gave both Brenda and me a dirty look.

This replacement exemplifies a diagnostic proposed by Reinhart and Reuland: "whenever a logophor is possible, a pronoun is just as possible" (Reinhart and Reuland 1993, p.684). While this does not imply that anything which can be replaced by a pronoun must be a logophor, it does state that all logophors are replaceable by pronouns. This in turn means that

anything which cannot be replaced by a pronoun must not be a logophor. This reasoning forms the basis of Reinhart and Reuland's definition of 'logophor', another term which has multiple definitions in the broader literature. Their definition is roughly synonymous with 'exempt anaphor' referring to any *self* pronoun which is acceptable despite not conforming to Condition A. Alternatively, the term 'logophor' can refer to an anaphor which is obligatorily bound by a discourse participant (either speaker or addressee). More generally, logophoricity in this sense can refer to any sensitivity to perspective or point of view in the discourse. It is in this sense that I will use the term logophor; the cases identified by Reinhart and Reuland's pronoun replacement test, I will refer to as exempt anaphors, and treat as truly exceptional to the treatment of *self* pronouns.

An issue not taken up in any great detail by Reinhart and Reuland is that when there is this free variation between the *self* and referential pronouns in non-argument positions, what is it that triggers the use of the *self* pronoun? Zribi-Hertz (1989), drawing on examples collected from a variety of fiction and non-fiction sources, concludes that such examples are indeed logophoric, taking a covert subject of consciousness as their antecedents. This view is refuted by Baker (1995) who, using a different set of literature, argues that the reflexive forms are used when the referent stands in contrast with some other entities in the discourse. König and Siemund (2000b) support Baker's view, going on to hypothesise that because these uses tend to occur in non-subject positions (i.e. non-nominative positions), they could merely be a result of the observed inadmissibility of intensified accusative pronouns in English:

(12) a. he himself, *him himself

b. they themselves, *them themselves

Under this analysis, the exempt anaphors are contrastive elements, being used where a full emphatic reflexive is barred.

König and Siemund also comment on Reinhart and Reuland's generalisation that the exempt anaphors are restricted to non-argument positions, giving the following as examples of *self* pronouns in argument positions of non-reflexive predicates as explicit challenges to Reinhart and Reuland:

(13) a. And that was exactly it, he thought, he really did not care too much what happened to himself.

b. They would talk of himself, he thought fondly.

c. It was time to put an end to the burning. But to do so would put an end to himself as well. (Koenig and Siemund 2000b, ex 32a-c)

These examples are all cited from either the authors' own corpus work, or re-cited from Zribi-Hertz in the case of (13a). They claim that in response to such examples, the argument/adjunct asymmetry described by Reinhart and Reuland would need to be abandoned, or these examples and others like them would need to be written off as not being part of the core grammar. The examples in (13) do meet the pronoun replacement test for exempt anaphors: they can all be replaced by pronouns with no loss of grammaticality. König and Siemund provide a rationale for this test in that they treat such unbound reflexive pronouns as being underlyingly similar to the structures in (12), with an incorporated non-nominative pronoun. Under this analysis, such cases are no longer challenges for Reinhart and Reuland, as these would not be reflexive markers, but emphatic pronouns.[3]

At this point, a set of questions emerge around the issue of determining whether one or the other of these approaches is superior. Should Chomsky's structural rules be abandoned in favour of the more semantic approach of Reinhart & Reuland? Is it even possible for one or the other to capture all cases? If neither is fully adequate, what are the data patterns which would need to be captured by an all-encompassing theory, and which data can be set aside as truly exceptional? To go about answering these questions, data from a broader selection of languages are required. In the next section, I discuss the rationale for the selection of languages covered in this thesis.

1.3 A Typology of Reflexives

As shown above just with a few English examples, there is no one standard way in which reflexivity is expressed. Looking across languages, it becomes clear that there are even more possibilities. Reinhart and Reuland address this issue mainly in the form of their discussion of Germanic long distance anaphors, in this case from Dutch:

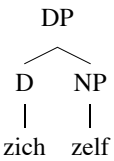
[3]Following this analysis through to its logical conclusion, any *self* pronoun in an accusative-marked position may have this non-reflexive-marking interpretation.

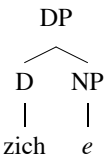
(14) a. Jan zag [jou achter zich/hem staan].
Jan saw [you behind SE/him stand
'Jan saw you stand behind SE/him.'

b. Jan haat zichzelf/*hem.
Jan hates himself/him
'Jan hates himself/him.' (Reinhart and Reuland 1993, ex 9-10)

The form *zich* here is what Reinhart and Reuland label as a simplex expression (SE) anaphor. The connection between *zich* and *zichzelf* is obvious from a morphological perspective, but the binding properties are different. Reinhart and Reuland go so far as to claim that the morphologically complex anaphors are universally locally bound, and the SE reflexives are universally long-distance bound.

Structurally, Reinhart and Reuland analyse the *self* pronouns of English and Dutch as complex DPs, and assign a parallel structure to the the SE anaphor[4]:

(15) a.


b.


Reinhart and Reuland further note that the SE anaphors are not only structurally analogous to referential pronouns, but they are also acceptable in the same binding domains. However, the SE anaphors lack ϕ features, and as such cannot have referential independence. These observations lead Reinhart and Reuland to propose a three-way typology of anaphoric expressions, based upon two criteria: Reflexivising Function (reflexive marking) and Referential Independence, defined as the ability to directly select an entity in the discourse.

As shown in Table 1.1, only the SELF anaphors are truly reflexive. Because the SE anaphors are obligatorily long distance, they can never enter into a co-argument relationship

[4]Reinhart and Reuland do not use DPs in their paper, but these structures are consistent with their description of the pronominal or SE elements occupying a determiner position.

Table 1.1: Reinhart and Reuland (1993) Typology of Anaphors

	SELF	SE	Pronoun
Reflexivizing Function	+	-	-
R(eferential independence)	-	-	+

with their antecedent, thus making reflexivity impossible. So while Reinhart and Reuland do put forward a typology of anaphors (in the broad sense of anaphor referring not just to those forms which fall under Condition A), they do not fully explore the possibility that there might be more than one way to express a reflexive relation, apart from their mention of inherently reflexive predicates.

A broader picture is painted in Déchaine and Wiltschko (2002b), where different forms of reflexive expressions are discussed. Their approach is both syntactic and semantic, as the different types of reflexives are differentiated according to the fine structure of pronouns elaborated in Déchaine and Wiltschko (2002a). They identify three levels: a simplex NP level, a level which projects to an intermediate ϕP, and finally a full projection to DP. As the reflexive expressions become more syntactically complex, there is a corresponding broadening in their usage, in that the forms with greater syntactic complexity are shown to have more functions. At each level, there is also a different semantic character to the reflexive relation.

The simplest level is the NP-reflexive, exemplified by Plains Cree. Here, reflexivity is not taken to be expressed by a referential DP, but rather by a nominal constant. The NP reflexive remains internal to the VP, undergoing a process akin to noun incorporation. As a result of this process, the valence of the predicate is reduced.

The basic noun incorporation pattern for Plains Cree is shown in the following minimal pair:

(16) a. kisîpêk-in-am(-w) wiyâkan.
wash-by.hand-I.TH-3rd dish
'S/he washes a/the dish.'

b. kisîpêk-in-iyâkan-ê-w.
wash-by.hand-dish-INTR-3SG
'S/he does the dishes.' or 'S/he washes a/the dish.' (Hirose 2003, ex 4.1a-b)

In (16a), the direct object *wiyâkan* stands alone at the end of the sentence, while in (16b) the noun has been directly incorporated into the verb. Also noteworthy here is that in (16a), there is indication on the verb of an inanimate theme, whereas in (16b), the verb is marked as intransitive, indicating that there has been a change of valence.

For Déchaine and Wiltschko, the Plains Cree reflexive is given a similar analysis. The reflexive marker *isô* indicates a direct fusion of argument postions, resulting in a reflexive reading:

(17) nipah-isô-w.
kill-REFL-3SG
'He kills himself,' (Déchaine and Wiltschko 2002, ex 19a)

Again, there is no distinct argument for the direct object here, and the sentence appears as an intransitive. As an intransitive, there is no room for an analysis which relies on a c-command relation between arguments.

The second level is that of the ϕP projection. ϕP reflexives are argued to have the semantics of bound variables, requiring a c-commanding operator to provide their reference. More generally, the ϕP category proposed by Déchaine and Wiltschko also covers referential pronouns, leading them to predict a lack of distinction in some languages between pronouns and reflexives. This is illustrated with a pair of examples from Haitian:

(18) a. Jean wè li.
Jean see 3SG
'Jean sees him/her/himself.' (Déchaine and Wiltschko 2002, ex 27)

b. Chyen$_i$ [m te kase pat li$_i$ a] te mòde m.
dog 1SG TNS break leg 3SG DEF TNS bite 1SG
'The dog whose leg I broke bit me.' (Déchaine and Wiltschko 2002, ex 28)

As shown in (18a), the pronoun *li* is ambiguous here between a referential and a bound reading which, as this is a single clause sentence, would be considered a case of reflexivity under the definition in (1). (18b) shows that *li* can be used as a resumptive pronoun, cited as a standard diagnostic for bound variable anaphora.

The category of ϕP is further broken down into two subcategories: local versus long distance cases. Looking first at the long-distance ϕP reflexive, Déchaine and Wiltschko examine the following data from Norwegian:

(19) a. Jon foraktet seg selv.
Jon despised REFL self
'Jon despised himself.'

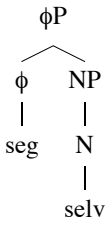
b. * Jon foraktet seg.
Jon despised REFL
Intended: 'Jon despised himself.' (Déchaine and Wiltschko 2002, ex 29a-b)

c. * Jon bad oss forakte seg selv.
Jon asked us despise REFL self
Intended: 'Jon asked us to despise him.'

d. Jon bad oss forakte seg.
Jon asked us despise REFL
'Jon asked us to despise him.' (Déchaine and Wiltschko 2002, ex 30a-b)

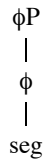
In Norwegian, the *seg selv* is used in local contexts, whereas the simplex *seg* occurs only in long distance environments. Syntactically, *seg selv* is analysed as a ϕP, with *selv* being an NP complement to a ϕ^0 head:

(20)



This syntactic structure captures the additional layer of complexity which differentiates *seg selv* from the Plains Cree case, but does not project up to the DP level which would give this full R-Expression status. The simplex form is treated as a head which projects all the way to a phrasal level:

(21)



In both its simple and complex forms, the reflexive in Norwegian must be bound; the difference comes in the (non-)locality of the binding relation.

As discussed in Reinhart and Reuland (1993) using parallel examples from Dutch, where there are such systems of simplex and complex reflexive expressions co-existing

within the same language, it is the complex one which is used in local domains, and the simplex which is used long distance. Déchaine and Wiltschko attribute this to the Blocking Principle of Williams (1997), which states that where two synonymous forms exist, the more specified one must be used. In this case, the availability of the more specified *seg selv* in the local context blocks the usage of *seg* in that position.

Under the terms defined in (1), the Norwegian *seg*, along with the earlier-mentioned Dutch *zich*, would not be classified as a reflexive, as it establishes a relationship between two arguments of different predicates, rather than working within one predicate. *Seg selv* however, does meet this requirement, as it is restricted to a local (clausal) domain. Indeed, the term "long-distance reflexive" seems in some sense to be an oxymoron, as reflexivity should be, under the definition in (1), a strictly local process. The treatment of these cases as being instances of bound variable anaphor allows for a reflexive-type reading to obtain in the local context, but allows for non-reflexive long distance uses as well.

The local ϕP reflexives are exemplified by the *s*-pronouns of Romance, a simple example of which can be drawn from French:

(22) Jean se voit.
Jean REFL see.
'Jean sees himself.' (Déchaine and Wiltschko 2002, ex 33a)

In the French example (and parallels from Spanish and Italian), reflexivity is marked through a clitic pronoun, though it is not the only function of the *s*-pronouns. Examining Spanish, the same *se* occurs in reciprocals, middles, inchoatives, and impersonal subject constructions:

(23) a. Juan y Maria se vieron.
Juan and Maria REFL see.3PL
'Juan and Maria see each other.'

b. Estros libros se venden bien.
these books REFL sell well
'These books sell well.'

c. Se abrió la puerta.
REFL opened the door
'The door opened.'

d. Se sabe que mentieron.
REFL know that lied.3PL
'It is known that they lied.' (Déchaine and Wiltschko 2002, ex 35a-d)

However, they go on to note that the *s*-pronouns are limited in their reflexive form to mark reflexivity for direct objects only, and that a phrasal form is used for PP-arguments. Furthermore, first and second person use distinct reflexive forms. In singular cases, there are unique accusative pronouns, but with the first and second person plural, the ordinary referential pronouns double as reflexives, shown again for French:

(24) a. Je me rase.
1SG 1SG.REFL shave.1SG.PRES
'I shave (myself).'

b. Tu te rases.
2SG 2SG.REFL shave.2SG.PRES
'You shave (yourself).'

c. Nous nous sommes vu.
2PL 2PL.REFL be.1PL see.PST
'We saw ourselves.'/'We saw each other.'

d. Vous vous êtes vu.
2PL 2PL.REFL be.2PL see.PST
'You saw yourselves.'/'You saw each other.'

Finally, they note that for the third person plural, a distinct form *les* is used for pronominal reference, while third person reflexives are *se*, regardless of number:

(25) a. Jean les a vu.
Jean 3PL have.3SG see.PST
'Jean saw them.'

b. Ils se sont vu.
3.PL REFL be.3PL see.PST
'They saw themselves.'/'They saw each other.'
(Déchaine and Wiltschko 2002, ex 38a-b)

However, there is an added complexity in that where the reflexive has a plural antecedent, there is ambiguity between a reflexive and a reciprocal reading, regardless of person, as shown in (24) and (25).

At the ϕP level, Déchaine and Wiltschko argue against the notion of there being a dedicated reflexive. From the Norwegian data, it is clear that a reflexive form (or at least a part of it) can be used in a non-local context, thus negating the claim that it is strictly reflexive. In the Romance cases, all instances of the reflexives remain local, but again, the same multiplicity of functions is observed with an outright ambiguity between reflexives and reciprocals, along with several other functions.

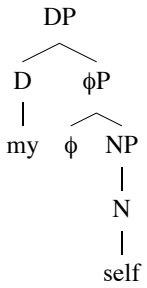
Finally, the DP reflexives are exemplified by English first and second person *self* pronouns:

(26) a. I_i hit myself$_i$.

b. You$_i$ hit yourselves$_i$.

Déchaine and Wiltschko assign to these reflexives the syntax of a possessive DP:

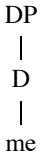
(27)



Once again, it is the articulated syntax and an appeal to the Blocking Principle which is used to account for the complementarity of reflexive and referential pronouns:

(28) a. * I_i hit me$_i$.

b.



The referential pronoun *me* is analysed as a simplex DP containing only a D^0 head, whereas *myself* contains a DP with a complement structure. Again, where the structures are synonymous, it is the more articulated one which wins out. The locality constraint on *self* pronouns is attributed to inalienable possession:

(29) a. I_i saw myself$_i$.

b. * I_i said that Lucy saw myself$_i$.(Déchaine and Wiltschko 2002, ex 8a-b)

(30) a. I raised my hand.

b. # I said that John raised my hand. (Déchaine and Wiltschko 2002, ex 9a-b)

(29) is a straightforward illustration of a Condition A effect, wherein the reflexive *myself* cannot be bound from outside its clause. Similarly, (30) shows that the inalienable possession reading for *my hand* does not survive across clausal boundaries. The sentence in itself is not ungrammatical, and the pronominal reference does not change, but the sentence does not carry the same reading that the speaker is in absolute possession/control of their own hand. Because the DP reflexives contain possessive pronouns, it is natural for Déchaine and Wiltschko to exploit this fact about inalienable possession in their explanation for why the DP reflexives must be locally bound.

These DP *self* pronouns are argued to be full R-expressions, a consequence of which is the fact that they are not exclusively limited to reflexive contexts. In addition to core coargument cases, two other non-reflexive uses of the *self* pronouns are reported. The first of these is the already-observed exempt usage. The second is a predicative use:

(31) You are not yourself today. (Déchaine and Wiltschko 2002, ex13a)

From this, they observe that the English *self* pronouns are not, as implied in the analysis of Reinhart and Reuland, restricted to representing a relationship between co-arguments, more in line with the König and Siemund analysis. Here, the *self* pronoun is predicated of the subject. This is argued to lead to the use of these pronouns as adjunct predicates in the emphatic or appositive reflexive construction:

(32) a. I myself saw Mary.

b. I saw Mary myself. (Déchaine and Wiltschko 2002, ex 14a-b)

Déchaine and Wiltschko make the claim that this is a parallel construction to other predicative adjuncts in English:

(33) a. John, tired, wrote the letter.

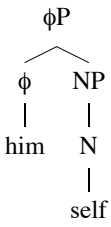
b. John wrote the letter tired. (Déchaine and Wiltschko 2002, ex 15a-b)

All of this is to illustrate that the DP reflexives are not strictly limited to reflexive contexts. A final consequence of this analysis is that as R-expressions, these are not, in fact, true reflexives. Rather, it is argued that the reflexive reading is a result of assigned coreference, a stipulation upon a predicate that while its arguments are distinct, they are externally equated.

Throughout their discussion, Déchaine and Wiltschko are careful to avoid mention of the third person reflexives in English, noting only in a footnote that the third person reflexive pronouns come out as ϕP reflexives under their analysis. The reason for this can be found in a closer examination of the D/ϕ/NP distinction, as presented in Déchaine and Wiltschko (2002a). There, it is argued that there is a split within the English pronominal system, with the third person pronouns receiving a different classification owing to their ability to participate in bound variable anaphora.

As bound variables, the third person reflexive pronouns would be analysed as ϕPs, rather than as full DPs; this is reflected in the lack of possessive syntax for the third person *self* pronouns:

(34)



However, the consequences of this split within the reflexive system are not explored in detail.

In total then, while the typology broadly defines three levels of reflexive, the NP, ϕP, and the DP, it in fact generates four distinct classes, in that the middle level, ϕP, was broken down between local and long-distance binding, shown in Table 1.2.

Note that defining the table along this local/long-distance dimension leaves two empty cells. A rationale for this can be found in examining the semantic correlates of the three different syntactic types of reflexive, summarised in (35):

(35) a. R[x, y], $x = y$ Assigned Co-reference (DP)

b. Opx[...x...] Bound Variable Anaphora (ϕP)

c. R[x, x] "True" Reflexive (NP)

Table 1.2: Déchaine and Wiltschko 2002b Typology of Reflexives

	Local	Long Distance
DP	English	N/A
ϕP	Romance	Norwegian
NP	Plains Cree	N/A

As noted, only the NP reflexive is considered to be true reflexivity, as it is the only case in which reflexivity is built into the predicate argument structure. For the ϕP reflexives, the identity of the two arguments has no direct connection to the predicate whatsoever, and it is a formalised relationship between the reflexive and its binder which creates the identity relationship, with reflexivity as defined in (1) obtaining only in local contexts. For DP reflexives, the mechanism is somewhat less clearly defined. While the semantic formulation of assigned co-reference is transparent, it is unclear where exactly this assignment originates. Based on the examples, it seems that the $x = y$ portion of the semantic form is not inherent to the predicate R, and so it must be contributed by one of the arguments. Thus, it appears that the DP reflexives in this analysis must also work on the predicate in some way, as do the NP reflexives. From this, both the DP and NP reflexive forms then are predicted to be restricted to local co-argument reflexivity. The middle level, where reflexivity is a result of bound variable anaphora thus interacts with the three logical possibilities for bound variables: strictly local, unconstrained, or strictly anti-local. Only the first two of these may generate reflexive meanings.

This typology makes for a good starting point, but it also raises some further questions. For English, a deeper exploration of the split in the pronominal paradigm is needed, to see whether this division between first and second on the one hand and third person on the other is tenable. Also, a closer look at the non-reflexive uses of the *self* pronouns is called for, to see how prevalent they are, and determine whether there are any more such functions. Finally, if it is indeed the *self* pronoun which contributes the assigned coreference reading, then an exact specification of the semantics of the *self* pronoun itself is required.

Looking to ϕP reflexives, one may question the suitability of the languages chosen. Keeping in mind the interaction with bound variable anaphora at this level and the defi-

nition of reflexivity in (1), the Norewegian *seg* examples, operating under an anti-locality condition, do not technically count as reflexive, as they illustrate a binding relation across two predicates. For the local cases, the Romance language reflexives do fit the paradigm, however there appears to be variation between languages, with not all non-reflexive uses attested for all languages, and a complex overlap between reflexive and non-reflexive pronouns, as well as with reciprocals. For the purposes of serving as an exemplar in a typological description, it would seem preferable to concentrate the analysis on a single language, with an aim to presenting a clearer overall picture. Still, two possibilities need to be explored for this level: a strictly local bound variable would be a natural choice for reflexive readings, while an unconstrained bound variable could serve as a reflexive in some, but not all, of its uses. Where bound by a co-argument, then such a variable would result in a reflexive reading.

Finally, the NP reflexives also leave some open issues. While the noun incorporation analysis of Plains Cree is straightforward enough, the presentation of the data in Déchaine and Wiltschko (2002b) is necessarily limited, and a more detailed analysis of the use of reflexives in the language may uncover new data patterns to account for. Specifically, given that there is an implication in this typology that the more articulated forms of the reflexive extend to more functions, in Plains Cree, it should emerge that those functions will be expressed using distinct non-reflexive forms. An examination of some of those functions and their expressions in Plains Cree may shed some light on why the ϕP or DP reflexives are able to take them over in other languages.

1.4 Modelling Reflexivity

In addition to facing the task of describing the use of reflexive forms in various languages, another key component of this thesis will be the final analysis in which the diverse syntactic expressions of reflexivity are shown to have parallel semantic forms. For this, a representation scheme which combines syntax and semantics is ideal. One such approach is Synchronous Tree Adjoining Grammar.

At the core of this formalism is Tree Adjoining Grammar, first formalised in Joshi et al. (1975), with the first major linguistics application coming in Kroch and Joshi (1985). In its barest foundation, TAG is a tree-rewriting system which has as its units elementary

trees, and two operations, Substitution and Adjoining, with which elementary trees are combined. As applied to linguistics, a lexicalised TAG is a TAG in which each elementary tree has one and only one lexical item. Elementary trees are then combined using the two tree-composition operations. These will be defined in more detail in Chapter 6.

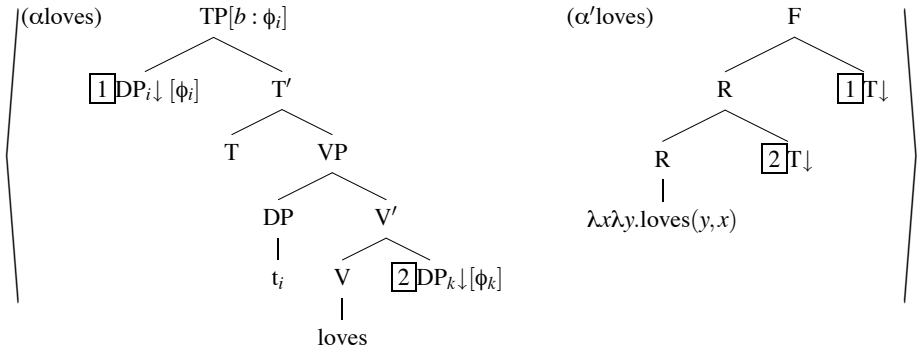
In Frank (2002), a TAG-based system of syntactic analysis is presented, showing that elementary trees built upon GB/Minimalist principles can be combined using the TAG operations to formulate a workable model of syntax with equal or greater explanatory power than Minimalism alone. Key to this analysis, and most relevant to the present discussion, is a basic constraint on the size of elementary trees. As in any lexicalised TAG, each elementary tree contains one and only one lexical head, but in defining a TAG θ-Criterion, Frank partially codifies the size of those elementary trees:

(36) TAG θ Criterion, (Frank 2002):
If H is the lexical head of elementary tree T, H assigns all of its θ-roles within T. If A is a frontier nonterminal node [substitution site] of elementary tree T, A must be assigned a θ-role in T.

This criterion stipulates that all and only the arguments of a given lexical head may be represented as substitution sites within the elementary tree headed by that lexical item. Thus, the argument structure of a given predicate will necessarily be reflected within a that predicate's elementary tree. Given that reflexivity as defined here is considered to be a relationship between two arguments of a single predicate, this formalism becomes a natural choice, as the structural domain of locality for reflexivity is hard-coded into the system: reflexivity must be expressed within a single elementary tree.

Synchronous Tree Adjoining Grammar (STAG) is an extension of lexicalised TAG, in which for each lexical item, there are now two elementary trees: one syntactic tree which is used in the derivation of the final output string, and a parallel tree from which the semantics is calculated (Shieber and Schabes, 1990). There is no set standard for the semantic representation; throughout the thesis I follow Han (2007) in representing the semantics using unreduced lambda expressions. Under Frank's θ Criterion, semantic trees are equivalently formalised, with a tree headed by a predicate again having substitution sites for its arguments. This is shown in (37):

(37)



Here is a simple tree pair for the English verb *love*. In the syntactic tree (αloves), the argument positions are empty DP nodes, with the downward-pointing arrow indicating that these are substitution sites. Note that the subject is base-generated in the [Spec, VP] position in this tree. Positing a [Spec, VP] node and showing movement from this position is entirely legal within TAG; tree-local syntactic movement is in fact necessary to account for certain phenomena (such as *Wh*-movement) within Frank's system. Note though that the formalism imposes a strict domain of locality on such movements in that they too must be contained within a single elementary tree.

The substitution sites for the arguments are marked with numerals in boxes. These numerals map onto nodes in the semantics tree, (α′loves), acting as derivational links. The semantic tree is much simpler, having only the denotation of *loves*, a predicate of type $<e,<e,t>>$, as a head of category R(elation) and substitution sites for two arguments of category T(erm). As DP arguments (their own elementary trees) are introduced to the syntax side, the semantic forms of those arguments, entities of type $<e>$, are substituted into the corresponding linked node on the semantics side. The crucial fact is that while there are two trees being derived, the derivational steps in the syntax and semantics are synchronous. The end result of the syntactic derivation is the syntactic string; the end result of the semantic derivation is a formula of type $<t>$. The semantic derivation converges only if, once all TAG derivation is completed, it is possible to complete the composition of the semantic tree.

The details of STAG derivation will be presented later in the thesis. This brief example is introduced here primarily to motivate the choice of STAG as a tool for analysing reflexivity. Reflexivity is an issue which lies squarely in the interface between syntax and semantics, a fact which is reflected in the formulation of the Reinhart and Reuland binding

principles which make reference to both levels. In presenting a simultaneous derivation of syntax and semantics in a framework that provides explicit links between the two modules, STAG is an ideal tool for modelling phenomena at this interface. Furthermore, TAG in general is an ideal tool for reflexivity, as one of the key constraints within the lexicalised TAG proposed by Frank is that there be constrained domain of locality which is definable in terms of a predicate's argument structure. The argument structure of a predicate, as argued in the previous section, is exactly the domain in which reflexivity should be confined.

1.5 Roadmap for This Thesis

This thesis is designed to be modular in nature; the intention is that readers who are interested in a particular chapter should be able to read that chapter in isolation without having to wade through the whole document. Using the Déchaine and Wiltschko typology as a guide, there are four possible types of reflexive, and thus this thesis consists of four language case studies: English, Korean, Shona, and Plains Cree. The selection of English and Plains Cree is intended to further expand upon the analysis presented by Déchaine and Wiltschko for those languages; in the case of English taking a closer look at the non-reflexive uses of the *self* pronouns, and in the case of Plains Cree, taking a closer look at some of those non-reflexive phenomena to which reflexive expressions often extend, and examining the alternative realisation of those phenomena. Korean and Shona, I argue, fill the two slots in the typology at the ϕP level, with Korean having an unconstrained bound variable whose use is sometimes reflexive, and Shona having a bound variable reflexive which is strictly local.

The languages selected for analysis are also intended to address an important gap in the STAG literature: cross-linguistic research. Publications making use of STAG, or even any lexicalised TAG, are almost exclusively restricted to Indo-European languages, particularly the languages of Western Europe. To the best of my knowledge, this thesis represents the first attempt at applying any form of TAG to a Bantu or an Algonquian language, representing an important test for the formalism. If STAG is to be advanced as a model for representing the syntax and semantics of natural language, then it should be shown to be robust enough to capture a diverse selection of languages.

The remaining chapters are as follows. Chapters 2 through 5 comprise the individual

language case-studies of English, Korean, Shona, and Plains Cree. Within each chapter, data come from a variety of sources, including published work, native speaker consultation, psycholinguistic experimentation, and corpus research. None of these chapters contain any STAG analysis; to keep the work as accessible as possible, the analysis in these chapters is framed in a GB/Minimalism approach. Chapter 6 then recaps the key data from each of the four case studies, re-casting the analysis within the STAG framework. Finally, Chapter 7 summarises the thesis, identifies remaining questions, and outlines future research directions.

Chapter 2

One Face, Many Functions
English *Self* Pronouns

This above all: to thine own self be true.
-Polonius. *Hamlet* by William Shakespeare.

In this chapter, I take a closer look at the use of *self* pronouns in English. After a brief summary of the English reflexive paradigm, and some key data presented in a discussion of the evolution of the analysis of *self* pronouns, I present the results of a corpus study which examines the use of *self* pronouns in written and spoken production. This corpus study leads to a more in-depth look at two of the more common non-reflexive uses: the aforementioned emphatic, and a manner adjunct use which occurs in a *by*-phrase. Because some non-argument *self* pronouns are potentially ambiguous between these two uses, a psycholinguistic experiment which seeks to identify some distinguishing characteristics of the two uses is introduced, and syntactic analyses for each of these two uses are proposed. Then, a semantic analysis is proposed, which treats the *self* pronouns as functions, and uses the selectional properties of those functions to account for the observed distribution of the *self* pronouns. The chapter concludes with a broader discussion of the status of English *self* pronouns, including a revisit of the issue of the anomalous status of the third person *self* pronouns, and some sociolinguistic of the the corpus research.

2.1 Background on English *self* Pronouns

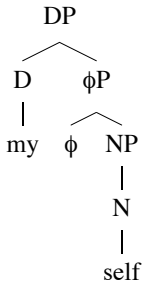
In this section, I briefly introduce the paradigm of English *self* pronouns, and review key analyses in existing literature. The full set of English *self* pronouns is presented in Table 2.1.

Table 2.1: English *Self* Pronouns

	Singular	Plural
1st Person	myself	ourselves
2nd Person	yourself	yourselves
3rd Person	him/her/it/oneself	themselves

The *self* pronouns are decomposable into a pronominal portion which is prefixed to the NP *self*, as in the Déchaine and Wiltschko analysis repeated below as (38):

(38)



With the exceptions of the third person *himself* and *themselves* (treated as ϕP by Déchaine and Wiltschko), the paradigm can be uniformly seen as a combination of a possessive pronoun and the *self* NP[1]. This possessive analysis is reinforced with examples making use of the more extended "own self" forms:

(39) The success of Iraq depends upon the capacity and the willingness of Iraqis to defend their own selves against terrorists. - George W. Bush, statement at NATO headquarters 22 Feb 2005.

[1]English *her* is ambiguous between an accusative and a possessive use, and the impersonal *it* and generic *one* have no differentiated possessive forms. If a possessive analysis were to be imposed *itself* and *oneself*, then these would underlyingly be *its self* and *one's self*, which could be argued to arrive at the observed forms by a process of phonological reduction.

Examples such as (39) reinforce the notion that there is a sense of possession inherent in the *self* pronouns, even for the third person cases. This underlying possessive syntax is used to motivate the necessarily local binding of English *self* pronouns, making the Déchaine and Wiltschko analysis somewhat unique in the wider literature, providing an external motivation for this often-discussed fact. In the next section, I take a brief look at the treatment of *self* pronouns in existing literature, where it becomes clear that for the most part the mission has been merely to correctly define the rules governing the position of *self* pronouns, with little emphasis placed on motivating those rules.

2.1.1 Reflexive Pre-history: Before 1981

Even before the theoretical advances which led to the proposal of the binding conditions in Chomsky (1981), the basic patterns of data with regard to *self* pronouns had been well-established in the literature. An oft-cited starting point for this line of research is Lees and Klima (1963). Prime among their observations was that *self* pronouns are sensitive to clausal boundaries:

(40) a. I told John to protect me. (Lees and Klima 1963, ex 15)

b. * I told John to protect myself. (Lees and Klima 1963, ex 16)

The examples in (40) are treated as originating from 'kernel' sentences, one per clause, in which the *self* pronouns begin as full nominal expressions. In explaining this data, Lees and Klima lay out a simple replacement rule wherein the second occurrence of a given nominal is transformed into the appropriate *self* pronoun if both are within the same kernel sentence. Despite the now-abandoned transformational machinery, the beginnings of the familiar binding conditions are apparent. This restriction of *self* pronouns to a kernel sentence which contains another instance of the same referent is a precursor to the definition of binding domain, and the interaction of their reflexive rule with a separate rule for generating referential pronouns is a first attempt at capturing the complementarity between those two forms. All in all, the sheer breadth of data presented in their paper presages much of the work that came over the following decades. Included in this are the classic alternation of reflexivity between the internal arguments of a *to*-dative versus a double object construction (later to be analysed by Barss and Lasnik (1986), arguing against a flat VP structure

for ditransitives) and data including *self* pronouns in embedded clauses which lack overt subjects, later to be treated as involving a PRO antecedent.

Less well-discussed are cases involving *self* pronouns in passive sentences:

(41) a. * Himself is shaved by John.

b. * John is shaved by himself.

The data in (41) are quite interesting. Lees and Klima mention these, with the reported judgements, as little more than an aside within the body text, speculating that such examples could be ruled out through constraints on passivisation. In present terms, (41a) can be explained as a straightforward Condition A violation in that the *self* pronoun is not bound. For Reinhart and Reuland, the situation is less clear; Principle A may not apply here, as the passive predicate can be argued not to have an external argument and is therefore not a syntactic predicate. Similarly, if the external argument is absent, this cannot be a reflexive semantic predicate, and the necessary condition for the application of Principle B is not met. However, even if both Principles A and B apply here, they are both met, as the necessary reflexivity and reflexive marking obtain (under the assumption that *himself* is coindexed with *John*. Looking at (41b), the situation is even more surprising. Here, *himself* has a c-commanding antecedent within its binding domain, so Condition A is met. Turning to Reinhart and Reuland, the reflexive here is in an adjunct position, so their Principles are silent on the issue of why this is ungrammatical. This example, and others like it, will be taken up later in the chapter.

2.1.2 Picture It: Representational DPs

Not even the most abbreviated discussion of *self* pronouns in English would be complete without at least a mention of the so-called "Picture NP" cases, or, more properly, representational determiner phrases (RDPs)[2]. This is an issue which Lees and Klima do not explore in any great detail, though it is clear they are aware of a sort of natural class of expressions, making reference to "constructions with *picture of*, *description of*, *book about*, *story of*, etc..." (Lees and Klima 1963, p 23). RDPs are a challenge in that *self* pronouns may

[2]Determiner phrase is the better choice, as the syntax of possessive structures necessarily comes into play.

sometimes appear further from their antecedents than otherwise expected, or they may be restricted to a far more local domain:

(42) a. Alan$_i$ said [that there was [a picture of himself$_i$] on the screen].

b. Alan$_i$ saw [Stephen$_j$'s picture of himself.$_{*i/j}$]

In (42a), the *self* pronoun is bound by *Alan*, even though there is a clause boundary separating them. In (42b), the *self* pronoun can only be bound by the possessor within the RDP, despite the fact that this is a mono-clausal sentence, and the subject should also be an available antecedent. Such examples are key as they motivate an important re-definition of binding domains.

Recalling the definition of Condition A, a *self* pronoun must be bound in its governing category. To account for the data in (42), the governing category is re-defined as a TP or DP containing an accessible subject. For (42a), the expletive *there* subject in the embedded clause does not meet this requirement (despite being in [Spec, TP]) and the long-distance binding of *himself* is permitted. For (42b), the presence of the possessor within the RDP establishes that the RDP itself is the binding domain. Under the Reinhart and Reuland analysis, *picture* becomes the relevant predicate. In (42a), the predicate only has a single argument, and the *self* pronoun is considered to be an exempt anaphor, while in (42b), both Principles are satisfied, in that the possessor provides the necessary conditions for the establishment of a syntactic predicate, and the reflexive marked predicate is reflexive at the semantic level. So, both a syntactic and an argument-structure based approach can deal with this data, though it appears that Reinhart and Reuland may be argued to do so more easily, as this is done without specifying different domains. Furthermore, the treatment of (42a) as an exempt anaphor correctly predicts that a pronoun would be equally acceptable:

(43) Alan$_i$ said [that there was [a picture of him$_i$] on the screen].

This alternation would not be expected in any analysis which assumes complementarity between *self* and referential pronouns.

In his discussion of RDPs, Runner (2007) proposes a test for exempt anaphora which capitalises on a subtle observation contained within Reinhart and Reuland (1993): an exempt anaphor should be capable of both bound variable anaphora and coreferential inter-

pretation, whereas a true reflexive *self* pronoun should behave only as a bound variable.[3] Runner uses sentences containing *only* to show that exempt anaphors in RDPs do show a "strict versus sloppy" ambiguity:

(44) Only Lucie liked the picture of herself. (Runner 2007, ex 38a)
Strict: Lucie is the only *x* such that *x* liked the picture of Lucie.
Sloppy: Lucie is the only *x* such that *x* liked the picture of *x*.

The strict reading is obtained when *herself* co-refers with *Lucie*, the sloppy when *herself* is bound by *Lucie*. This test verifies that (42a) can be treated as an example of exempt anaphora:

(45) Only Alan$_i$ said [that there was [a picture of himself$_i$] on the screen].
Strict: Alan is the only *x* such that *x* said there is a picture of Alan on the screen.
Sloppy: Alan is the only *x* such that *x* said there is a picture of *x* on the screen.

In (45), the same ambiguity emerges. With independent proof that the *self* pronoun in (42a) has the behaviour of an exempt anaphor, Condition A is "off the hook," and no new explanations are required to account for the example. Recalling the discussion of exempt anaphors from the previous chapter, it is examples such as this which are at the core of the case for considering exempt anaphors in English as logophoric.[4] This fits the pattern in that the antecedent *Alan* is the source of the reported statement.

However, there are other more subtle phenomena which neither Chomsky nor Reinhart and Reuland can easily deal with. The first of these concerns a pair of examples credited to Kuno (1987):

(46) a. Mary$_i$ isn't interested in anybody's opinion of herself$_i$.
b. * Mary$_i$ isn't interested in John's opinion of herself$_i$. (Runner 2007, ex 16a-b)

According to Runner, this contrast is not explained under the Reinhart and Reuland principles, which would predict both sentences to be ungrammatical. Similarly, Condition A

[3]Though he uses this test, Runner does admit that the observation on which it is based is disputable.

[4]Most likely, this is the reason the two terms appear to be used interchangeably, as there is a significant overlap in English.

should rule both of these out as the *self* pronouns are not bound within their binding domains. The reason for this contrast is described as an issue of referentiality; *anybody* does not block the potential binding between the reflexive and the subject in the same way as *John*. However, unlike the expletive example in (42a), it is not the case that *Mary* is the only (or even the closest) potential antecedent. An alternative reading for (46a) is available where *herself* is bound by *anybody*, which would be in line with both Chomsky and Reinhart and Reuland. Again, a simple pronoun replacement test can show that under the reading where *Mary* appears to bind *herself*, this is actually a case of exempt anaphora:

(47) Mary$_i$ isn't interested in anybody's opinion of her$_i$.

What makes this seem exceptional though is that it is apparently evidence of exempt anaphora in an argument position of a transitive predicate. However, this is not too different from the cases of intensified pronouns (treated as exempt) discussed by König and Siemund.

A second troublesome alternation can be seen between verbs of perception and creation when combined with an RDP. The examples in (48) are adaptations of a pair from Tenny (2003) which is credited to Jackendoff (1972):

(48) a. I$_i$ told the story about myself$_i$ that John likes to hear.

b. * I$_i$ hate the story about myself$_i$ that John always tells.

Here, the reflexive is acceptable where the RDP is the theme of a verb where the antecedent has agency in the telling, but not when the antecedent is merely perceiving the story. For both Chomsky and Reinhart and Reuland, these should both be fine, as Condition A is satisfied on the one hand, and on the other, the intransitive predicate *story* is not subject to the Reinhart and Reuland Principles. This also runs counter to the Zribi-Hertz notion of a subject of consciousness, as the reflexive is ruled out in (48b) where the antecedent's own mental state is being reported. To account for this apparent change in the binding domain between the two examples of (48), Tenny proposes that experiencers raise to a higher position in the left periphery. In this case, the raising would take place in (48b), drawing the antecedent too far away from *myself* to satisfy Condition A (though for the analysis to be tenable, there would have to be an exclusion of binding by the trace of this movement). Runner gives a parallel example to this showing an alternation between RDPs

under *know* versus *not know*, but the judgements are reversed, and it is the affirmative case which is acceptable, and the negated case which is unacceptable. Binding by a knower is at least more in line with the notion of a subject of consciousness, but it then raises the question of why a knower is not an experiencer, as Tenny would predict an experiencer to not provide a local antecedent. One could hypothesise that in (48), a preference for a source antecedent in an RDP is the actual source of the contrast, making the teller an acceptable antecedent, but not the hater. Where there is no overt source or goal, as would be the case for a verb like *know*, then the subject-of-consciousness constraint could be active, explaining why an experiencer is in fact a possible antecedent.

The current state of the art in work on RDPs can be credited to Kaiser et al. (2009), which reports on a series of experiments on reflexive and referential pronouns embedded within non-possessed and possessed RDPs, examples of which are repeated for reference:

(49) a. Peter told Andrew about the picture of {him/himself} on the wall.

b. Peter told Andrew about Greg's picture of {him/himself} on the wall.
(Kaiser et al. 2009, ex 8)

For the non-possessed cases, Kaiser et al. claim that *self* pronouns should be expected to default toward subject binding, whereas referential pronouns should have an anti-subject bias in the same sentence, despite the Condition B violation. Experiments were conducted to test whether these biases would interact with the kind of verb effects seen in (48). Designating these semantic effects as the "source" and "receiver" hypotheses, Kaiser et al. test whether the *self* pronouns are more likely to take source antecedents while referential pronouns are more likely to take receiver antecedents. In (49), the sentence structure is congruent with these two hypotheses: the subject is also a source, and the indirect object is a receiver. Kaiser et al. constructed experiments testing the sentences in (49), along with sentences where *told* is replaced with *heard from*, which makes the source/receiver dimension incongruent with the subject/object dimension.

In their experiments, the test sentences such as those in (49) are presented aurally, along with a picture of the scene being described alternating the figure shown in the picture on the wall. Forced-choice tasks where participants must select an antecedent are combined with eye-tracking to monitor the participants' fixations on the image presented. The use of eye-tracking is justified as it "provides a more sensitive measure of participants' interpretations

and therefore helps to minimize the problem of an effect being masked by the forced-choice situation" (Kaiser et al. 2009, p 71). The addition of eye-tracking revealed three important findings. The first of these is that overall, with *self* pronouns, participants tended to look more to the sentential subject, and with pronouns, looks tracked more toward the indirect object. This confirms the existence of the expected structural bias. Secondly, eye-tracking indicated that very early on, the semantic effects are present for both *self* and referential pronouns, within this overall structural effect. In both cases, there is an initial jump in looks to the referent favoured by the semantic source/receiver bias, followed by a period of convergence, and then returning again to a gaze pattern consistent with the semantic bias. Finally, the gaze pattern showing the effect of the semantic bias was present even in cases where participants ignored that bias and made selections based on the structural bias. So, while the forced choice task suggests that the semantic bias is weaker for *self* pronouns, the eye-tracking results show that the participants' reactions to the stimuli are still impacted by this effect, even though it is not reflected in their final responses.

This discussion of RDPs indicates that while structure, be it articulated in terms of syntactic c-command or co-argument relations at a semantic level, is a good predictor of the distribution of *self* pronouns in English, it is not the only factor. Room must be made for other semantic effects, as shown in this discussion of the *self* pronouns' sensitivity to information source, or agency. Even when such factors do not dictate the final interpretation, they are present in the interpretation process. This consideration may allow apparently exempt cases such as (42a), repeated below as (50) to be brought in line with a more inclusive theory:

(50) Alan$_i$ said [that there was [a picture of himself$_i$] on the screen].

If indeed there is some semantic bias leading *self* pronouns in English to prefer antecedents which are sources of information, then that bias could be used to explain why it is that the reflexive in this sentence is able to "escape" Condition A. To test this, it is simply a matter of re-casting the example with a different verb:

(51) ? Alan$_i$ heard [that there was [a picture of himself$_i$] on the screen].

To my native speaker judgement, there is a degradation of the sentence when the *self* pronoun is bound by a receiver subject of a higher clause. Thus, it appears to be the case that

some violations of the structural binding theory can be attributed to non-structural semantic factors.

Runner (2007) notes that the interplay of *wh*-movement and *self* pronouns within RDPs has played a role in the development of the Minimalist model of grammar, based upon the following examples adapted from Chomsky (1995):

(52) a. John wondered [which picture of himself] Bill saw.

b. John wondered [which picture of himself] Bill took.

Chomsky notes that in both cases, the *self* pronoun is ambiguous, with either *Bill* or *John* serving as potential antecedents. However, (52b) is doubly ambiguous, in that *took* can be understood as literal "taking" or as the idiomatic "take a picture", and under the second reading, the ambiguity of the *self* pronoun disappears, with *Bill* being the only possible antecedent. Under Chomsky's analysis, the ambiguity of the *self* pronoun comes through the construction of two possible LF configurations for (52b):

(53) a. John wondered [which x, x a picture of himself] [Bill took x]

b. John wondered [which x] [Bill took [x picture of himself]].

For each of these LF representations, Condition A would dictate a different antecedent for *himself*, John or Bill, respectively. However, only (53b) yields the idiomatic reading of "take a picture", as it preserves the structure of the idiom at LF. This then predicts the lack of ambiguity for the *self* pronoun, as there is no LF representation which preserves the idiom, and allows John to bind the *himself*. This line of argumentation serves as proof that the binding conditions must apply at LF, which in turn serves as an argument for the abolition of S-Structure from the Extended Standard Theory of the 1980's. Looking back at this in light of the Kaiser et al. findings, it appears that an appeal to the sourcehood or agency of *Bill* in (52b) under the idiomatic reading could likewise explain the unavailability of *John* as an antecedent without having to delve into this argument of different structures at LF. It is for precisely reasons such as this which Runner advocates the study of RDPs; because facts about *self* pronouns have motivated various structural analyses such as the line of work originating from Barss and Lasnik through to this argument from Chomsky. If it turns out that there is an alternative analysis for these data, then that could call into question the validity of the research built on those data. Essentially, his point is that a structural analysis based on binding facts is only as solid as the existing analysis of binding.

2.1.3 Long Distance? Not Exactly

RDPs are not the only cases in which *self* pronouns appear to take antecedents across clause boundaries in English. This section will conclude with a discussion of some of these cases, including questions, ECM, control, and raising structures.

In questions that involve multiple clausal embedding, a *self* pronoun may appear to have its antecedent in a higher clause:

(54) Who$_i$ did Dara claim [that John predicted [would humiliate themselves$_i$?]]

At first glance, the antecedent for *themselves* is two clauses higher than the reflexive, violating Condition A. Locality is restored through the mechanism of cyclic *Wh*-movement, in which the trace (or copy, under some theories of movement) at the origin of *who* serves as a local antecedent for the reflexive. This likewise would provide the necessary external argument for a syntactic predicate, bringing the example in line with the Reinhart and Reuland principles as well.

Exceptional Case Marking (ECM) sentences also present a challenge to binding conditions:

(55) a. Sandy$_i$ wants herself$_i$ to win the round.

b. * Sandy$_i$ wants her$_i$ to win the round.

As shown in (55), an ECM sentence in which the embedded clause subject is co-indexed with the matrix subject must use a *self* rather than a referential pronoun. As such, this cannot be written off as a case of exempt anaphora, and must be accounted for within the binding conditions.

Under the simple definition of treating a clause boundary as a binding domain, this clearly runs afoul Condition A. There are two possible solutions to this problem, the first of which is an analysis that tweaks the definition of binding domain so as to consider the entire sentence to be one binding domain. This is consistent with the fact that the clause boundary is already transparent to case assignment; this same transparency could extend to the binding domain. The second solution is to follow the route of re-casting ECM as a raising to object construction, where the embedded clause subject raises into the upper clause for reasons of accusative case assignment in Spec-Head configuration with the case

assigner, the matrix clause *v*. After this movement, the embedded clause subject would land in a position where binding from the matrix subject is unproblematic.

For Reinhart and Reuland, the situation is more complex. Under their definition of a syntactic predicate, the embedded clause subject counts as an argument of the matrix clause by virtue of the case assignment, no matter how it may be represented structurally. As such, the upper predicate is reflexive-marked and reflexive, so Principle A is met. The problem is that in their terms, the embedded subject is also the external argument of the embedded syntactic predicate, thus reflexive-marking a non-reflexive predicate. To escape this violation of Principle A, Reinhart and Reuland are forced to posit an LF verb movement operation by which the embedded clause verb undergoes head movement to the matrix clause, resulting in a combined predicate (*win-want*) in the case of (55a), which is appropriately reflexive and marked as such.

A situation parallel to the ECM cases can be found in cases such as (56):

(56) a. Jimmy$_i$ would like very much for himself$_i$ to win.

b. * Jimmy$_i$ would like very much for him$_i$ to win.

As discussed in Reinhart and Reuland (1993), this example is handled under Conditions A and B by yet another redrawing of the binding domain to include the entire sentence. With this, the reflexive is the only possible choice for the embedded clause subject. However, there is no clear way for Reinhart and Reuland's principles to account for the data. They themselves note that there is no way to construe *Jimmy* and *himself* as co-arguments, and yet the embedded clause is a reflexive-marked syntactic predicate. At this point, Reinhart and Reuland state that because of the marginality of this structure, it is not necessary to account for it. Specifically, they cite the availability of a subject control equivalent as evidence that there is no need to account for the structure in (56).

While it might solve the issue of (56) by eliminating the reflexive altogether, subject control can also create structures having the appearance of long-distance binding of a reflexive:

(57) a. Carrie$_i$ promised [PRO$_i$ to control herself$_i$.]

b. * Carrie$_i$ promised [PRO$_i$ to control her$_i$.]

For Condition A, the requisite local antecedent is provided by PRO. For Reinhart and Reuland, the crucial decision is made by Principle B, as PRO, being caseless, would not create the syntactic predicate necessary to apply Principle A. Counting PRO as a semantic argument though, the embedded predicate is reflexive, and thus must be reflexive-marked, ruling out the referential pronoun.

A similar case which has the appearance of long-distance binding involves raising predicates:

(58) a. Jo$_i$ seems [t$_i$ to despise herself$_i$.]

b. * Jo$_i$ seems [t$_i$ to despise her$_i$.]

Again, this is resolved with an appeal to movement, with the trace of *Jo* in the embedded clause serving as the local antecedent for *herself*, in line with Condition A. Similarly, Reinhart and Reuland's Principles are met; the embedded predicate is reflexive, and appropriately reflexive-marked.

More challenging are cases where the *self* pronoun is an argument of the raising predicate:

(59) a. Rich$_i$ seems to himself$_i$ [t$_i$ to outperform his rivals.]

b. * Rich$_i$ seems to him$_i$ [t$_i$ to outperform his rivals.]

Where the experiencer of *seem* is coidexed with the raised subject, a *self* pronoun must be used, as shown in (59). For Condition A, this is not a problem, as the raised subject binds the reflexive within the matrix clause. The situation is again more complicated for Reinhart and Reuland. Case assignment makes *Rich* a part of the matrix syntactic predicate, so Principle A is satisfied. Principle B, on the other hand, looks for a reflexive semantic predicate to apply; there being none in (59a), it goes through. This is a subtle case where there is nothing in the Reinhart and Reuland analysis which forces the presence of a *self* pronoun. While the Chomskyan Condition B rules out (59b) owing to the fact the moved subject c-commands the experiencer within its binding domain, neither of the Reinhart and Reuland principles rule out this example. Principle A applies only to reflexive-marked syntactic predicates, and is thus blind to (59b). Similarly, Principle B does not apply to (59b) either, as there is no reflexive semantic predicate. To rule out this example, along with other instances of infelicitous use of referential pronouns, Reinhart and Reuland must

augment their system with a purely structural constraint on A-chain formation. They argue that in (59b), {*rich, him, t*} all form an A-chain. The example is ruled out by the following condition:

(60) GENERAL CONDITION ON A-CHAINS
A maximal A-chain (α_1, ... ,α_n) contains exactly one link –α_1– that is both +R and Case-marked.

Here, +R stands for referential independence, which Reinhart and Reuland ascribe to both referential pronouns and R-Expressions. Thus, the Chain Condition rules out (59b) by virtue of there being two +R members of the chain. By ruling out referential elements, the Chain Condition defines structural positions in which only a *self* pronoun is permitted. In light of the Déchaine and Wiltschko definition of DP reflexives as having R-Expression status, the Chain Condition will need to be modified if not completely abandoned.

Despite the mileage which Reinhart and Reuland derive from adopting their Principles, which are shown to account for some cases of antecedentless *self* pronouns which cannot be captured by Condition A, they too are forced into a position of relying on a notion of syntactic structure to explain the distribution of *self* pronouns. Still, as shown by the recent work on RDPs, even relying upon syntactic structure and predicate-argument structure will not be sufficient, as it has been argued that external semantic factors such as source versus receiver will play a role in the licensing of *self* pronouns. Source and receiver can be manifested as a part of the argument structure of a predicate, but they are crucially not a part of the RDP predicate containing the *self* pronoun. Thus, this survey of the literature indicates that the semantic relation between arguments, syntactic structure, and external semantic information must all be available when formulating principles to explain the distribution of *self* pronouns.

2.2 Corpus Study

While the review of existing literature provides a snapshot of the various phenomena surrounding *self* pronouns in English which have caught the attention of researchers, this may not be an accurate gauge of the actual use of *self* pronouns. Indeed, the use of the term "exempt" suggests a state of affairs wherein the majority of uses of *self* pronouns conform

to the patterns laid out by Condition A and the Reinhart and Reuland principles, but there is no verification of this. Exempt anaphors often appear in popular media:

(61) a. And I think that myself or any other player is allowed to play those clubs because they're approved – end of story.

b. He will help Stevie, Dirk [Kuyt] or myself to score goals because he gets a lot of assists.

c. The year which is about to end has been a positive one for myself and for Milan, but 2010 will be even better.

The above examples represent just a sampling of the exempt anaphors uncovered through a rough search for instances of the word *myself* in one month's worth of stories on the www.cbc.ca/news website. The methodology is by no means exhaustive, but it shows that these uses of *self* pronouns are not hard to find. Indeed, Zbiri-Hertz describes a corpus of over 100 examples of Condition A violations gathered in her casual reading, and C. Baker, along with König and Siemund, all present large sets of such examples harvested from literature and the British National Corpus. What is missing from these though is a sense of proportion: they report only on the exceptions to Condition A or the Principles of Reinhart and Reuland, without making any reference to data which *do* conform to those analyses. To get a more accurate sense of how frequent these and other uses of *self* pronouns are, a more detailed study is undertaken.

2.2.1 Selecting and Analysing the Corpus

For this project, the Treebank 3 (Marcus et al., 1999) corpus was used. Specifically, samples from two different sections of Treebank 3 were used. First, the entire Wall Street Journal (WSJ) corpus, amounting to 1,000,000 words of printed text was analysed. This analysis was then repeated using a portion of the Switchboard (SWB) corpus. Switchboard consists of approximately 3,000,000 words of text, transcribed from five minute telephone conversations held between strangers and moderated by an automated computer system. Here, a 545 conversation subset was used, amounting to roughly one-third of the overall corpus, to achieve a comparable word count with the WSJ corpus.

For both corpora, perl scripts were written to identify *self* pronouns. Due to the structure of the corpus data, a slightly different method was used in each case. For the WSJ corpus,

sentences containing *self* pronouns were extracted. For SWB, entire conversational turns were extracted, so long as there was at least one instance of a *self* pronoun somewhere in that turn. In some cases, there were extracts (sentences or turns) which contained more than one *self* pronoun; each was considered a separate token for the analysis. Seven instances of the phrase *do-it-yourself* (as in a do-it-yourself home repair manual) were eliminated from the analyses, taken to be frozen complex expressions rather than instances of *self* pronouns. In total, this yielded a total of 496 tokens from the WSJ corpus and 575 from the SWB. These were annotated for such variables as ϕ features, theta role, and associated predicate. Additionally, a set of functional categories for the *self* pronouns was defined, and used as the basis for further analysis.

In all, there were eleven distinct categories of reflexive use defined in this study. Each will be discussed in turn, along with illustrative examples from the corpus itself. In some cases, these categories were defined at the beginning of the study, while some of the classifications evolved through the course of the analysis. The categories defined here, along with the criteria for inclusion, were applied to the corpus as a whole on a final pass through the data.

The first category is the canonical one, where the *self* pronoun appears in an argument position, having the same reference as one of its coarguments:

(62) ...because the government has not converted **itself** into a modern, democratic, "developed nation" mode of operation. (wsj_1120: 2)

Closely related to these cases are those where the *self* pronoun is again in an argument position, but of a structure involving more than one predicate. This category covered cases of ECM and control:[5]

(63) a. But they believe **themselves** to be serving. (wsj_2412: 77)

b. I have to force **myself** PRO to do it. (SW2603.DFF: A.69)

This category also included cases where the *self* pronoun was in the embedded clause, with PRO as its antecedent. As such, these could have been considered co-argument cases of reflexivity in the lower clause, but due to the added complexity of the structure, and

[5]Raising would have also been included here, but the corpus contained no instances of *self* pronouns being used in sentences with a raising predicate.

the seeming long-distance nature of the relation, they were left in the bi-clausal group. Similarly, the example of Object Control above is simply a matrix clause reflexive, but for the sake of unity, all control clauses were kept together. Also included here were cases where the reflexive is modified by a second predicate:

(64) I really don't see **myself** going back to school or anything. (SW2744.DFF: A.152)

While this example does not fit into the strict ECM/Raising/Control paradigm in terms of having an non-finite embedded clause, there is still a sense in which *myself* is an argument of *going* while staying as a part of the matrix syntactic predicate by virtue of case assignment. This type of example actually makes up a sizeable portion of this overall categorisation.

The next category was the representational DPs. Sentences placed into this category fit the basic schema of containing a DP wherein the *self* pronoun was the complement of some representational noun:

(65) ...people whose ignorance and intellectual incompetence is matched only by [their good opinion of **themselves**]. (wsj_1286.mrg: 10)

As shown in this example, this class is not strictly limited to pictures *per se*, but any DPs of similar structure. In order to fit into this category, the antecedent did not have to be local to the DP as a specifier, but the presence or absence of the possessor was noted.

The other major exception to the standard binding theory, exempt anaphora, was also categorised from the beginning:

(66) And, uh, a great disappointment for some people like **myself**. (SW2379.DFF: B.16)

In order for tokens to be placed in this category, a replacement test was used. Only those *self* pronouns which could be felicitously replaced with a ϕ-feature equivalent referential pronoun were included. Tokens for which this test failed were placed into some other category; none of the other *self* pronouns in the study are thus considered exempt.

Another category which the Déchaine and Wiltschko discussion would predict to find was the emphatic use of the *self* pronoun:

(67) The classroom **itself** operated on the periphery of this awful system... (wsj_1315.mrg: 30)

As described earlier, the *self* pronoun is not contributing any new content to the sentence, it is merely putting additional emphasis on its antecedent. This use of the *self* pronoun was found in the expected two forms: either adjacent to its antecedent (shown above), or with the *self* pronoun appearing at the end of the sentence.

One category which emerged through the course of the analysis was that of manner adjuncts:

(68) ...find an indoor pool where either you can do this by **yourself**... (SW2382.DFF: B.32)

In this case, the *self* pronoun appears in a *by*-phrase, adding information to the manner in which the action took place. Specifically, it adds that the action was carried out alone. As with the emphatic case, there is an additional form for this usage as well, in which the *by* preposition is elided. Crucially, unlike locative adjuncts (the classic "snake beside himself" sentences) these fail a pronoun replacement test, and so cannot be considered exempt.

Classified separately were cases of other prepositional phrase modifiers containing *self* pronouns:

(69) Many of the affluent aren't comfortable with **themselves**. (wsj_2366.mrg: 43)

Some of these cases had close to argument status, occasionally representing a beneficiary. Crucially, this category only included cases where the sentence's meaning would be changed if the *self* pronoun were replaced with a referential pronoun. In the case of (69), the referent for the people with whom the affluent are not comfortable would necessarily change if *themselves* were changed to *them*. Baker (1995) reports a similar judgement, and is reluctant to consider beneficiaries as exempt anaphors, mentioning their close-to-argument status.

Another type of modifier that was identified was the case where the *self* pronoun is in a prepositional phrase modifying a nominal:

(70) ...gives you some space for **yourself**... (SW2072.DFF: B.46)

Here, the PP is modifying the nominal *space* rather than the predicate itself. Again, this cannot be considered exempt, as a pronoun replacement test fails.

Another category involved cases where the reflexive was in a copular complement:

(71) I am beside **myself** (wsj$_0$403.mrg: 5)

As in the other cases, tokens were only placed in this category if they did not pass the test for exemption.

Two final categories were created to account for cases which did not fit into any of the other categories. The first of these was for cases which appeared to be dysfluent uses of a *self* pronoun:

(72) Do **youself** have children wi-, who are or have been through the public school system? (SW2828.DFF: B.13)

While this sentence is improved by the replacement of *youself* with *you*, this does not strictly fit into the exempt case. The exempt cases are still considered to be grammatical with the *self* pronoun being a matter of stylistic choice. Here, my native speaker intuition is that there is something wrong with the sentence.

Finally, there was an 'other' category for cases which did not fit any of the above criteria. These however turned out to be quite exceptional:

(73) a. ...all the firm has to do is "position **ourselves** more in the deal flow." (wsj_0604.mrg: 38)

b. ...always try to stump Jesus, try to give him something that would contradict **himself**. (SW2260.DFF: B.44)

In the first of these examples, from the written corpus, *ourselves* does not strictly speaking have an antecedent, but because it appears in a fragmentary quotation, it can be assumed that in the original quotation there was a proper antecedent (i.e. the sentence being quoted most likely started with 'we.'). The second case is more difficult to diagnose, though it is not as strikingly ungrammatical as the sentence in (72). Perhaps there is an ellipsis at work here, concealing structure above the *contradict* clause which would render the sentence easier to parse, and provide an antecedent for the reflexive.

2.2.2 Findings

The counts for each corpus are summarised in Table 2.2. Looking at the first two categories, the Coraguments and the Multiple Predicate cases, combining them yields the total number of tokens wherein the reflexive appeared in an argument position. What is immediately striking about this is the fact that the total amounts to only 60% of the occurrences in the written corpus, and just over half of the occurrences in the spoken corpus. This poses clear challenges to the Reinhart and Reuland analysis, in that it makes clear that close to half of the uses of *self* pronouns in English cannot be accounted for using a purely coargument-based analysis of their distribution.

	WSJ		SW	
	Tokens	%	Tokens	%
Coarguments	241	48.59	216	37.57
Multiple Predicates	64	12.90	77	13.39
Representational DP	4	0.60	1	0.17
Exempt Anaphor	1	0.20	21	3.65
Emphatic	142	28.63	121	21.04
Manner Adjunct	21	4.23	96	16.70
Other Adjunct	13	2.62	23	4.00
Nominal Modifier	7	1.41	12	2.09
Copular Construction	2	0.40	3	0.52
Dysfluent	0	0	3	0.52
Other	1	0.20	2	0.35
TOTAL	496		575	

Table 2.2: Distribution of *Self* Pronouns across WSJ and SWB Corpora

Looking next at the more widely-discussed counter-examples, there is again an unexpected finding. Across both corpora, with a total of 1071 tokens, there were only five instances of *self* pronouns within an RDP. Of these, only two were bound locally within the DP, the others being possessorless and bound within their clause. Similarly, there were only 22 instances of exempt anaphors, under the criteria laid out above. These are unexpectedly low numbers given the amount of attention these types of examples receive in the general

literature. One might expect them to occur frequently, but instead, the results suggest that these are quite rare phenomena.

After the A-positions, what emerge to be the most common uses of the *self* pronouns turns out to be the emphatics and manner adjuncts. While in both corpora, the emphatics are more numerous than the manner adjuncts, their distributional patterns between the spoken and written corpora are opposite. The emphatics are more frequent in the written corpus, while the manner adjuncts are more frequent in the spoken, by a much wider margin.

The remaining categories were relatively infrequent. Out of all this, the most unexpected finding was the prevalence of the emphatics and manner adjuncts. Furthermore, these proved to be among the hardest to classify, due to the existence of alternate forms for each. It is to this issue which I now turn.

2.3 Emphatics and Manner Adjuncts

In this section, I take a closer look at the two most frequent non-reflexive uses of the *self* pronouns: the emphatics and the manner adjunct. First of all, owing to their frequency, they are worthy of further study; while it is true that RDPs are worthy of attention because of their importance in the development of binding theory, they are a comparatively rare phenomenon. Secondly, these examples are particularly challenging to Reinhart and Reuland, as they are not argument positions, and yet the use of a referential pronoun is impossible:

(74) a. The classroom **itself** operated on the periphery of this awful system... (wsj_1315.mrg: 30)

b. * The classroom it operated on the periphery of this awful system...

(75) a. ...find an indoor pool where either you can do this by **yourself**... (SW2382.DFF: B.32)

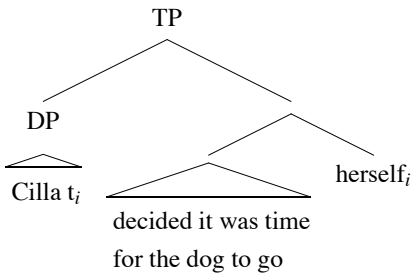
b. * ...find an indoor pool where either you can do this by you...

Looking first at the emphatics, recall that they can appear in more than one position. In the canonical use, the emphatic *self* pronoun occurs immediately following its antecedent as in (74a). However, it can also appear at the end of a sentence:

(76) Cilla t_i decided it was time for the dog to go herself$_i$.

Following the analysis of Bickerton (1987), this can be treated as a simple extraposition process, schematised in (77):

(77)



Assuming that this extraposed position is low enough in the syntax, this presents no problem for Condition A; the *self* pronoun will still be c-commanded by a subject antecedent.

Turning to the manner adjuncts, there is variation in their usage:

(78) Chesney$_i$ filled Schmeikel's bowl himself$_i$.

The *by* preposition in this construction appears to be optional. However, this does nothing to the position of the *self* pronoun and so has no effect on the binding relations.

In and of themselves, these variable uses of the non-argument *self* pronouns seem to be unremarkable. However, they do conspire to create potentially ambiguous sentences as shown in (79):

(79) a. David$_i$ obtained the contraband himself$_i$.

b. David$_i$ himself$_i$ obtained the contraband.

c. David$_i$ obtained the contraband by himself$_i$.

As shown, there are two possible underlying structures for (79a). It could be the result of an extraposition in a sentence such as (79b), or it could be a manner adjunct with an unpronounced preposition, deriving from (79c). This potential for ambiguity proved to be problematic in the corpus study, as it made it difficult to classify bare sentence-final non-argument *self* pronouns. In coming to an analysis of these structures, an important first step is being able to tell them apart.

2.3.1 Disambiguating Sentence Final *self* Pronouns

To solve this problem, a test was devised, based in part upon the extraposition analysis proposed by Bickerton (1987) for the emphatic case. If the sentence-final emphatic is formed through extraposition, it should be impossible to pronounce the emphatic *self* pronoun at both the trace position and the extraposed position:

(80) * Vera$_i$ herself$_i$ was a member of the bridge club herself$_i$.

However, if the sentence final *self* pronoun were a manner adjunct, then this should be a possible construction under the assumption that manner adjuncts and emphatics have separate derivations, with the first *herself* emphasising the subject, and the second indicating the manner. So, if a sentence with a final *self* pronoun tolerates the insertion of an additional subject-adjacent emphatic, then the sentence final *self* pronoun should be a manner adjunct. If it does not tolerate the additional *self* pronoun, then the sentence-final *self* pronoun is a displaced emphatic. While such sentences are definitely marked from a pragmatic point of view, they are grammatical, and with the right context, they seem perfectly fine:

(81) Les$_i$ himself$_i$ fixes his cars by himself$_i$, though he advises his friends to use a mechanic.

Looking at the pair (80) and (81), another contrast emerges. In the ungrammatical (80), the predicate is non-agentive, whereas the example in (81) uses an agentive predicate. This suggests a connection between the manner adjuncts and agentivity of the associated predicate. This connection is noted by Gast and Siemund (2006) who claim that some (but not all) bare sentence-final non-argument *self* pronouns can have the same interpretation as when the *self* pronoun appears in a *by* phrase. They describe this as an "oblique reflexive expressing actor-oriented intensification" (Gast and Siemund 2006, p 363).

An experiment was devised to test both the hypotheses that sentences should be able to tolerate two of these non-argument *self* pronouns, each filling a separate role, and to test explicitly whether there is a connection between agency and the unambiguous *by* phrase.

The collection of grammaticality judgements for sentences such as (80) and (81) is not a simple task, as the sentences seem so unnatural to native speakers that it would be difficult to determine whether a negative response is a result of actual ungrammaticality, or just pragmatic oddness. For this reason, a more sensitive task is required.

This experiment is based upon a magnitude estimation (ME) task, in which participants are required to evaluate a series of new stimuli with respect to a fixed modulus. Originating in work on psychophysics used to test sensitivity to stimuli such as light or sound intensity (Stevens, 1975), this methodology has been adapted to grammaticality judgements. Participants are presented a modulus sentence, and asked to score its naturalness. Then, stimuli are presented, with the task being to score the stimuli relative to the modulus. Keller and Asudeh (2001) use this method to test binding judgements of native English speakers, showing that their findings are not significantly different from experiments using a more straightforward forced-choice task.

In this case, all experimental stimuli took the shape of sentences which ended with a non-argument *self* pronoun. In all, there were four different configurations of non-argument *self* pronouns, crossed with two predicate types, agentive versus non agentive. In total, this yielded eight stimulus types, shown in (82) and (83):

(82) Agentive Stimuli

a. Jim painted the house *himself*.

b. Jim painted the house *by himself*.

c. Jim *himself* painted the house *himself*.

d. Jim *himself* painted the house *by himself*.

(83) Non-Agentive Stimuli

a. Will was a subscriber *himself*.

b. Will was a subscriber *by himself*.

c. Will *himself* was a subscriber *himself*.

d. Will *himself* was a subscriber *by himself*.

As shown, the sentence final *self* pronoun appears both with and without the *by* preposition, and again both with and without the earlier incident of an emphatic *self* pronoun. For each stimulus type, there were three tokens, yielding a total of 24 experimental stimuli. An equal number of filler sentences were created, with a broader range of sentence types, and grammaticality. To distract from the *self* pronouns at the end of the experimental stimuli, all the fillers also ended with *self* pronouns, though some of these were argument reflexives.

(84) Sample Fillers

a. Robin saw a lizard beside herself.

b. Jean wanted to distinguish himself.

c. * Will was talking to herself.

d. * Who the fact that was a crook proved itself?

Ideally, in order to encourage the participants to give as wide a range of responses as possible, there should be a maximal amount of variety in the stimuli. Because it is difficult to make a sentence seem "more grammatical," there is not much room for this in the construction of grammatical filler sentences. Where this is possible is in the ungrammatical fillers. Here, the examples show an error based on a lack of agreement, and one based upon a subjacency violation. To balance these out, grammatical fillers with similar structure were also used.

Care must also be taken in the choice of a modulus sentence. In an experiment like this, the ideal modulus is not completely grammatical, but not so immediately perceptible as ungrammatical as some of the ungrammatical fillers. Again, finding this middle ground is important in order to encourage participants to make as wide a distinction as possible between grammatical and ungrammatical stimuli. For this experiment, it was decided that the best choice for a modulus would be a sentence containing a superiority error in which an adjunct *wh* phrase has moved over a subject *wh*:

(85) Why did who criticise himself?

Note that the modulus also ends with a *self* pronoun, though its ungrammaticality is not connected at all to that pronoun.

Participants begin the experiment by assigning a numeric score to the modulus sentence. Then, the stimulus sentences are assigned scores relative to the modulus score. In this case, participants are asked to judge the modulus based upon its naturalness as a sentence of English. They are then instructed to assign relative scores to the experimental stimuli; more natural stimuli receiving higher scores, and less natural ones receiving lower scores. Using the original modulus score, the raw scores from each participant can be converted into ratio scores and then normalised by taking the decadic logarithm (following Keller and Asudeh), so that the scores can be used for statistical comparison.

Whereas prior implementations of this experimental protocol have used pencil and paper approaches, the decision was made in this case to use an electronic method implemented using SFU WebSurvey, an internet-based survey software. This has the benefit of recruiting a larger number of participants in a short period of time without requiring them to visit a laboratory or classroom, and greatly facilitates the data analysis process. To orient participants to the task, a two minute instructional video was prepared, introducing the magnitude estimation methodology, first using line lengths as a simple illustration, then showing how it can be applied to sentences. Still images from the instructions, showing the comparison of lines to sentences are shown in Figure 2.1. On the left, participants are shown how numbers are used to rate new lines against the modulus at the top. On the right, sentences are presented, and participants are asked to perform the same rating task. To avoid any influence on the responses, no numerical values are ever presented beside sentences in the instructions. To get familiar with the web interface, participants were then required to evaluate the sample sentences from the video against a sample modulus before proceeding into the experiment itself. A sample of the experiment display is shown in Figure 2.2. Throughout the experiment, participants had access to the modulus sentence shown on screen for each trial. Each test sentence appeared on screen individually, with participants entering their numeric ratings in a text field. Test trials were presented in a randomised order which was kept constant for all participants.

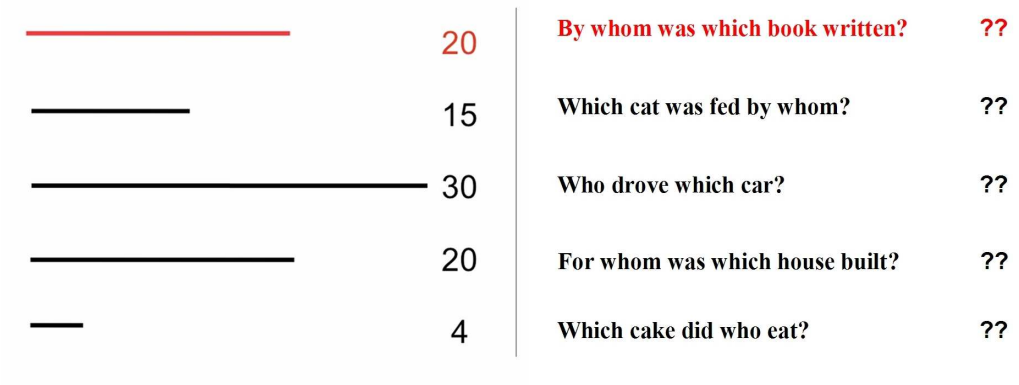


Figure 2.1: Screen shots from the magnitude estimation task instructions

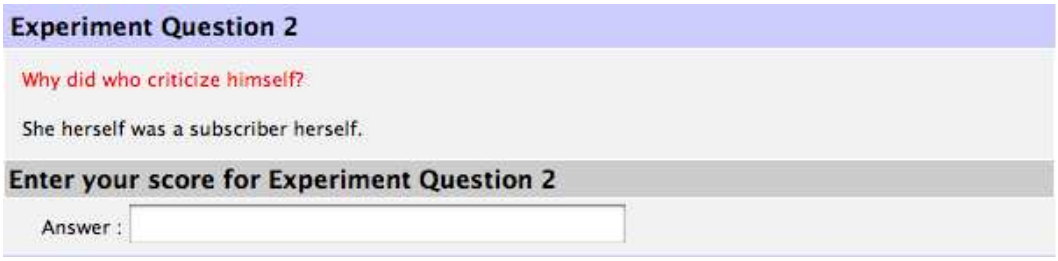


Figure 2.2: Participant view of magnitude estimation task experiment interface

Participants were recruited through social networking websites, as well as through undergraduate classes at Simon Fraser University. Participants were all required to be native speakers of English who had been born and raised in a primarily English-speaking environment. In all, 30 participants were recruited for this experiment, each being entered into a draw for a $100 gift certificate. A within-subjects experimental design was used, with all participants evaluating all types of experimental stimulus.

As a first measure of the effectiveness of the method and implementation, all participants' scores for the filler sentences were examined. Because some participants had given scores of zero to some of the ungrammatical stimuli, it was impossible to carry out a full statistical analysis on the filler data.[6] However, a simple inspection of the filler results shows that participants are able to make distinctions between grammatical and ungrammatical sentences using this method. Recalling that a score of 1.0 indicates that a sentence is judged to be as natural as the modulus (85), the average ratio score for the grammatical fillers was 1.71, while the average ratio score for the ungrammatical fillers was 0.99. The ungrammatical fillers even segment out into a distinct pattern of scoring, with subjacency violations and cases of a *self* pronoun lacking a c-commanding antecedent being scored markedly worse than superiority violations or agreement mismatches.

Average ratio scores for all eight conditions are shown in Figure 2.3. After all mathematical transformations, a 2-way ANOVA was conducted on the average scores for each experimental stimulus. A significant main effect for predicate type was found ($F_{(1,29)}$=287.39, $p < 0.001$), indicating that the presence of agentive versus non-agentive predicates had

[6]This problem was identified early in the data collection, and the instructions were subsequently modified, but due to uneven data sets, no comparative analysis of the filler versus test items was conducted.

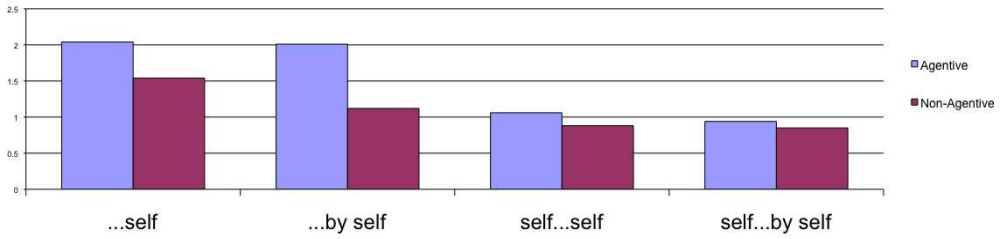


Figure 2.3: Mean Ratio Scores for each Condition in Magnitude Estimation Task Experiment

an effect on the overall scores. Specifically, the non-agentive predicates were dispreferred in all conditions. Similarly, a main effect of *self* pronoun configuration was found ($F_{(3,29)}$=235.67, p <0.001), indicating that the different patterns of *self* pronouns found in the paradigms in (82) and (83) had a significant effect on the scores. Finally, a significant interaction between these two factors was found ($F_{(3,29)}$=16.547, p <0.001), indicating that the choice of predicate type interacts with the configurations of *self* pronouns.

Looking closer at the scores between the different stimulus types, it was found that there was a significant difference between the scores for the agentive versus non-agentive sentences ending with *by self* ($t_{(29)}$=3.98, p=0.004). Specifically, the non-agentive sentences were significantly less acceptable (an average ratio score of just 1.12) than the agentive sentences with *by self*. Though the margin was not as wide, a significant difference was also found for the *self...self* sentences ($t_{(29)}$=2.76, p=0.01), again with the agentive sentences being more acceptable than the non-agentive ones.

The results of the ANOVA are not too surprising, providing experimental proof for the generalisations drawn from the previous corpus work, establishing that there is indeed an interaction between the agentivity of a predicate and the configurations of non-argument *self* pronouns it can support.

More conclusive are the individual comparisons between conditions. The first result enhances the findings of the ANOVA analysis, making it clear that the *by* phrase manner adjunct is significantly less acceptable with a non-agentive predicate. This not only provides a more solid empirical basis for the use of predicate type as a factor in deciding ambiguous cases in the corpus work, but it also provides independent evidence for propos-

ing a connection between this type of adjunct and the argument structure of the predicate. Recalling the initial discussion of making a choice between competing binding theories, this would be an indication that the best method may be to take a hybrid approach, as it would seem that this particular non-argument *self* pronoun requires a specific argument or θ-position to be present in order to be licensed.

The second significant comparison, showing that there is a significant difference between sentences with the double *self* pronoun again based on predicate type is also important. First of all, by establishing that in a controlled experiment, native speakers can distinguish between grammatical and non-grammatical uses of sentences with two bare non-argument *self* pronouns, the test used to diagnose ambiguous cases in the previous corpus work has experimental validation. Having established that the non-agentive sentences could not support a manner adjunct, the *self...self* cases with a non-agentive predicate could only be produced in a derivation wherein an extraposed emphatic was pronounced at both the trace position and the landing site, a derivation which should be blocked. The only alternative would be to assume that the two instances of the emphatic had different origins. Furthermore, this experiment indicates that even without the *by* preposition, native speakers are sensitive to the argument structure distinction at work.

Beyond the statistical analysis though, there is one more general trend in the data worthy of note. For all the sentences with the doubled *self* pronouns, their acceptance level was quite low, with only the agentive *self...self* cases crossing the line of being rated on average more acceptable than the modulus. This could in part be due to the fact that the sentences were presented in a written form, with no context. Recalling the original example (81), the addition of a context makes the judgement clearer. Given the result that the *by* phrase manner adjunct on its own was poorly rated with the non-agentive predicates, a simpler classification scheme would be to follow the lines of Gast and Siemund, assuming that for sentences with an ambiguous *self* pronoun in the sentence final position, the predicate could resolve that ambiguity: if the predicate is non-agentive, a sentence-final bare *self* pronoun is most likely to be an extraposed emphatic, with clear manner adjuncts being widely rejected. If the predicate is agentive, most likely the sentence final *self* pronoun would be a manner adjunct, as there was no significant difference in the acceptance judgements of the sentence final *self* pronouns with or without the *by* when modifying an agentive predicate; they were rated virtually identically in the experimental results.

Recalling the initial questions, this experiment has not conclusively demonstrated that sentences with the doubled *self* pronoun are possible, owing to the relatively low acceptance levels of all four stimulus types containing the doubled *self* pronouns. This leaves open the possibility that there may be some room to challenge the notion that the emphatics and manner adjuncts are distinct constructions, in that their co-occurrence is marginal at best. Still, even within these marginal cases, participants were able to make a significant distinction based on predicate type, showing that there is indeed some interaction between these non-argument *self* pronouns and the verb. If it were the presence of the two *self* pronouns making the sentence ungrammatical, then such a distinction should not occur: all uses should be equally bad regardless of the predicate. In the case of the manner adjunct, there is a clear path to follow, in that it appears there is going to be some connection with agency. It is to this that I next proceed.

2.3.2 External Argument Adjuncts

Recalling the previous discussion, the manner adjunct *self* pronouns were considered to be unusual owing to their inability to be replaced by a referential pronoun. Under the terms of Reinhart and Reuland, this should mean that they are not exempt anaphors. However, there is nothing in Principles A and B which predicts that this adjunct must be a *self* pronoun:

(86) Roger$_i$ baked the cake by *him$_i$/himself$_i$.

The predicate here is clearly not reflexive, and thus Principle B would not require reflexive marking. Furthermore, it is not clear at this point that the manner adjunct could provide such reflexive marking, despite the experimental evidence that this position is connected with agency. The Chain Condition might be seen as a possible solution, but recall that it makes reference to A-chains, and could only be invoked if there was reason to believe the external argument adjunct (EAA) was an A-position. The situation is little better for the standard binding theory, as the question becomes one of determining why some adjuncts are exempt from binding theory, but not this one. Another more recent investigation of *self* pronouns in non-argument positions (Xue and Popowich, 2002) describes a binding condition which states that *self* pronouns not bound by a co-argument (or, in this case, lacking a co-argumet) must be bound by the minimal c-commanding subject. This accurately describes the licensing of the *self* pronoun in (86), but still does not account for the fact the

reflexive is obligatory. For both Reinhart and Reuland and Chomsky then, the best hope of explaining the data would be to establish the argumenthood of this apparent adjunct.

In looking at the adjunct PP in (86), this adjunct bears a striking similarity to another common adjunct PP in English, shown in (87):

(87) The cake was baked by Roger.

This is a classic passive *by* phrase, where the complement of the preposition is a DP referring to the agent in the described event of cake-baking. Similarly, the DP complement in the adjunct from (86) refers to the agent, though redundantly. In the corpus, the majority of cases where the manner adjunct occurred were with agentive predicates, both transitive and intransitive. There were no instances of possessive *have*, or psych predicates such as *believe*. Overall, the distribution suggests that there is a correlation between the use of this adjunct and agentive predicates, a correlation which was verified in the ME experiment.

The investigation of the passive *by* phrase begins again with a look at the corpus. Here, the same one million words of spoken English from Switchboard were examined. There was no analysis of the written corpus, because it contained substantially fewer instances of the manner adjunct; with only 21 written instances, versus 96 spoken, it was decided that the spoken corpus would be the best place to look for correspondences between manner adjuncts and the passive *by* phrase. In all, 142 instances of a passive *by* phrase were found in the Switchboard corpus, occurring, not surprisingly, with ordinary transitive predicates. Similar to the results for the manner adjuncts, there were no cases of the passive *by* phrase on possessive *have* or verbs of mental states.

One distinction was that the passive *by* phrase occurred on more predicates where the internal object was an experiencer, rather than a theme:

(88) a. I was absolutely enthralled by it. (SW2262.DFF: B.34)

b. We have been really frustrated by our gardening attempts here. (SW2935.DFF: B.8)

While there were no cases in the corpus of these verbs occurring with a reflexive *by* phrase in the active, such sentences are not impossible:

(89) a. You$_i$ enthralled me by *you$_i$/yourself$_i$.

b. John$_i$ frustrated me by *him$_i$/himself$_i$.

These sentences, which sound marginal to some speakers, could certainly be used in the right context. For example, (89a) could be directed to an actor in a production with a large ensemble cast, who has been modestly declaring that any enthrallment on the part of the audience is due to the performance of the cast as a whole. A particularly devoted fan of this one actor could utter (89a), only with the reflexive, meaning that the one actor was enthralling without the support of the rest of the cast. A similar context could be constructed for (89b), again with the intended meaning only coming through with the reflexive, not a pronoun.

Using these observed similarities as a starting point, Storoshenko (2009) argues for the unification of active and passive *by* phrases under the term External Argument Adjunct (EAA). The EAA is defined as an adjunct to *v*P, containing a referent which obligatorily refers to the external argument of a predicate. In the active case, that external argument is present, and in the passive, it is not. Here, I will provide an abbreviated version of the argumentation used to motivate the EAA.

In further establishing this similarity between the active and passive *by* phrases, the binding facts should also be examined. If the passive *by* phrase is somehow an argument, only *self* pronouns should be used where there is a co-referential subject. If the *by* phrase complement is not an argument (or even argument-like), then a free variation should be expected, along the lines of an exempt anaphor. As shown below in (90), neither of these predictions holds true:

(90) a. * Bob$_i$ was hit by himself$_i$.

b. * Bob$_i$ was hit by him$_i$.

c. ? Bob$_i$ was hit by Bob$_i$.

The judgements here are clear: neither a *self* nor a referential pronoun may appear in a passive *by* phrase where the sentential subject binds the complement of the PP. Of all three sentences in (90), the best is actually (90c), which is an apparent Condition C violation. However, the acceptability of this case is most likely a result of interaction with focus phenomena, negating an earlier claim that some other person may have hit Bob.

The binding facts for the EAA can be captured under the standard binding theory, with the following considerations. Firstly, adopting the analysis of Canac-Marquis (2005), bind-

ing domains can be re-defined in terms of derivational phases: phase boundaries demarcate binding domains. Secondly, following on recent work by Chomsky (2005), *v*P in an active clause is a phase boundary, whereas *v*P in a passive clause is not. To keep in line with Reinhart and Reuland, it would be necessary to extend this obligatory co-reference of the complement to full argument status. This is somewhat tenable under a Jaeggli (1986) style analysis of the passive where the external theta role is assigned to the EAA position (recalling that θ-role assignment is a part of the definition of a syntactic predicate), but it will be harder to maintain that the EAA in the active is also an argument of the predicate, there being no additional theta role.

Looking first at cases with the EAA in an active clause, an external argument is introduced at [Spec, *v*P], though this is later moved (or copied) to [Spec,TP]. Still, the trace (copy) at [Spec, *v*P] is the key element here. Because active *v*P is a phase boundary, the *v*P is a binding domain. The EAA, which is by definition co-referential with the agent at [Spec, *v*P], is c-commanded by the trace of the external argument within its binding domain. Thus, the only form of DP which will be acceptable here under Condition A is a reflexive, as Conditions B and C would rule out pronouns and co-referential R- Expressions:

(91) a. Roger$_i$ [$_{vP}$ t$_i$ baked the cake by himself$_i$].

b. * Roger$_i$ [$_{vP}$ t$_i$ baked the cake by him$_i$].

c. * Roger$_i$ [$_{vP}$ t$_i$ baked the cake by Roger$_i$].

In this case, the EAA does not introduce a new referent; rather it conveys the additional meaning that the action was carried out by the agent alone. This can be seen in the observation that the EAA does not allow for any instances of partial co-reference:

(92) a. * Sophie built the house by themselves.

b. * We built the house by myself.

By comparison, locative *by* phrases allow for number mismatches:

(93) a. I shovelled the sidewalk by us.

b. We saw the snake by me.

However, these number mismatches are degraded with *self* pronouns, particularly when there is a singular subject and a plural *self* pronoun; the converse case with a plural subject and singular reflexive is not problematic at all:

(94) a. ? Daniel saw the snake by themselves.

b. Amy and I saw the snake by myself.

While the exact constraints on the locative by phrase are not clearly established in these examples, it is evident from the data in (92)-(94) that locative *by* phrases are not subject to the same strict one-to-one relation between their complement and the external argument as the EAA.

Evidence that the active EAA is distinct from the locative *by* phrase can be found in *wh*-extraction diagnostics. Because the EAA is obligatorily co-referential with the external argument, it is obligatorily bound in active cases, where the external argument c-commands the EAA position. As such, *wh*-extraction from this position should be impossible, as the resulting *wh*-question would result in a crossover violation. This is indeed borne out:

(95) a. * Whom did Roger bake the cake by?

b. * By whom did Roger bake the cake?

The sentences in (95) could be acceptable as locative readings, asking in whose vicinity Roger baked the cake, but they cannot be interpreted as asking a question answerable by (86), stating that Roger baked the cake without outside assistance. If anything, a question formed from the manner adjunct would most likely be a *how* question, with no remnant of the PP or its complement. This suggests that the meaning of this phrase as a whole does not in itself carry any independent reference to the external argument, merely that the form used to express that meaning contains this bound form. Conversely, a question based on a locative is somewhat better:

(96) a. ? Whom did you see a snake by?

b. ? By whom did you see a snake?

Setting aside stylistic questions in the choice of the accusative pronoun, and whether or not to pied-pipe the preposition, these are more acceptable than the examples in (95), indicating that *wh*-extraction from the EAA (95) is more clearly ungrammatical than the adjunct island violations of (96). Again, though there is more work to be done in refining the exact distinction, it is enough to note the existence of this distinction to make the claim that the active EAA is distinct from a locative *by* phrase. In these locative cases, the fact that the

complement of *by* can be replaced by the *wh* pronoun suggests that the complement here *does* refer to an individual, whereas doing so in (95) loses the manner meaning.

Turning to the passive, there is no external argument at [Spec, vP], making the EAA's referential content new information. According to Chomsky (2005), passive vP is not a phase, therefore not a binding domain: the whole clause becomes the binding domain. Because of this extension of the binding domain, the promoted internal argument at [Spec, TP] c-commands the EAA within its binding domain; Condition A would once again require a *self* pronoun. However, as was seen in (90), repeated below as (97), this is not the case:

(97) a. * Bob$_i$ was hit by himself$_i$.

b. * Bob$_i$ was hit by him$_i$.

c. ? Bob$_i$ was hit by Bob$_i$.

As was already observed, a *self* pronoun may not appear in a passive EAA. This is borne out by the examples of passive EAA from the spoken corpus, where there was only one instance of a passive EAA containing a *self* pronoun, where the EAA appears in a passive follow-up question to a prior active question where the agent *they* of the first question is the antecedent for the *themselves* in the follow-up.

(98) Do they drive in? Are they driven by themselves? (SW2866.DFF: A.89)

This example may be regarded as an exceptional case, as an arguably more natural alternative to (98) would be (99):

(99) Do they drive in? Do they drive themselves?

Here, both questions are active, and the same meaning is retained. Further investigation would be needed to determine whether the kind of structure in (98) is really productive, or represents a "one off" speech error.

In this, a potential solution to the problem of (97) is presented. Meaningwise, there is no difference between the sentences in (100), though the active is grammatical while the passive is not:

(100) a. Bob$_i$ hit himself$_i$.

b. * Bob$_i$ was hit by himself$_i$.

(100b) is not a binding theory violation, so there must be something else constraining this structure. According to the Blocking Principle of Williams (1997), distinct linguistic forms with identical meanings are not permitted; one form, the more articulated one, will supersede the other. In this case, the active sentence with both arguments present the *self* pronoun in an argument position blocks the passive with the *self* pronoun in the EAA. This accounts for the facts in (97). While Condition A is satisfied, the availability of an equivalent active form blocks (97a) and (97b) is handled by Condition B; the additional focus interpretation preserves (97c).

In considering the passive EAA and *wh*-extraction, the role of the implicit external argument becomes crucial. First of all, the Blocking Principle does not seem to apply, as questioning the EAA complement in the passive does not sound as ungrammatical as in the active, even despite the existence of a synonymous active question:

(101) a. Who baked this cake?

b. ? By whom was this cake baked?

More surprising about (101b) is that it appears to be at worst marginal, if not completely grammatical. If the examples in (95) are so strongly ungrammatical due to a crossover violation, then (101b) would indicate that this crossover violation is not present when extracting from the passive EAA. Recall that in the passive EAA, the *by* phrase complement is co-referential with the implicit external argument; while this argument is present in the discourse, the apparent lack of a crossover violation indicates that no implicit argument is present in the syntax, covertly binding the EAA position.[7]

While this analysis can account for the distribution of *self* pronouns in the EAA position, the precise character of this adjunct is not so simple to specify. In the passive, the EAA contributes a new referent, while for the active sentences, the reference itself is redundant, and the EAA contributes a sense of the action taking place without any outside assistance. The question is whether there is any difference in felicity in the following discourses:

[7]What remains unclear is the reason behind the contrast between (100) and (101). Because the observed ungrammaticality in (100) is attributed to a blocking effect, one possibility is that the questions in (101), which would be seen as denoting the sets of possible answers, are not synonymous in the same way as the declaratives of (100).

(102) a. # Martha$_i$ found the solution by herself$_i$. It was found by Martha and The Doctor together.

b. # The solution was found by Martha. Martha and The Doctor found it together.

For both the active and the passive in (102), there is a sense that continuing the discourse with a statement that Martha did not act alone is pragmatically odd. That inserting *actually* at the beginning of the second sentence for both examples in (102) improves the dialogue reinforces the sense that the two sentences are contradictory. The passive version could be paraphrased by (103):

(103) The solution was found by Martha and only Martha.

This paraphrase suggests a possible avenue for analysis of the EAA, as *only* is a focus sensitive particle. The EAA could invoke a set of alternative agents, and exclude them through a covert exclusivity operator. The mechanics of this analysis will be presented later in the chapter; what still remains is to discuss the cases where the EAA appears without its preposition. So far, I have assumed that there is no discernible change in meaning, and that this is a deletion process by which the *by* is dropped from the EAA. However, an alternative theory is proposed in the literature on emphatics. Gast and Siemund, in discussing such adjunct prepositional phrases across languages, describe them as standing for the agent. English *by* phrase examples are given alongside this discussion, though from their perspective, the version with the *by* phrase is not related to the version without, and no formal analysis of the *by* phrase is presented. The analysis proposed above gives a treatment of the *by* phrase, showing how in both active and passive sentences, the *by* phrase complement can be treated as obligatorily co-referential to the agent, in effect serving as an argument position. That the *self* pronoun is obligatory in the active can be explained in terms of either Condition A, or under the Chain Condition.

Before turning to the semantics of the English *self* pronouns, I summarise the literature on the emphatics, showing how they can be brought in line with this treatment of the EAA, and addressing the issue of whether or not there is a meaning change when the *by* is absent.

2.3.3 Emphatic *Self* Pronouns

Unlike the manner adjuncts, there is no paucity of analyses for the emphatic *self* pronouns, though there is little consensus to be found. Some of these analyses have already been mentioned in passing, so I will begin by tying together and comparing the various analyses.

Going back first to the analysis presented in Déchaine and Wiltschko (2002b), there it is claimed that the emphatic *self* pronoun is similar to an adjective predicated of its antecedent. This was shown with the examples repeated below:

(104) a. I myself saw Mary.

b. I saw Mary myself. (Déchaine and Wiltschko 2002, ex 14a-b)

(105) a. John, tired, wrote the letter.

b. John wrote the letter tired. (Déchaine and Wiltschko 2002, ex 15a-b)

Based on the positional similarity between *myself* and *tired*, they describe the *self* pronoun as an identificational predicate, along the same lines that *tired* is a depictive predicate. To support this claim, they cite an example from Baker (1995) showing that it is not possible to have two different emphatics in the same sentence:

(106) * Fred himself is not usually as alert as Karen herself. (Baker 1995, ex 42d)

This is explained by an independent constraint on identificational predicates which are argued to impose a uniqueness restriction. That is, only one element in a sentence should carry this predicate. The ungrammaticality of (106) then follows from this constraint, indicating that *himself* and *herself* are identificational predicates.

This uniqueness constraint does not limit the emphatic to subject positions. There is independent evidence that the emphatic can appear on non-subjects, but extraposition to the sentence-final position is apparently blocked from non-subject DPs:

(107) a. John gave Mary herself the book.

b. * John gave Mary the book herself.

(107b) has the appearance of a clash of ϕ features; the immediate response is to correct (107b) by having the emphatic agree with the subject *John*. Assuming that the emphatic is subject to Condition A, then this could indicate that the extraposed position is somewhere

outside *Mary*'s c-command domain. Extraposition from subjects is fine though, meaning that the extraposed position is below [Spec, TP]. Despite Bickerton's (1987) description of the emphatic as a complement to its antecedent, it remains hard to conceive of *herself* as an argument of *Mary* in any way. As such, the Reinhart and Reuland Principles have nothing to say on this issue, and the chain condition does not apply either, being limited to A-chains.

Similar to the analysis of Déchaine and Wiltschko which described the emphatic as identificational, Bickerton likewise describes these as having the meaning "NP_i and no one *but* NP_i" (Bickerton 1987, p345). A more formalised analysis comes in Baker (1995), using the term 'intensive NP.' Based on his example (106) above, he argues that one of the functions of the emphatic is to mark the most discourse-prominent referent in a sentence. This is given in one of two conditions he establishes:

(108) CONDITION OF RELATIVE DISCOURSE PROMINENCE (Baker 1995, ex 24)
Intensive NPs can only be used to mark a character in a sentence or discourse who is relatively more prominent or central than other characters.

(106) thus violates this constraint by virtue of the fact that there can only be one most-prominent referent. His other constraint runs along similar lines to the Bickerton quote:

(109) CONTRASTIVENESS CONDITION (Baker 1995, ex 19)
Intensive NPs are appropriate only in contexts in which emphasis or contrast is desired.

Contrast is implied in Bickerton's "and no one *but*," making Baker's analysis consistent with that of Bickerton and the later Déchaine and Wiltschko account of identification.

Returning to the issue of non-subject positions, Baker claims that for non-nominative pronouns, such as accusative pronouns in object positions, the prominent and contrastive pronoun is actually a bare *self* pronoun, which he argues accounts for a number of locally free *self* pronouns in argument positions:

(110) But Marianne, who saw his agitation, and could easily trace it to whatever cause best pleased herself, was perfectly satisfied, and soon talked of something else. (Baker 1995, ex 14a)

Baker's claim is that the *herself* in (110) has the same discourse function as the ungrammatical *her herself* would. König and Siemund (2000a) use diachronic evidence to show that English *self* pronouns grew out of a combination of object pronoun + intensifier, so it is not unexpected that in object positions the older form is retained without doubling the pronoun.

There is a sense throughout these analyses that the emphatic has a similar function to the EAA, indicating that only the antecedent of the emphatic is relevant. There does not seem to be a restriction to agency for the emphatic though, as in (107a), the natural reading is that Mary is the only recipient of the book, falling in line with this exclusive interpretation.

However, this is not always the case. Looking back to (104), the reading is in fact one of *inclusion*; the sentence can be paraphrased as *Even I saw Mary*. This inclusive reading also emerges in some corpus examples:

(111) a. Humana, which wants to acquire one of the new machines itself, is on the record as opposed to the proposal. (wsj_0416.mrg:14)

b. Uh, believe me, I do that myself. (SW2062.DFF: A.31)

In both these cases, the prior text establishes some other entity which shares the same property of either wanting one of the new machines, or making contributions to a retirement savings program.

This alternation is observed in König and Siemund (2000a), who give the following paradigm:

(112) a. The director himself will talk to us. (adnominal)

b. I am a little short of cash myself. (adverbial, inclusive)

c. Mary earned that money herself. (adverbial, exclusive)

In discussion of the first case, (112a), König and Siemund say that here, the director is being juxtaposed against other possible speakers, and make similar claims for parallel examples. With regard to this construction they make the following conclusion: "Relating a given value to a set of salient alternative values is a characteristic property of focusing devices, to which intensifiers therefore clearly belong" (König and Siemund, 2000a, p. 42). A more formal definition appears in König and Gast (2006):

(113) DEFINITION OF ADNOMINAL INTENSIFIERS (König and Gast, 2006, ex (4))
Adnominal intensifiers are expressions that are used to relate the referent x of a given (co-)constituent to a set of alternative referents $Y = \{y_i, y_2 ... y_n\}$, such that each element $y_i \in Y$ can be identified relative to x.

In this definition though, there is no statement of whether the relation is one of exclusion or inclusion, merely that an alternative set of referents is invoked. This is not a problem though, as at first glance, it appears that the extraposition analysis can still be maintained. Both (112b) and (112c), where the emphatics were sentence-final, can be recast and have the same meanings:

(114) a. I myself am a little short of cash.

b. Mary herself earned that money.

In (114a), the reading still comes through that the speaker is making some statement of shared poverty (either way it is constructed, this sounds like a natural response to being asked for a loan), and (114b) still has the meaning that it was Mary alone who earned the money. Thus, the emphatic which remains adjacent to its antecedent can be either inclusive or exclusive, just as the extraposed one can have either function. This is an important observation, as it is largely ignored in the literature, which regards the emphatic that remains adjacent to its antecedent as exclusively exclusive. Even a cursory internet search can quickly yield examples which provide enough context to make it clear that the meaning is inclusive rather than exclusive in some cases:

(115) a. ...this will not be covered under third party insurance- it is literarily there to protect other road users when you yourself are on the road.
(www.insurancesearch.co.uk/car-insurance/, accessed 21-Mar-2010.)

b. I've read through that handbook for the recently deceased. It says: 'live people ignore the strange and unusual.' I myself am strange and unusual.
(Lydia Deitz - *Beetlejuice*, 1988)

In both examples (115), the referent modified with the emphatic shares the described property with others in the previously established context. In the first case, *other road users* explicitly spells this out, while in the second case, a generic reference is made to *the strange and unusual*, a set to which the speaker is adding herself. Clearly, this usage exists.

Returning to König and Gast, they note that there is a connection between the predicate and this inclusive/exclusive alternation. They claim that the exclusive use is not permitted in stative contexts, the contexts where inclusive emphatics are primarily used. However, the versatility of the inclusives was shown in the (111) corpus examples, where the predicates are *want* and *do*. This also relates back to the results of the magnitude estimation task, in that the manner adjunct appears to have a similar function as the exclusive emphatic. Recall that the *by...self* phrases, which have the same exclusivity reading, were significantly more acceptable with non-stative predicates.

Repeating the examples from (112), adding in the observed exclusive adjacent emphatic yields the paradigm in (116):

(116) a. I myself am a little short of cash. (adjacent, inclusive)

b. The director himself will talk to us. (adjacent, exclusive)

c. I am a little short of cash myself. (extraposed, inclusive)

d. Mary earned that money herself. (extraposed, exclusive)

There is no positional distinction in the inclusive and exclusive functions, as both readings are possible in adjacent or extraposed cases. This means that an analysis which derives the extraposed case from the adjacent case is still tenable. In turning to predicates, it may be possible to retain a connection between agentivity and the exclusive reading, but the inclusive reading appears to be possible with any predicate, given the right context.

The observations are even more complex though, and Gast and Siemund (2006) ultimately argue that there are distinctions to be drawn between the adjacent and extraposed cases. This distinction is based upon three diagnostics, looking at the exclusive emphatics in conjunction with indefinite antecedents, inanimate antecedents, and sentential negation. They first note that modifying an indefinite with the adjacent emphatic, as in *a president himself* can sound unnatural, whereas an extraposed case would be fine. However, they then argue that the adjacent cases are also acceptable in a context where the existence of a set of presidents has already been established and juxtaposed with another set of people (like vice-presidents).

They also claim that the exclusive adjacent cannot be used with an inanimate DP. This can be easily challenged though:

(117) I saw the original itself.

(117) is felicitous in a context where the speaker is denying having seen one of a known set of copies of an artwork. Rather, the speaker has in fact seen the original.

The third diagnostic comes in the context of negation:

(118) a. Max himself did not mow his lawn, but his brother Bill did.

b. Max did not mow his lawn himself, but his brother Bill did.
(Gast and Siemund 2006, ex 17-18)

Their discussion of the sentence pair in (118) focuses on the sloppy reading of *his*, where there are two lawns in consideration. (118a) directly contrasts the brothers, stating that Max's lawn was not mowed, but Bill's was. In (118b), both lawns are mowed, and the contrast is in the fact that Bill mowed his own lawn, while Max, one way or another, had someone else do it. Based on these data, Gast and Siemund argue that there is no connection between the adjacent and the extraposed emphatics; at least those with the exclusive reading. However, stepping back, the first of their two diagnostics are inconclusive; they themselves argue that using an adjacent exclusive emphatic with an indefinite antecedent is fine in the right circumstances, and it seems that the same can be said of inanimate antecedents. The negation data does show that there is a scopal distinction to be made, in that the extraposed emphatic, anteceded by the subject, does appear to be under the scope of sentential negation.

Recall now the case of (107), reproduced below as (119):

(119) a. John gave Mary herself the book.

b. * John gave Mary the book herself.

Earlier, I claimed that this example could be used to deduce (roughly) the position of the extraposed emphatic, arguing that the extraposition of an object emphatic would be ruled out because it was moving outside the c-command domain of its antecedent. In light of the claim by Gast and Siemund, (119b) is ruled out because it should agree with *John*, being in the position of an actor-oriented emphatic. (119a) remains exclusive, having a similar reading to those cases where the emphatic is directly attached to a subject. From this, we can determine that while the exclusive reading of the adjacent emphatic is still associated

with non-stative contexts, it is not explicitly connected to agency in the same way as the actor-oriented emphatic at the end of the sentence.

Finally, it should be noted that nowhere in the cited literature is any connection drawn between the sentence final actor-oriented emphatic (what I have been calling the exclusive extraposed) and a *by* phrase, except to note that such oblique constructions are common for languages which do not have a form which can stand on its own. Certainly there is no claim that one derives from the other. Given the strong evidence linking the *by-self* phrases to the passive *by* phrase, I continue to consider them as separate, though they will have similar semantic forms.

A question not addressed by Gast and Siemund is to run the same three diagnostics used on the exclusives on the inclusive reading for the emphatic. Like the exclusives, inclusives are awkward when adjacent to an indefinite, but an extraposed usage is fine.

The second diagnostic was that of animacy:

(120) a. Jack himself was under the table.

b. The cup itself was under the table.

For both examples in (120), the inclusive reading is fine, given the right context. For (120a), Jack could be hiding under a table with other associated people already established to be doing so. In the case of the cup, it would simply be a matter of first noting some other associated items (teapot, saucer, milk jug, etc,...) to be under the table, and (120b) follows naturally.

Finally, the interpretation of the inclusive emphatic does not seem to vary when negation is brought into play:

(121) a. I myself am not a fan of The Doctor.

b. I am not a fan of The Doctor myself.

Again, both sentences in (121) can be read as agreeing with a prior statement, this time about not liking The Doctor. What is key is that there is no sense of a scope distinction as seen in (118). Putting all the observed facts together yields Table 2.3.

The first major observation to come out of this table is that there is no difference in the usage of either the exclusive or the inclusive adjacent. Both have the same distributional qualities. Where differences emerge is in the extraposed cases. Specifically, the exclusive

Table 2.3: Emphatic *Self* Pronoun Distinctions

	EXCLUSIVE		INCLUSIVE	
	Adjacent	Extraposed	Adjacent	Extraposed
Indefinite Antecedent	?	✓	?	✓
Inanimate Antecedent	✓	✓	✓	✓
Scopes over negation	✓	✗	✓	✓
Agent Oriented	✗	✓	✗	✗

extraposed appears to be under negation, and is actor-oriented, whereas the inclusive extraposed is above sentential negation, and does not have this same actor-orientation, coming as it often does in stative environments.

Circling back to Bickerton, it does then seem that *some* of the emphatics can be analysed as moving from their antecedent-adjacent position to the end of the sentence. However, they are the ones having precisely the opposite meaning as what he describes; only the inclusive emphatics can be given this movement analysis. There are no selectional differences based upon the position of the inclusive emphatic, unlike the exclusive, and the lack of effect of negation can be explained by having the inclusive extraposed emphatic interpreted at its base position. The difference in meaning for the exclusive cases in (118) can be captured if one is base-generated under the scope of negation while the other is not. If there was a derivational relationship between the two exclusives, one would expect the restriction on actor-oriented antecedents to be common to both.

To sum up, it appears that the emphatics can be broken down into four different categories, and none of them are related to the EAA. Having described the distribution of close to 94% of the instances of *self* pronouns in the corpus, I will now turn to the semantic account.

2.4 Semantics of English *self* Pronouns

Semantically, the *self* pronouns have two basic functions: one is the reflexive usage, expressing an identity between arguments, and the other is a focus-sensitive interpretation

expressing either an inclusive or exclusive relationship between the referent of the *self* pronoun and a group of contextually-defined alternatives. Having approached the *self* pronouns from a primarily syntactic direction so far, I now advance an analysis which works from the semantics, using different semantic forms to constrain the positions in which those pronouns can appear. As discussed in Chapter One, co-argument *self* pronouns in English can be seen as operators on a predicate, adding assigned co-reference between arguments. Here, I develop that analysis to cover not only local cases, but also cases where the *self* pronoun appears to operate across clause boundaries. While this will require a multiplicity in the semantic definitions of the *self* pronoun, the goal of this section will be to present an analysis which keeps this multiplicity as minimal as possible. Also, all of these uses are unified in that they carry the assigned co-reference identity. This will distinguish the argument cases from the emphatics, which will make use of an identity function in concert with focus.

2.4.1 Relating Arguments

Recalling the classification of Déchaine and Wiltschko, *self* pronouns have the status of full R-Expressions (contra Reinhart and Reuland). However, their proposed semantic form for reflexivity is that of assigned co-reference:

(122) $\mathrm{R}[x, y], x = y$

In this formulation, the relationship between the *self* pronoun and its antecedent is expressed as a specification on the *predicate*, rather than anything specific about one of the arguments. Because this identity relation is not inherent to the predicate, it must be contributed by the *self* pronoun. Thus, while the *self* pronoun may have the status of an R-expression, in line with its status as a DP, it also appears that a *self* pronoun modifies the predicate of which it is an argument. The referential character of the *self* pronoun is preserved in that a referential pronoun is built into the semantic form of the *self* pronoun, one half of the identity relation added to the predicate. The implementation of such a semantic approach to English reflexivity, which manipulates the predicate rather than making specifications about the syntactic positions of the arguments themselves, appears to be more in line with a Reinhart and Reuland type approach. First, I explore simple mono-clausal implementations of this analysis before moving on to more complex cases.

A semantic form for a basic reflexive is shown in (123):

(123) $[\![\text{REFL}]\!] = \lambda P \in D_{<e,<e,t>>} \lambda y.P(x)(y) \wedge x = y$

The *self* pronoun is in itself a function from $<e,<e,t>>$ to $<e,t>$, re-writing a two-place predicate into a one place predicate with the new identity relation added. The x variable in (123) would be replaced by the appropriate referential content A sample derivation is presented in (124):

(124) $\lambda P \in D_{<e,<e,t>>} \lambda y.P(x)(y) \wedge x = y \;\; (\lambda y \lambda z.\text{likes}(z,y))$
$\lambda y.\text{likes}(y,x) \wedge x = y$

As shown, the resulting function after combination of the transitive verb with the *self* pronoun is a one place function with the identity added to the predicate. The input to the function labelled REFL automatically precludes any possibility that the *self* pronoun will wind up in a subject position, as subject positions are standardly complements to one place rather than two place predicates. Schematised as in (123) and (124), the *self* pronoun looks like a detransitiviser, but crucially, there is referential content as well.

This formulation removes the need for any specific mechanics of agreement between the *self* pronoun and its antecedent:

(125) * John likes yourself.

The sentence in (125) would be derived from a specific form of the *self* pronoun, already carrying the second person pronoun:

(126) $[\![\text{yourself}]\!] = \lambda P \in D_{<e,<e,t>>} \lambda y.P(\text{yourself})(y) \wedge \text{yourself} = y$

Looking at (126), there would be no way for the two arguments of (125) to felicitously combine, as they would never satisfy the equivalence stipulation, under the pragmatic assumption that speakers not refer to their interlocutors in the third person, or make other such errors.

However, this formulation is not without its problems. A first one is that the analysis does not account for all reflexive *self* pronouns in ditransitive predicates:

(127) John$_i$ introduced himself$_i$ to Mary$_j$.

Simplifying, (127) can be represented semantically as in (128):

(128) $\lambda x \lambda y \lambda z$.introduce(z,y,x)(Mary)(himself)(John)

Here, after Mary has undergone lambda conversion, replacing the *x* variable, the result is a transitive predicate, and the definition of the reflexive formulated above would apply without problem, merely changing the referential content of (126) to *himself*. However, problems arise in other variations on (128):

(129) a. John$_i$ introduced Mary$_j$ to himself$_i$.

b. John$_i$ introduced Mary$_j$ to herself$_j$.

In both sentences of (129), the *self* pronoun is the most deeply embedded argument of the ditransitive predicate, which does not match the input to the definition in (123). Further complicating the issue, the *self* pronouns have different antecedents corresponding to different arguments in (129a) and (129b). Thus, even just to capture the mono-clausal cases of argument reflexivity, an account which defines reflexivity as a function on a predicate will require at least three different specifications.

Matters are more complicated for *self* pronouns bound across clauses. First, ECM clauses which were handled simply by Condition A, but required a verb movement for Reinhart and Reuland:

(130) Sandy$_i$ wants herself$_i$ to win the round.

In examining (130), a repetition of (55a), the first question is to consider what the underlying predicates are. First, there is the transitive *win* predicate, and the *want* predicate which takes an experiencer and a proposition. It is possible to render this within the semantics of assigned co-reference, though the mechanism is rather complicated. Firstly, note that with *herself* in a subject position, the form presented in (123) cannot be used. Continuing on the line that the *self* pronoun is a function which works on a predicate, a new form must be derived which works upon a one place predicate. Based on the generalisation that *self* pronouns are valid only in subject positions of non-finite embedded clauses, a form of the *herself* can be proposed which passes the assigned coreference up to the matrix predicate, as shown in (131):

(131) $[\![$to win the round$]\!] = \lambda z.\text{win}(z,\text{the round})$
$[\![$herself$]\!] = \lambda Q_{<e,t>}\lambda P_{<t,<e,t>>}\lambda x.P(Q(\text{herself}))(x) \wedge x = \text{herself}$
$[\![$herself to win the round$]\!] = \lambda P_{<t,<e,t>>}\lambda x.P(\text{win}(\text{herself},\text{the round}))(x) \wedge$
$x = \text{herself}$

In some sense mirroring the covert verb movement analysis from Reinhart and Reuland which created a combined predicate to account for ECM cases, the *self* pronoun takes both the embedded and matrix predicates as arguments, completing the embedded predicate with *herself*, and assigning the identity relation to the matrix predicate.

Turning to raising predicates, there are two different structures to deal with. The first of these is the case wherein the *self* pronoun appears in the lower clause:

(132) a. Jo$_i$ seems [t$_i$ to despise herself$_i$.]

b. * Jo$_i$ seems [t$_i$ to despise her$_i$.]

In (132), a repeat of (58) from above, the analysis is that the lower clause is the domain of reflexivity, with the movement trace providing the necessary antecedent for *herself*, and external argument for the definition of a syntactic predicate. No changes are needed to the semantic definition of the reflexive here, under the assumption that a separate semantic mechanism will resolve the identity between the trace and its antecedent.

The second case is where the *self* pronoun is an argument of the raising predicate itself, shown in the earlier example (59a):

(133) Rich$_i$ seems to himself$_i$ [t$_i$ to outperform his rivals.]

To implement this reflexive relation on a purely semantic level requires the following assumptions. The most intuitive semantic form to give the *seems_to* predicate is a function which takes two arguments: a proposition and an experiencer. The propositional argument would be *t_i to outperform his rivals*. Though some mechanism for interpreting the trace needs to be defined, there are no variables in this proposition which can enter into the kind of identity relation defined in (123), and no room for the kind of manipulation seen in (131). Thus, what would seem to be required here is some form of the *self* pronoun, specific to the raising predicate, which would be able to access the index on the trace within the proposition, and assign that same index to the experiencer argument of *seems to*, such

that both arguments are identified with the same antecedent. Not only does such a formulation sound far-fetched at best, but it dispenses with the assigned co-reference reading of reflexivity: there are simply not enough variables to work with.[8] Given how poorly a semantic account deals with this particular case, it is not surprising that this is one of the key cases leading Reinhart and Reuland to adopt a structural constraint (the Chain Condition) on the distribution of *self* pronouns. Note however, that under a full implementation of the Déchaine and Wiltschko system of pronominal structure, the Chain Condition would not be the correct solution either, as the *self* pronoun and its antecedent would not necessarily share the same indexing, as they would be considered distinct R-Expressions. As such, this structure cannot be given a treatment under the present assigned co-reference analysis.

Turning next to RDPs, a purely semantic approach to *self* pronouns in possessed DPs is fairly easily implemented. Consider the basic case where a *self* pronoun must be bound by the possessor:

(134) Alan$_i$ saw [Stephen$_j$'s picture of himself$_{*i/j}$]

Treating *picture* as a two-place predicate, the same formulation of the the *self* pronoun as in (123) will work. *Himself* defines an identity relation between the two arguments of *picture*.

Far more problematic are cases such as (48), repeated below as (135):

(135) a. I$_i$ told the story about myself$_i$ that John likes to hear.

b. * I$_i$ hate the story about myself$_i$ that John always tells.

Dealing first with the grammatical (135a), if indeed the *self* pronoun is to be viewed as working on predicates, then the starting point must once again be to consider which predicate is being modified. *Story*, here is a one-place predicate, while *told* is a two-place predicate, and the identity relationship would need to hold between the subject of *told* and the single argument of *story*. In a sense then, this is semantically parallel to the ECM case. Indeed, the semantics of the *self* pronoun presented in (131) will account for this case with only the minor alteration of taking the *told* predicate of type $<e<e,t>>$ rather than a predicate of type $<t<e,t>>$ as the second argument.

[8] Aside from the issue of incompatible semantic types, this lack of open arguments also militates against having the *self*-pronouns in the subject positions of monoclausal sentences.

Accounting for the contrast in the pair in (135) requires additional modifications though. Recalling the discussion surrounding this particular example, and the work of Kaiser et al., the difference in judgements is predicted by a sensitivity of the *self* pronoun to the role played by the antecedent: *self* pronouns favour having antecedents which have source θ-roles. Thus, *myself* can be bound by the teller of the story in (135a) but not the hearer in (135b). The most straightforward implementation of this would be to add a caveat to the equivalence statement in the reflexive definition from (131):

(136) $[\![\text{myself}]\!] = \lambda Q_{<e,t>}\lambda P_{<e,<e,t>>}\lambda x.P(Q(myself))(x) \wedge x = \text{myself} \wedge \text{source}(x, Q(\text{myself}))$

(136) adds a two-place source relation to the definition of the *self* pronoun, which when applied to *story* requires that the subject from the verb be the source of the item described in its one-place predicate complement. The addition of this element to the definition of *myself* accounts for the contrast in (135). For the *hate* case, while the identity between the *self* pronoun and the subject holds, the source relationship does not, as it is John telling the story in (135b). These semantic manipulations capture restrictions on the use of *self* pronouns by restricting the functions which can act as inputs to the reflexive *self* pronouns. Inelegant though this solution may appear, it captures the observed data. Similar extensions could be implemented to account for the various other observed phenomena connected with RDPs, such as preference for source of consciousness or agents.

As shown in this section, it is possible to account for most instances of reflexive *self* pronouns using a semantic form which operates on a predicate, adding the condition of assigned co-reference between two arguments. However, even for mono-clausal cases, three forms of such a reflexive are needed. A fourth form is needed for ECM cases where the *self* pronoun straddles two predicates, though with minor adaptations, this same form can be used for *self* pronouns in possessorless RDPs. The basic monoclausal form can also be used for possessed RDPs, and for raising cases where a *self* pronoun appears in the lower clause. The only problematic structure for this analysis of the *self* pronouns is the case where the *self* pronoun appears as an argument of a raising predicate. However, in Chapter 6, it will be shown that using Tree Adjoining Grammar, this last problem can be solved, taking advantage of that formalism's account of raising which crucially does not have a trace of the antecedent in the lower clause.

2.4.2 Emphatics and the EAA

The emphatic *self* pronouns, inclusive and exclusive, adjacent and extraposed, along with the EAA cases, will have a substantially different analysis. These are focus-sensitive operators, affirming either an inclusivity or an exclusivity relationship between the antecedent of the reflexive and some contextually-salient group of alternatives. In this section, I present a working semantic analysis of these forms, essentially re-casting the work of Gast (2006) within a simple Rooth-style alternative semantics.

Similar to the observations of Déchaine and Wiltschko, Gast begins with the underlying assumption that a reflexive is essentially an identity function which he defines as ID:

(137) $\mathrm{ID}(x) = x$

(138) $[\![\text{John himself}]\!] = \mathrm{ID}([\![\text{John}]\!]) = [\![\text{John}]\!]$

This function, he concedes is truth-conditionally trivial, its output being identical to its input, but he argues that it is in interaction with other aspects of the grammar that this function becomes meaningful. Specifically, Gast makes use of focus marking:

(139) The president [himself]$_F$ will open the meeting.

Following the basic principles of Rooth (1992, 1996), this focus marking, through a combined manipulation of pitch accent, duration, or loudness, presupposes a set of alternatives to the focus-marked element.

Recalling the denotation of *himself* from (137) where the emphatic was defined as an identity function, focus interpretation presupposes the existence of alternative functions, rather than alternative individuals. Gast's proposal is that there is a single available alternative, an alterity function OTH:

(140) $\mathrm{OTH}(x) = y \wedge y \neq x$

As defined by Gast, OTH is a choice function, returning a non-specific member of a set of type-equivalent entities distinct from the input. Thus, it is the effect of focus which presupposes the viability of OTH as an alternative to ID, and it is the function of OTH to generate potential alternatives to *the president* in (139).

To account for the various positions and meanings of the emphatics, Gast develops an analysis which derives all the adjacent and extraposed cases (inclusive and exclusive) from

a single underlying form. In particular, his analysis for the exclusive extraposed emphatic involves a two-step movement operation which mimics the effects of quantifier stranding. He treats the exclusive as being base-generated in the adjacent position at [Spec, *v*P]. When the external argument moves to [Spec, TP], the emphatic is left behind. Then, the verb moves to T^0, pied-piping the internal argument, leaving the unmoved emphatic now at the end of the sentence. This movement is motivated by a general constraint which favours the placement of focused elements sentence-finally. Problematic here is the fact that this analysis predicts the possibility of ungrammatical examples of quantifier stranding:

(141) a. [All the kids]$_i$ t$_i$ mowed the lawn.

b. [The kids]$_i$ [all t$_i$] mowed the lawn.

c. * [The kids]$_i$ [mowed the lawn]$_j$ [all t$_i$] t$_j$.

In (141), three permutations of the same sentence are shown, with (141a) being a canonical case, and (141b) being a typical treatment of quantifier stranding. (141c) implements the further movement proposed by Gast, which is ungrammatical with a quantifier, which should be able to be focused.

Furthermore, Gast's analysis does not allow for the possibility of an inclusive adjacent reading, which I have already shown to exist in English. Finally, this single underlying origin analysis precludes any possibility of combining an adjacent and an extraposed emphatic in the same sentence. Again recalling the ME experiment, these were judged to be marginal, but native speakers could make statistically significant distinctions between cases, suggesting that using two of these non-argument *self* pronouns in the same sentence should be possible. In light of these concerns, I will advance an analysis in this section which provides distinct underlying forms, though retaining Gast's core idea of the ID function.

The simplest case to deal with is the exclusive adjacent. First of all, this is the structure which most clearly makes use of focus, which Gast claims to be on the emphatic. Secondly, the sentence in (142) only presupposes alternatives:

(142) Jack himself$_F$ mowed the lawn.

Here, alternatives to Jack are presupposed, but these have the feeling of a possible contrast, rather than an explicit negation of alternatives. This can be achieved using Gast's original

analysis for such cases. The emphatic instantiates the ID function; under focus, alternatives to that function are introduced. Thus, (142) asserts that ID(Jack) mowed the lawn, and presupposes alternatives in which OTH(Jack) mowed the lawn. To keep this applicable to non-subjects, the simplest approach is to treat this as a modifier directly on a DP, regardless of its position.

However, noting the agency and aktionsart restrictions on the sentence-final uses, a constraint can be placed on the exclusive extraposed, such that it can only appear as a modifier to an agentive *v*P. Specifically, this would attach at the v' node, modifying the predicate before the external argument is encountered in composition. Unlike the adjacent, the extraposed, which is essentially an adverbial, will have truth-conditional consequences:

(143) $[\![\text{himself}]\!] = \lambda P \lambda z.P(z) \wedge \forall y[P(y) \rightarrow y = \text{ID}\ (z)]$

The formula in (143) is based upon a Rooth-style semantics for the focus-sensitive particle *only*. It leaves the predicate unaltered, but adds the condition that for all entities which satisfy that predicate, they must be identical to the asserted external argument. This has the effect of ruling out alternative agents, rather than presupposing contrasting alternatives. Furthermore, this is accomplished without positing any extraneous syntactic movement, and places the exclusive extraposed[9] emphatic under the scope of negation.

Turning to the Inclusive cases, Gast provides a complicated syntactic form in order to derive the Inclusive reading.[10] A simpler treatment is proposed by Sæbø (2009), based upon the example in (144):

(144) As Elizabeth Brinker cares for her mother, she knows she herself$_F$ is [at risk of inheriting]$_F$ Alzheimer's disease. (Sæbø 2009, ex 20)

Working through the analysis, Sæbø uses much of Gast's machinery, again treating the emphatic as instantiating the ID function. Just as for Gast, when focused, an alternative function stands in the place of *herself*, and an alternative predicate replaces *at risk of inheriting*. The presupposition is spelled out in (145):

(145) $\phi = R(\text{Alzheimer's})(f(x))$

[9] At this point, the notion of extraposition is removed from the analysis, but the terminology is retained in hopes of minimising confusion.

[10] Interested readers are referred directly to Gast for the full analaysis. As noted above, his analysis does not give an account for the inclusive adjacent emphatics.

This reads as 'there is some relation R, an alternative to *at risk of inheriting*, which holds between Alzheimer's and the entity picked out by some function f, an alternative to ID, acting on x, Elizabeth Brinker.' Here, the context provides an immediate verification of the presupposition, in that Elizabeth Brinker's mother already suffers from Alzheimer's, verifying that the presupposition is true. According to Sæbø, this is all that is required for the inclusive reading to emerge: consistency between established background and the focus presupposition. By extension, where this consistency is absent, either because the background is inconsistent or incomplete, the default interpretation would be one of contrast. Thus, in both the exclusive and inclusive adjacent cases, the emphatic can still be treated as the ID function under focus, with the background providing the differentiation between the inclusive and exclusive uses.

Lastly, there is the inclusive extraposed:

(146) Jack is a drinker himself.

As in the case of the exclusive extraposed, this is better seen as a propositional modifier, rather than simply as a case of shifting the emphatic to the end of the sentence. However, because the inclusive reading relies on being embedded in a background where alternatives are already established, there is no need for additional machinery to ensure that reading comes through. The only fomalisation required is to ensure that the inclusive reading is associated with the subject, which can be done through modification at T′ with the following form:

(147) $[\![\text{himself}]\!] = \lambda P \lambda z.P(\text{ID}(z))$

When the emphatic is focused, this will once again presuppose an alternative relation, the output of which becomes the argument of P. The subject-only restriction can account for another prior contrast:

(148) a. John gave Mary herself the book.

b. * John gave Mary the book herself.

Recalling the original discussion around this example, (148b) has the feeling of a gender mismatch, where the extraposed emphatic can only connect back to *John*. While the inclusive extraposed is insensitive to aktionsart, it is still restricted to subjects:

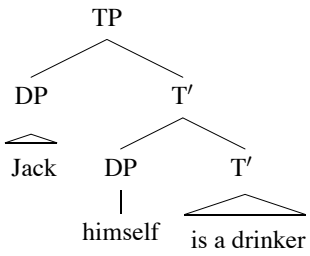
(149) * Sally is fond of many boys, including Steve and Jeff. Sally is fond of Patrick himself.

In (149), a context is constructed which establishes that there are pre-existing members of a set of boys liked by Sally. An exclusive reading is ruled out because the background provides examples which show Patrick is not alone in having Sally be fond of him. An inclusive reading *should* emerge, but does not. Instead, the sentence is ungrammatical. Note also that even replacing *himself* with *herself* will not fix (149) in the given context:

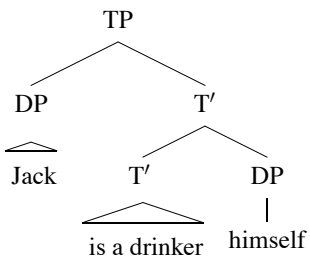
(150) # Sally is fond of many boys, including Steve and Jeff. Sally is fond of Patrick herself.

The gender mismatch is gone, but the presupposition generated by *herself* is that there are alternative people fond of Patrick, not that there are alternative people of whom Sally is fond. The fact that Patrick cannot be inclusively modified in (149) suggests that the inclusive reading is limited to subjects, regardless of position. This can be accomplished in one of two ways: either by making the adjacent version sensitive to the DP to which it attaches, possibly by seeking out nominative case, or by positing two possible structures for the inclusive emphatics:

(151) a.



b.



Like some adverbs, this inclusive can be posited to alternate between a left- or right-adjoined T′ modifier. In both cases, the semantic form is as in (147).

Looking back to (148a), the exclusive adjacent can appear on a non-agent. This suggests that a form which restricts that modifier to only agents would be too severe. The tendency for the exclusive adjacent to associate with agents can be explained by the fact there is an exclusive extraposed emphatic which is more clearly agent-oriented; there is a connection between agentivity and exclusivity. However, the constraint is not so strong as that on the inclusive emphatics, which appear to be strongly subject-oriented. The final state of the analysis is summarised in Table 2.4.

Table 2.4: Semantic Forms and Restrictions on Emphatics

	Modifies	Semantics	Restrictions
Exclusive Adjacent	DP	$\lambda x \mathrm{ID}_F(x)$	None
Exclusive Extraposed	*v*P	$\lambda P \lambda z.P(z) \wedge \forall y[P(y) \rightarrow y = \mathrm{ID}\ (z)]$	Ext. Arg.
Inclusive (Both)	TP	$\lambda P \lambda z.P(\mathrm{ID}_F(z))$	Subjects

With these semantic forms in mind, it is possible to return to the cases from the magnitude estimation task experiment, where participants were able to make a significant distinction between the sentences in (152):

(152) a. Jim himself painted the house himself.

b. Will himself was a subscriber himself

While neither of the forms in (152) received very high acceptability ratings, (152a) was judged to be be significantly more acceptable than (152b). The reason for this becomes clear when considering the proposed syntactic and semantic forms. Recall that the original experiment was designed to test whether agency had an impact on the acceptability of these forms; in (152a) the predicate is agentive, and therefore the sentence final emphatic can be the exclusive extraposed. In (152b), there is no agent, and thus the sentence final position can only be an inclusive extraposed. Keeping in mind that the exclusive adjacent tends toward attachment to agents, the adjacent emphatic in (152b) is also likely to be inclusive, which should be impossible given that both the adjacent and the extraposed inclusives are treated as variations of a single modifier, adjoining either to the left or right. Furthermore, even if the adjacent in (152b) were the exclusive, the presuppositions generated would have

to be consistent with the background which licensed the inclusive, making the contrast necessary for the exclusive reading impossible.

For (152a) on the other hand, it is possible to have both. The sentence-final emphatic is the exclusive extraposed; it adds a semantics similar to *only*, stating that the external argument is the only one who completed the action. Crucially, there is no presuppositional background at play here, as the exclusive extraposed is formulated to be truth-conditional. This allows for the inclusive adjacent to be applied, presupposing that there are others in the background who share the same predicate. (153) provides a context and makes one small change to (152a) to bring this out:

(153) All of Jim's neighbours painted their houses themselves. Not wanting to be outdone, Jim himself painted his house himself.

This is a coherent dialogue in which the first *himself* has the inclusive reading, while the second has the exclusive, a possibility which cannot be generated under Gast's analysis, which specifically predicts that sentences should never be able to have more than one emphatic, though (153) shows this to be fine.

A consequence of the proposed analysis is that it can allow for the generation of a sentence with two extraposed emphatics:

(154) ? John painted his house himself himself.

The inner emphatic, reflecting a lower syntactic attachment, would carry the exclusive reading, while the outer one would carry the inclusive. I mark this with a question mark because, in the context of (153), I do not find this to be jarringly ungrammatical, and can still get the reading. Still, it cannot be denied that the form in (153) is better, possibly on stylistic grounds.

Independent evidence for the two distinct syntactic positions can be found in a pair from Gast:

(155) a. I write a report every week myself.

b. I write a report myself every week. (Gast 2006, ex93-94)

In his analysis, Gast treats these as both being inclusive, and provides a discussion about the difference between them being solely a matter of different positions for *every week*.

However, my interpretation of these sentences is that (155a) is indeed inclusive, but that (155b) can be *exclusive*, reporting that the speaker alone has to write a report every week. (155b) could be the lament of an office-worker whose colleagues never share in the work of weekly report-writing. This follows from the proposed syntactic analysis, in that the exclusive would attach low, while the inclusive would attach high, predicting they would fall on different sides of an adverb.

The structural distinction can also be found in a combination of the emphatics with depictive modifiers:

(156) a. John plays the piano nude himself.

b. ? John plays the piano himself nude.

In (156a), the emphatic can only have the inclusive reading, with John being one of a number of contextually-established nude piano players. Again, the inclusive reading is corresponding to a relatively high syntactic position in the right periphery. For (156b), the inclusive reading is impossible; the sentence is marked as marginal because it comes across as a somehow awkward attempt at expressing the exclusive meaning. As with the relative positioning of an emphatic with an adverbial in (155), the data show that the inclusive reading is associated with a syntactically higher position than the exclusive.

Having given a semantic account of the various types of emphatice, it remains to provide an account of the external argument adjunct. This will be similar to the exclusive extraposed in that it attaches to *v*P. However, it will be different in that the semantic form will need to be loose enough to account for the fact that in the passive, the EAA does not appear with a *self* pronoun. The EAA essentially to re-states the external argument of a predicate. In the active, a *self* pronoun must be used, and the meaning imparted is that there was no additional external contribution to the action. In the passive, there has been no indication at all of the external agent, so it appears as new information.

The EAA itself can be instantiated as a function from $\ll e,\langle e,t\rangle\gg, \langle e,t\rangle\gg$, taking in an entity and a one-place predicate, returning a one-place predicate:

(157) $[\![\text{by}_{EAA}]\!] = \lambda x \lambda P \lambda z.\ \text{perform'}(x, P(z))$

Applied in a passive case, assuming an analysis of the passive in which the predicate enters the derivation already passivised, the derivation of (158a) can be sketched as in (158b):

(158) a. The house was painted by John

b. ⟦was painted⟧ = λ*x*.was_painted(*x*)
⟦by John⟧ = λ*P*λ*z*. perform(john,*P*(*z*))
⟦was painted by John⟧ = λ*z*. perform(john,was_painted(*z*))

The definition in (157) makes use of a predicate *perform*, which relates a given predicate to an entity which performs the action denoted in that predicate. There is no formal connection between the performer of the action and the open argument of the input to perform; this is what allows the EAA to attach equally to active and passive predicates. In an active case, that open argument will be the external argument, but in the passive, it will be the pre-posed internal argument:

(159) ⟦The house was painted by John⟧ = perform(john, was_painted(the house))

The final semantic form gives the interpretation that John performed the action of painting the house. The final consideration is to account for the use of the *self* pronouns in the active cases.

To answer this, I look back to the core cases of reflexivity, using the Déchaine and Wiltschko analysis which treated the *self* pronoun as a full DP. For co-argument reflexivity, the analysis was that the *self* pronoun takes a predicate as its argument, adding the requirement of assigned coreference between two arguments, and treating the *self* pronoun as one of two independent entities. The same can work here, treating the *self* pronoun as a simple type *e* entity, similar to its exempt usage:

(160) a. John painted the house by himself

b. ⟦painted the house⟧ = λ*y*.painted(*y*, the house)
⟦by himself⟧ = λ*P*λ*z*. perform(himself,*P*(*z*))
⟦painted the house by himself⟧ = λ*z*. perform(himself, painted(*z*, the house))
⟦John painted the house by himself⟧ = perform(himself, painted(john, the house))

The semantics is such that the person who performs the action of a transitive predicate will be the same as the external argument of that predicate. Otherwise, an incoherent form would result, where an action would be performed by someone other than its agent. In a

sense, this is a covert instance of assigned co-reference, where a co-referential relationship is forced. Unlike co-argument reflexivity though, the *self* pronoun is not contributing that meaning; the *self* pronoun here is serving as a DP which enters into this externally-constructed relationship. The reason a *self* pronoun is forced here can be found in the syntax, under the assumption that a *self* pronoun is the only DP which can tolerate local binding, Conditions B and C still being in effect.

With this last form, the discussion of the semantics of the English *self* pronouns has come full circle, returning to the original Déchaine and Wiltschko treatment of the *self* pronouns as DPs. For coargument reflexivity, referential content is present, though embedded in a larger function which worked on the predicates with which the *self* pronouns are merged. These definitions extend to account for cases where the *self* pronoun appears to be bound across clauses, and even to include the source requirement uncovered in psycholinguistic experiments on RDPs. For the emphatic uses, the semantic character of the *self* pronoun changes quite radically, coming in the form of an identity function which applies either directly to a DP or embedded within adverbial uses. This semantic distinction is reflected in the syntax in that all the emphatics have the syntax of modifiers rather than arguments. Finally, in its exempt uses, the *self* pronoun will be a DP with no reflexive or identity machinery whatsoever. While this analysis cannot be credited for its parsimony, the facts are captured.

All except one, that is. The present semantic analysis cannot account for (133), repeated below as (161):

(161) Rich$_i$ seems to himself$_i$ [t$_i$ to outperform his rivals.]

While it is true that the simple type *e* version of the *self* pronoun, needed for the exempt cases, could fill the empty argument position of *seem*, there is no way in the present framework to force the identity of the experiencer of *seem* and the agent of *outperform*. This will be taken up in Chapter 6.

2.5 Loose Ends

To close this chapter on the *self* pronouns of English, there are still two issues which have gone unaddressed, both falling under the general category of language change. This section

should be considered speculative, but it could shed some light on perceived irregularities in the use of *self* pronouns in English.

The first issue is the unique status of the third person forms *himself* and *themselves*. According to the pronominal structure proposed by Déchaine and Wiltschko, the English third person *self* pronouns should be treated as bound variables, because they do not fall into the scheme of being constructed from possessive pronouns. This distinction, I feel, would be more important to make if it were the case that the English *self* pronouns as a whole had a consistent semantic form across all their uses. As seen in the discussion from the previous section, the *self* pronouns have multiple functions, from changing verbal predicates to themselves instantiating an identity function. Because there is such a multiplicity of function, it may not be that problematic that there is a slight variation in form.

In the case of *hisself*, the third person singular form expected under a fully possessive paradigm, one could attempt to argue that there is some phonological constraint against the [ss] string which led to the adoption of the irregular *himself*. Of course, this cannot account for the avoidance of *theirselves*. Taking a more diachronic view, König and Siemund note that historically, *self* was an isolated intensifier which appeared in conjunction with accusative pronouns:

(162) Hannibal...hine selfne mid atre acwealde.
'Hannibal killed himself with poison.' (König and Siemund 2000a, ex 14)

This example from Old English shows *selfne* standing beside the accusative *hine*, marking what König and Siemund describe as an unexpected co-indexation of subject and object. By the era of Middle English, pronouns and the *self* intensifier have merged, and the newly formed pronouns even go through a period of being able to function as a clausal subject:

(163) a. Hymself drank water of the wel...

b. Since of ourselves, ourselves are choleric.
(König and Siemund 2000a, ex 17a-b)

The case in (163a), from *The Canterbury Tales*, is most informative in that it shows non-possessive *hymself* being used in what looks like a lost referential function, in parallel with the possessive *ourselves*, here taken from *The Taming of the Shrew*. While this does not account for the discrepancy in the set of *self* pronouns, it at least provides evidence that they

have developed in parallel, sharing a function historically which is now lost. The continued shared function though, is not surprising.

Furthermore, there is the possibility that the question is somewhat moot, as suggested by this example from the Switchboard corpus:

(164) and he did it hisself, so most of that, by now, is falling apart. (SWB2332.DFF A.63)

Such uses of *hisself* and *theirselves* are commonly cited as features of southern American dialects of English. The same phenomenon is observed by Cheshire et al. (1993) in a survey of English usage among students in British metropolitan areas. This is described as a levelling process, regularising the paradigm of *self* pronouns. Should this take hold, a uniform patten may yet emerge.

König and Siemund also bring up a phenomenon which they label as 'creeping reflexives':

(165) a. On behalf of myself and USAir, we would like to thank you for...

b. I think if somebody would have called and asked, both myself and my husband would have been willing to talk. (König and Siemund 2000a, ex 22a-b)

In both of these cases, *myself* is used where *me* would be expected. The term *creeping* is used because there is a sense that the *self* pronouns are expanding their territory, being used in new non-reflexive contexts. They go on to note that this is most prevalent in first and second person cases, which is echoed in the corpus research. All of the exempt anaphors found in the Switchboard corpus are either first or second person, and fully one third of those occur in exactly the same sentence:

(166) How about yourself?

Recall that the Switchboard corpus collects telephone calls between strangers who have never spoken before. In this context, rather than asking *How about you?* speakers may be using the *yourself* as a polite form, somehow softening the request for information.[11] The two sentences in (165), particularly the USAir example, have this same feeling of the

[11]This is based on my own self-observed use of *yourself* in this way. More study is clearly needed.

self pronoun being used to engender a sense of formality which would not be present with *me*. Should this be a trend which continues over time, with the *self* pronouns being used in more non-reflexive or even non-emphatic positions, the data may become so murky as to make the multifarious uses of the English *self* pronouns even more difficult to disentangle.

Chapter 3

The Accidental Reflexive
Korean Long Distance Anaphora

They're not local!
-Tubbs Tattysyrup. *The League of Gentlemen*.

In this chapter, I turn my attention to Korean. Not only does Korean have a set of reflexives which appear to be derived from the referential pronoun paradigm, but it has two simplex forms *casin* and *caki*, which can in turn compose into the compound *caki-casin*. After an overview of the *casin*-based forms, which will cover some similar ground as the non-reflexive uses of the English *self* pronouns, I will focus on *caki*, arguing that it is a bound variable, which can create a reflexive-like meaning when it enters into a binding relationship with a co-argument. Thus, *caki* fits nicely into the category of a ϕP reflexive, having the semantics of a bound variable, and able to enter into local binding relationships, in addition to long-distance uses.

3.1 Korean Reflexives based on *casin*

In beginning a discussion on the reflexive system of Korean, the first challenge is is to determine what exactly are the core cases of reflexives in the language. To look at the

existing literature, this is a matter of some debate. Working from the strict definition of a reflexive from Chapter One, where reflexivity is restricted only to co-arguments, then the best candidate would be the set of compound reflexives, made up of a referential pronoun and the morpheme *casin*, which are generally taken in the literature to be exclusively local:

(167) a. Na$_i$-nun na-casin$_i$-ul piphan-ha-yess-ta.
1SG-TOP 1SG-SELF-ACC criticise-do-PST-DECL
'I criticised myself.'

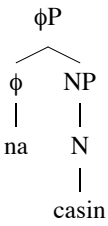
b. Ne$_i$-nun ne-casin$_i$-ul piphan-ha-yess-ta.
2SG-TOP 2SG-SELF-ACC criticise-do-PST-DECL
'You criticised yourself.'

Thus, it is with these that I shall begin.

3.1.1 *pronoun-casin*

As shown in the above examples, the *pronoun-casin* reflexives make use of the same pronoun forms as when the pronouns are used referentially, as opposed to the English case where the reflexive pronouns are (generally) formed from the genitive form. Furthermore, the Korean referential pronouns can have a bound variable interpretation, particularly across clauses (Kang, 1988), giving them the status of ϕ heads. This allows the Korean reflexive pronouns to be uniformly treated as ϕPs in the Déchaine and Wiltschko schema:

(168)



Under this analysis, *casin* has an NP status similar to English *self*. It is worth noting that this structure appears at first to run afoul the standard definition of Korean as a head-final language, but there is additional evidence for the claim that nominal projections in Korean may be head initial: specifically from demonstrative (169) and possessive (170) phrases:

(169) a. ku namca
that boy
'that boy'

b. i cadongcha
this car
'this car'

(170) a. John-uy cip
John-POSS house
'John's house'

b. na-uy sinpal
1SG-POSS shoes
'my shoes'

In (169), demonstratives *ku* and *i* precede the nouns they modify, and in (170), the possessive marker *uy*, appears between possessor and possessee, exactly as in an English head-initial possessive structure. Based on this, a head-initial structure for the reflexives fits into the general pattern of Korean nominal structures.

Under Reinhart and Reuland's account, it could be *casin* which provides the necessary reflexive marking. Taking a GB/Minimalist approach, these reflexives can be accounted for under Condition A, in that they can only be felicitously used when bound within the same clause (Cole et al., 1990; Moon, 1995):

(171) Bill$_i$-i [John$_j$-i ku-casin$_j$-lul miwe-ha-n-ta-ko]
Bill-NOM John-NOM 3SG-SELF-ACC hate-do-PRES-DECL-COMP
mal-ha-yess-ta.
say-do-PST-DECL
'Bill said John hates self.' (Moon 1995, ex 9e)

Moon uses this example to argue that the reflexive *ku-casin*, along with all other members of that paradigm can only be locally bound. This claim is not borne out in the literature though, as multiple authors report examples where *pronoun-casin* forms find non-local antecedents:

(172) a. John$_i$-i Tom$_j$-ul [ku-casin$_i$-i Chicago-ey w-assulttay]
John-NOM Tom-NOM 3M-self Chicago-DAT come-when
manna-ess-ta.
meet-PST-DECL
'John met Tom when he (John) came to Chicago.'
(H.-B. Lee 1976, ex 25b)

b. Sensayngnim-un haksayng-tul-eykey [ku-casin-i ka-kess-ta-ko]
teacher-TOP student-PL-DAT 3M-self-NOM go-will-DECL-COMP
malha-yess-ta.
say-PST-DECL
'The teacher said to the students that he would go.' (H.-B. Lee 1976, ex 25b)

In H.-B. Lee (1976), *ku-casin* is described as obligatory in local contexts (versus *ku*), but still optional otherwise. He however provides no detailed account of where the *pronoun-casin* forms are used non-locally.

A more recent look into the long-distance use of *pronoun-casin* is presented in G. Lee (2000), where it is argued that *ku-casin* (along with all the other *pronoun-casin* reflexives) can be a long-distance anaphor. Part of his argument makes use of examples such as those in (173):

(173) a. Yenghui$_i$-nun [Bob$_j$-i kunye-casin$_{i/*j}$-ul coha-ha-n-ta-ko]
Yenghui-TOP Bob-NOM 3F-self-ACC like-do-PRES-DECL-COMP
sayngkak-ha-n-ta.
think-do-PRES-DECL
'Yenghui thinks that Bob likes herself.' (G. Lee 2000, ex 6)

b. Sarah$_i$-nun [swunye-nim$_j$-i [nay$_k$-ka kunye-casin$_{i/j/*k}$-ul
Sarah-TOP nun-HON-NOM 1SG-NOM 3F-self-ACC
pwulsin-ha-koiss-ta-ko] sayngkak-ha-n-ta-ko] malha-yess-ta.
distrust-do-PROG-DECL-COMP think-do-PRES-DECL-COMP say-PST-DECL
'Sarah said that the nun thinks that I distrust her herself.'
(G. Lee 2000, ex 7)

What is striking about these examples (along with the H-B Lee 1976 examples) is that they obviate local binding: in both cases in (173), *kunye-casin* appears in the direct object position of an embedded clause with a ϕ-feature incompatible subject. Rather than rejecting the sentences outright, Korean speakers are able to extend the binding domain when needed. More interesting is the gloss for the second example, where *kunye-casin* is glossed as the English 'her herself'.[1] G. Lee's contention is that in these long-distance cases, the reflexives have an emphatic reading, which also appears in mono-clausal examples where there are no potential antecedents:

[1] Ungrammaticality of the English form notwithstanding.

(174) a. Ku-casin kuttay ku cangso-ey issci-an-ass-ta.
3M-self that time the place-DAT be-NEG-PST-DECL
'He himself was not at the place at that time.' (G. Lee 2000, ex 3)

b. Ne$_i$-nun ku-casin$_j$-ul salang-hay-yaha-n-ta.
you-TOP 3M-self-ACC love-do-must-PRES-DECL
'You must love him himself.' i.e. the person, not their wealth, fame, status, etc... (G. Lee 2000, ex 4)

Again, these examples a reflexive without antecedent remains grammatical. (174b) most clearly brings out this emphatic reading, where the person is contrasted against his desirable attributes. However, G. Lee reports that this emphatic reading is also found where local binding is possible:

(175) Yengswu$_i$-nun [Chelswu$_j$-ka ku-casin$_{*i/j}$-ul nemwu mitnun-ta-ko]
Yengswu-TOP Chelswu-NOM 3M-self-ACC too much trust-DECL-COMP
sayngkak-ha-n-ta.
think-DO-PRES-DECL
'Yengswu thinks that Chelswu has too much confidence in himself.' (G. Lee 2000, ex 2)

Here, while G. Lee reports the expected local binding, he describes such sentences as having an "English" flavour, and claims that native speaker consultants preferred to replace *ku-casin* with *caki-casin*, a clear local anaphor. This is in line with his observation that the *pronoun-casin* forms were introduced into the language in the late 1930s by novelists under the influence of Western literature. O'Grady (1984) makes the same observation about the use of *ku* in isolation as a pronoun; if indeed (pronoun) *ku* and *ku-casin* are relatively recent additions to the language, then confusion around their use is perhaps not so surprising.

The claim that the *pronoun-casin* forms are emphatics does not originate with G. Lee's analysis. Jayaseelan (1997) makes a similar claim based on the observation that *casin* can attach to proper nouns:

(176) John-casin-i o-ass-ta.
John-self-NOM come-PST-DECL
'John himself came.' (Jayaseelan 1997, ex 43)

In this function, and in attachment to pronouns, Jayaseelan claims that the attachment of *casin* to a (pro)noun brings about a contrastive focus reading. Gast and Siemund (2006)

make a similar observation, defining *casin* as the appositive intensifier in Korean. Lee (2004) notes that such forms retain their R-expression status in that they are still sensitive to Condition C effects:

(177) Ku$_i$-nun [kunye$_j$-ka Yengswu-casin$_{*i/*j/k}$-ul sinloy-ha-n-ta-ko]
3M-TOP 3F-NOM Yengswu-self-ACC trust-do-PRES-DECL-COMP
malha-yess-ta.
say-PST-DECL
'He said that she trusts Yengswu himself.' (G. Lee 2004, ex 17)

Because *Yengswu-casin* is obligatorily free, G. Lee concludes that *casin* does not endow any anaphor-like features on the proper noun *Yengswu*. Taking this usage back to *pronoun-casin*, Jayaseelan goes on to claim that such forms are not anaphors at all, merely modified pronouns, and thus non-local usage is to be expected. G. Lee instead argues that the *pronoun-casin* reflexives are ambiguous between a local reflexive reading, and a non-local emphatic one. He further suggests that this distinction is detectable in speech, as the emphatic form has a distinct pause between the two morphemes, a pause which is not present in the reflexive usage.

3.1.2 Monomorphemic *casin*

This form is possibly the least-studied (or at least least-written-about) of all the Korean reflexives, the simple *casin* in isolation. As stated above, Jayaseelan notes that *casin* can attach to proper nouns with an emphatic effect, but has no examples of *casin* used on its own. Gast and Siemund do not present any data on Korean, but they group Korean among languages which use different forms for reflexivity and appositive emphasis. According to their classification, *casin* should be exclusively dedicated to this emphatic function, while reflexivity is the domain of *caki*. My own consultation with native speakers has confirmed that the emphatic use of *casin* in this fashion is indeed productive:

(178) John-casin-un ppang-ul mek-ess-ta.
John-self-TOP bread-ACC eat-PST-DECL
'John himself ate bread.'

Furthermore, consultants have agreed that this usage has an exclusive reading, signalling that John is alone among bread-eaters (everyone else made different food-choices). Judge-

ments of sentences where *ku-casin* replaces *John-casin* in (178) show the same types of readings, where *ku* is a referential pronoun finding its antecedent in the discourse context while *casin* has the same emphatic function, suggesting that *casin* may have a usage similar to the exclusive adjacent emphatic described for English.

H.-B. Lee has only two examples using *casin* in isolation, making no mention of this exclusive reading:

(179) a. Sensayngnim$_i$-i casin$_i$-uy calmos-ul molu-n-ta.
teacher-NOM self-GEN mistake-ACC not know-PRES-DECL
'The teacher does not know his own mistake.' (H.-B. Lee 1976, ex 1b)

b. Chelswu$_i$-ka Yengho$_j$-lul casin$_i$-i Pusan-ey i-ssultay
Chelswu-NOM Yengho-ACC self-NOM Pusan-DAT go-when
manna-yess-ta.
meet-PST-DECL
'Chelswu met Yengho when he went to Pusan.' (H.-B. Lee 1976, ex 2c)

While no attention is paid to the long-distance nature of (179b), H.-B. Lee is clear in his claim that *casin* in isolation is derived from the reflexive *ku-casin* through an operation of pronoun-deletion. He lists alternative versions of the sentences in (179) as not changing in meaning when *casin* is replaced with any of *ku*, *caki*, *caki-casin*, or *ku-casin*. This analysis predicts the forms in (179) to be possible, whereas it is unclear whether G. Lee's analysis does so, and it seems as though Gast and Siemund would expect such an example to be impossible. Jayaseelan addresses this issue by suggesting that bare *casin* is actually *pro-casin*, where the covert *pro* gets its reference from context, and *casin* retains its emphatic function.

A more recent account dealing specifically with *casin* and Japanese *zibun* comes from Son (2003). Son begins his discussion of *casin* noting that it is equally acceptable for local and long-distance cases:

(180) John$_i$-i [Mary$_j$-ka casin$_{i/j}$-ul pinan-ha-yess-ta-ko]
John-NOM Mary-NOM self-ACC blame-do-PST-DECL-COMP
sayngkak-ha-n-ta.
think-do-PRES-DECL
'John thinks that Mary blames self.' (Son 2003, ex 6a)

The equal acceptability of local and long-distance possibilities for bare *casin* are also noted in Cole et al. (1990):

(181) a. Chelswu$_i$-nun [Inho$_j$-ka casin$_i$-ul sarang-han-ta-ko
Chelswu-TOP 1SG-NOM self-ACC love-do-DECL-COMP
sayngkak-ha-n-ta.
think-do-PRES-DECL
'Chelswu thinks Inho likes himself.' (Cole et al. 1990, ex 19b)

b. * Chelswu$_i$-nun [nay$_j$-ka casin$_i$-ul sarang-han-ta-ko
Chelswu-TOP 1SG-NOM self-ACC love-do-DECL-COMP
sayngkak-ha-n-ta.
think-do-PRES-DECL
'Chelswu thinks I love myself.' (Cole et al. ex 21b)

Judgements for (181) are given as reported in the original. In the first example, only the long-distance reading is cited, though technically speaking either should be possible. (181b) is cited as an example of the blocking effect familiar from research on Chinese *ziji*, which can be long-distance bound, but only when all potential antecedents match each other's ϕ features. Again, the possibility of local binding is not addressed in the original example, but there is no person-feature restriction on *casin*, so it should be the case that *casin* could take the embedded clause subject as an antecedent in (181b), making the sentence grammatical. J.-M. Yoon (1989) claims that *casin* can only be locally-bound by first or second person antecedents; long distance binding by such antecedents is not possible. Taking this claim alongside (181b), suggests that the sentence should indeed be grammatical if locally bound. Furthermore, if *casin* is indeed subject to the blocking effect, then long-distance binding by a first or second person antecedent would only be possible if all intervening antecedents had matching ϕ-features, a structure which Yoon may not have tested.

The presence of the blocking effect is key to Cole et al.'s argument that *casin* falls within the scope of their treatment for long-distance anaphors. They argue that like *ziji*, *casin* undergoes an LF movement to its clause-local Infl head, where ϕ feature agreement takes place. Long distance effects arise from successive-cyclic raising through Infl heads. Because this is restricted to Infl heads, subject-orientation is a consequence of this analysis, the anaphor only being able to be anteceded by an element in the specifier position of an Infl head. This treatment of monomorphemic *casin* is incompatible with the Jayaseelan account, which does not come with this additional consequence of subject-orientation. Son provides one example of *casin* with a non-subject antecedent:

(182) John$_i$-eykey casin$_i$-uy emeoni-ka t$_i$ simpurum-ul siki-ess-ta.
John-DAT self-GEN mother-NOM errand-ACC make-PST-DECL
'To John, self's mother made an errand.' (Son 2003, ex 24b)

Here, *casin* has a dative antecedent, derived from scrambling. A broader selection of such examples appears in I.-H. Lee (1978):

(183) a. John$_i$-ka Tom$_j$-ul [casin$_{*i/j}$-ka cikcep Chicago-ey
John-NOM Tom-ACC self-NOM in person Chicago-DAT
o-la-ko] myenglyeng-ha-yess-ta.
come-FUT-COMP order-do-PST-DECL
'John ordered Tom that self in person come to Chicago.'

b. John$_i$-i Tom$_j$-ul casin$_{i/j}$-uy cip-eyse chacana-yess-ta.
John-NOM Tom-ACC self-GEN house-DAT find out-PST-DECL
'John found out Tom in self's house.' (I.-H. Lee 1978, ex 4c-d)

Unlike Son's example, these from I.-H. Lee do not rely on scrambling. (183a) reads like an example of object control, where *casin* is taking the place of PRO, potentially explaining why the matrix object is listed as the only possible antecedent. (183b) is different again in that there is no sense of a control structure, and the sentence is reported to be ambiguous between subject and object antecedent readings for *casin*. Crediting the example to C. Lee (1973), I.-H. Lee also presents the following:

(184) [Sue$_i$-ka casin$_j$-ul palapo-nun kes]-ka Joe$_j$-eykey culkew-ess-ta.
Sue-NOM self-ACC look at-ADNOM comp-NOM Joe-DAT please-PST-DECL
'Sue's looking at self was pleasing to Joe.' (I.-H. Lee 1978, ex18)

C. Lee deals with this apparent backward anaphora by positing an underlying structure in which *Joe* is a topic/subject of the psych predicate, with a movement operation deriving the observed word order. C. Lee's analysis of *casin* does not take into account the full set of data though, as he only describes *casin* as a reflexive particle for first and second person, apparently discounting *ku-casin* or attachment of *casin* to proper nouns.

One of the few sources to directly acknowledge the confusion around this issue is Li and Takahashi (1995), who offer the following:

(185) a. Chelswu$_i$-ka Inho$_j$-eykeyso [Sunpyo$_k$-ka casin$_{i,*j,k}$-ul
Chelswu-NOM Inho-from Sunpyo-NOM self-ACC
salang-ha-n-ta-ko] tul-ess-ta.
love-do-PRES-DECL-COMP hear-PST-DECL
'Chelswu heard from Inho that Sunpyo loved self.'

b. Chelswu$_i$-ka Inho$_j$-eykeyso [Sunpyo$_k$-ka casin$_{?i,?j,k}$-ul
Chelswu-NOM Inho-from Sunpyo-NOM self-ACC
salang-ha-n-ta-ko] tul-ess-ta.
love-do-PRES-DECL-COMP hear-PST-DECL
'Chelswu heard from Inho that Sunpyo loved self.'
(Li and Takahashi 1995, ex 15a-b)

The two examples in (185) are identical, save for the reported indexing possibilities for *casin*. For some speakers, both the local and long-distance subjects were considered equally possible, but for others, there was degradation of the acceptability of the matrix subject, which came along with a degraded (but not categorically ruled out) acceptability for the matrix dative. So consistent are the reported judgement patterns across multiple examples and informants that Li and Takahashi go so far as to label a Dialect A and Dialect B, with A being the more conservative (185a) and B the more liberal (185b). So while there is some evidence of inter-speaker variation, all these data call into doubt the claim that *casin* is strictly subject-oriented, and therefore call into doubt the Cole et al.'s particular analysis of long-distance anaphora.

Lastly on the subject of *casin*, there is one matter which arises from Son's choice of comparing *casin* with the Japanese *zibun*: logophoricity. Sells (1987) describes *zibun* as a logophor, having similar perspectival sensitivity similar to what Kaiser et al. describe for English *self* pronouns in RDP contexts. If *casin* is a Korean analogue to *zibun*, one would expect these same effects to emerge. My own consultations with native speakers show this not to be the case:

(186) a. John$_i$-un Bill$_j$-eykey [Sue-ka casin$_{i/*j}$-ul miwa-ha-n-ta-ko]
John-TOP Bill-DAT Sue-NOM self-ACC hate-do-PRES-DECL-COMP
mal-ha-yess-ta.
tell-do-PST-DECL
'John told Bill that Sue hates self.'

b. John$_i$-un Bill$_j$-lopute [Sue-ka casin$_{i/*j}$-ul miwa-ha-n-ta-ko]
John-NOM Bill-from Sue-NOM self-ACC hate-do-PRES-DECL-COMP
tul-ess-ta.
hear-PAST-DECL
'John heard from Bill that Sue hates self.'

In (186), the sentences remain exactly the same, save for the alternation between the verbs and the roles (source versus goal) of the matrix clause arguments. If *casin* were logophorically sensitive, then (186b) should emerge with Bill as the antecedent, he being the source of the information. Looking back at the Li and Takahashi examples, the conservative speakers matched this pattern exactly, though it is worth noting that the speakers who relaxed subject orientation for *casin* with the verb *tul* 'hear' (Dialect B) did not do so for *malha* 'tell', suggesting there may be a logophoric effect after all. Still, the dominant pattern aligns with strict subject-orientation, suggesting that *casin* is not a Korean equivalent of the logophoric Japanese *zibun*. This conclusion is supported by the Gast and Siemund typology, which gives the Japanese/Korean pairings of *zibun*/*caki* and *zisin*/*casin*.

3.1.3 *caki-casin*

Turning to *caki-casin*, this is generally held to be restricted to local binding only, again with an example from Moon:

(187) Bill$_i$-i [John$_j$-i caki-casin$_i$-lul miwe-ha-n-ta-ko]
Bill-NOM John-NOM SELF-SELF-ACC hate-do-PRES-DECL-COMP
mal-ha-yess-ta.
say-do-PST-DECL
'Bill said John hates self.' (Moon 1995, ex 9d)

Based on this, Moon applies the same strict locality condition to *caki-casin* as other members of the *pronoun-casin* paradigm. This is echoed by G. Lee who explicitly describes *caki-casin* as a local anaphor.

However, this too is challenged. H.-B. Lee reports a paraphrase of (179b) which has *caki-casin* as an embedded subject bound from the matrix clause. Apparent long-distance binding of *caki-casin* is also noted in J.-M. Yoon (1989):

(188) John$_i$-un [Mary$_j$-ka [caki-casin$_{*i/j}$-i ttokttok-ha-ta-ko]
John-TOP Mary-NOM self-self-NOM smart-do-DECL-COMP
sayngkak-ha-n-ta-ko] malha-yess-ta.
smart-do-PRES-DECL-COMP say-PST-DECL
'John said that Mary thinks that self is smart.' (J.-M. Yoon 1989, ex 7)

In (188), *caki-casin* can take an antecedent from a higher clause, but it can only do so from the next highest clause; in this double-embedding structure, *caki-casin* must still be bound by the nearest subject. J.-M. Yoon uses this to argue that even the cases where *caki-casin* in an embedded subject position is bound from a higher clause reflect a strict clause-local binding constraint. Jayaseelan presents a similar example where *caki-casin* has an incompatible local antecedent:

(189) John$_i$-un [i sakon$_j$-i caki-casin$_{i/*j}$-ul yumyonghake
John-TOP this event-NOM self-self-ACC famous
mantil-ess-ta-ko] sayngkak-ha-n-ta.
make-PST-DECL-COMP think-do-PRES-DECL
'John thinks that this event made self famous.' (Jayaseelan 1997, ex 49)

In this example, *caki-casin* cannot be locally-bound as it requires a human antecedent, the nearest being the matrix topic.

More recently, J.-H. Kim and J.H. Yoon (2009) report an experiment in which sentences containing *caki-casin* and a non-local antecedent are presented:

(190) a. Heera$_i$-nun [tongchanghoy$_j$-ka [caki-casin$_i$-i taumcwu-ey
Heera-TOP alumni assoc-NOM self-self-NOM next week-DAT
kyelhinhana-nun sasil]-ul imi palphyohay-ss-ta]-ko
get married-REL fact-ACC already announce-PST-DECL-COMP
malhay-ss-ta.
say-PST-DECL
'Heera said that the alumni association already announced the fact that she would get married next week.'

b. Jieuni$_i$-ka Sanghoon$_j$-eykey [ipen hakki-ey-nun caki-casin$_i$-i
Jieuni-NOM Sanghoon-DAT this semester-LOC-TOP self-self-NOM
kkok iltung-ul ha-lke-la-ko malhay-ss-ta]-ko
for sure 1st place-ACC do-ASP-DECL-COMP say-PST-DECL-COMP
na-nun al-ko iss-ta.
1SG-TOP know-COMP be-DECL

'I know that Jieun said to Sanghoon that she would be at the top of her class this semester.' (J.-H. Kim and J.H. Yoon 2009, ex 13-14)

J.-H. Kim and J.H. Yoon report no significant difference in acceptability judgements for cases with long-distance binding of *caki-casin* in (190), the distinction being that there is a subject intervening between *caki-casin* and its antecedent in (190a) but not (190b). However, in (190a), *tongchanghoy* is again not a suitable antecedent for *caki-casin*, which J.-H. Kim and J.H. Yoon note resists having inanimate or group-denoting antecedents. Sentences where *caki-casin* has a potential antecedent in a local relationship are not included in the study. As such, while J.-H. Kim and J.H. Yoon have shown that *caki-casin* can have long-distance antecedents, their experiment does not prove that this long-distance reading is possible in the presence of an acceptable local antecedent. However, their experiment makes use of a secondary comprehension task which presents an ellipsis followup to the target items:

(191) Aera-to kulekey malhay-ss-ta.
Aera-too so say-PST-DECL
'Aera said so too.' (J.-H. Kim and J.H. Yoon 2009, ex 19)

Paralleling the strict/sloppy diagnostic for referentiality proposed by Runner, (191) would be presented following (190a), with participants being asked to choose whether it would be interpreted as Aera reporting an announcement of Aera's marriage (sloppy reading), Heera's marriage (strict reading), or neither. For both of the cases in (190), respondents reported a majority of strict readings, demonstrating that long-distance *caki-casin* is behaving in a referential role. This suggests that *caki-casin* does behave differently in long-distance contexts, despite the analysis proposed by J.-M. Yoon. J.-H. Kim and J.H. Yoon also provide the following examples of *caki-casin* in a local context:

(192) John-i iywu epsi caki-casin-ul miwe-ha-n-ta-ko (na-nun
John-NOM reason without self-self-ACC hate-do-PRES-DECL-COMP 1SG-TOP
tul-ess-ta). Kulentey yocum Bill-to kulenta-tela.
hear-PST-DECL by the way these days Bill-too do so-hear.
'(I heard that) John hates self without reason. I heard that these days Bill does so too.' (J.-H. Kim and J.H. Yoon 2009, ex 21)

For the ellipsis sentence in (192), only a sloppy reading is reported, with Bill hating himself rather than John. This, J.-H. Kim and J.H. Yoon claim, is evidence that locally, *caki-casin* can only have a bound interpretation. Taking this along with their experimental findings merely suggests that when clausemate binding is not possible, *caki-casin* has recourse to an alternative means of finding an antecedent. Echoing J.-M. Yoon, Li and Takahashi show that when there are equivalently-possible local and long-distance antecedents, the local is the only one selected:

(193) Chelswu$_i$-ka Inho$_j$-eykeyso [Sunpyo$_k$-ka caki-casin$_{*i,*j,k}$-ul
Chelswu-NOM Inho-from Sunpyo-NOM self-self-ACC
salang-ha-n-ta-ko] tul-ess-ta.
love-do-PRES-DECL-COMP hear-PST-DECL
'Chelswu heard from Inho that Sunpyo loved self.'
(Li and Takahashi 1995, ex 15a)

More importantly, this judgement was consistent across both the dialect groups identified by Li and Takahashi: for all speakers, *caki-casin* was strictly local.

Non-local examples of *caki-casin* under a viable local antecedent can be found though:

(194) a. Joe$_i$-nun [Sue$_j$-ka caki-casin$_{?i/j}$-ul cohahako-iss-ta-ko]
Joe-TOP Sue-NOM self-self-ACC like-PST-DECL-COMP
sayngkak-ha-n-ta.
think-do-PRES-DECL
'Joe thinks that Sue likes self.' (C. Lee 1988, ex 10)

b. Susan$_i$-un [John$_j$-i caki-casin$_i$-ul cohaha-n-ta-ko]
Susan-TOP John-NOM self-self-ACC like-PRES-DECL-COMP
sayngkak-ha-n-ta.
think-do-PRES-DECL
'Susan thinks that John likes self.' (G. Lee 2004, ex 4)

Both C. Lee and G. Lee present almost the exact same example of *caki-casin* in an embedded clause with a third person subject taking the matrix subject as its antecedent. However, both are reporting the same emphatic usage; C. Lee suggests that his example could be disambiguated using either *Joe-casin* or *Sue-casin*, while G. Lee notes that the matrix-antecedent in his example arises when a pause is placed between *caki* and *casin*, signalling the emphatic use of the *casin* affix. This correlates with the findings from the J.-H. Kim and

J.H. Yoon study which suggested that non-local uses of *caki-casin* were not in fact cases of binding. Here, non-local uses of *caki-casin* are treated as emphatics, which Jayaseelan has claimed are not anaphoric. That J.-H. Kim and J.H. Yoon reported sloppy readings for (192) where *caki-casin* was locally bound gives further credence to the G. Lee position that *caki-casin* and *pronoun-casin* are all ambiguous between local reflexive and potentially long-distance emphatic uses.

The other major constraint on *caki-casin* worth noting is a ϕ-feature constraint: *caki-casin* is limited to third-person antecedents. This, however, is attributed to *caki*, as *casin* is known to be able to take first and second-person antecedents. For Jayaseelan, to bring *caki-casin* in line with his analysis of *casin* as an emphatic affix, he calls *caki* a third person pronoun, and treats it as non-anaphoric. Because at this point the discussion on *caki-casin* is hinging on the nature of *caki*, it is to that form which I now turn.

3.2 *Caki*

As seen in the previous section, there are two broad positions on the status of *casin*: one camp sees *casin* as a (possibly *the*) reflexive marker in Korean, while the other sees *casin* as an emphatic marker. A similar state of affairs exists for *caki*, with the camps breaking down into claiming that *caki* is a pronominal versus claiming it is an anaphor. Here, I will first sum up the case for claiming *caki* is a pronoun, then move into its treatment as an anaphor, closing with evidence for a bound variable treatment, including arguments presented in Storoshenko (2007). The literature on *caki* is extensive, and fraught with contradictory (and sometimes replicated) analyses. This section lays out those contradictions in detail, with the goal of giving an exhaustive account of the data that need to be accounted for.

3.2.1 The Pronominal Argument

One of the earliest statements to the effect that *caki* should be treated as a pronoun is found in I.-H. Lee, who notes, along the same lines as Jayaseelan, that in its formation of a third-person reflexive *caki-casin*, *caki* has a function that is pronoun-like. However, Lee uses the term “pseudo-pronoun” making it clear that he does not view *caki* to be on par with referential pronouns like *ku*.

In his argument that *caki* should be seen as a pronominal, Jayaseelan presents three more pieces of evidence. The first of these is that *caki* is anti-local, as would be expected of a pronoun under Condition B:

(195) ?? John$_i$-un caki$_i$-lul miwe-ha-n-ta.
John-TOP self-ACC hate-do-PRES-DECL
'John hates self.' (Jayaseelan 1997, ex 45)

Here, Jayaseelan shows that local binding of *caki* is questionable. As a response to other cited examples where local binding of *caki* is fine, he notes that Condition B effects do not seem to be strong in Korean regardless of the pronoun:

(196) Nay-ka nay-lul miwe-ha-n-ta.
1S-SUBJ 1S-ACC hate-do-PRES-DECL
'I hate me.' (Jayaseelan 1997, fn 23, ex ii)

Jayaseelan argues that if speakers accept the clear Condition B violation with first person *nay* repeated in the sentence, then local binding of *caki* cannot be taken as evidence that *caki* is not a pronoun, and the reported degraded reading leads to an anti-local analysis.

Jayaseelan's second argument that *caki* should be considered a pronoun comes from its plural form *caki-tul*:

(197) Pierre$_i$-ka Marie$_j$-eykey caki-tul$_{i+j}$-uy sacin-ul poyocwu-ess-ta.
Pierre-NOM Marie-DAT self-PL photo-ACC show-PST-DECL
'Pierre showed Marie selves' photograph.' (Jayaseelan 1997, ex 46a)

In (197), *caki-tul* refers to the collective of Pierre and Marie, showing split antecedence, a hallmark of pronominals. This particular argument recurs in the literature, with the same facts being used to reach the same conclusion in Huang (2000).

The third and final argument for the treatment of *caki* as a pronoun is that it takes discourse referents:

(198) Pierre$_i$-nun Marie-lul hosangha-ess-ta. kunye-nun caki$_i$-ka akki-ko
Pierre-TOP Marie-ACC think of-PST-DECL 3F-TOP self-NOM cherish-and
salang-ton sonye-i-ess-ta.
love-RELAT young girl-be-PST-DECL
'Pierre thought of Marie. She was the young girl self cherished and loved the most.' (Jayaseelan 1997, ex 47)

In this example, *caki* is the subject of the relative clause, finding its antecedent in prior discourse. Taken together with apparent anti-locality and split-antecedence, Jayaseelan views this potential for extra-sentential reference as the last proof that *caki* should be considered a pronoun.

Cole et al. (1990) provide a separate line of argumentation that reaches the same conclusion. Recalling their earlier example in which *casin* showed a blocking effect, that same effect is not present for *caki*:

(199) Chelswu$_i$-nun [nay$_j$-ka caki$_i$-lul sarang-han-ta-ko] sayngkak-ha-n-ta.
Chelswu-TOP 1SG-NOM self-ACC love-do-DECL-COMP think-do-PRES-DECL
'Chelswu thinks I love self.' (Cole et al. ex 24)

As shown in (199), unlike *casin*, *caki* does not manifest any blocking effects. That is, the presence of a first-person local antecedent does not block *caki* from taking the long-distance third person subject as its antecedent. Taking this along with anti-locality, Cole et al. conclude that *caki* is a pronominal, and does not constitute a counter-example to their theory of long-distance anaphora.

The Cole et al. argument is somewhat theory-internal in that it hinges upon acceptance of their movement-to-Infl account of long-distance anaphors. The other three observations: anti-locality, split-antecedence, and sentence-external reference are the three most-commonly cited reasons for treating *caki* as a pronoun. These can be listed among the claims made for the felicitous use of *caki*:

Table 3.1: Distribution of *Caki*: First Summary

Environment	Status
Local Antecedent	Ungrammatical
Non-Local Subject Antecedent	Grammatical
Split Antecedent in Plural	Grammatical
Discourse Antecedent	Grammatical

3.2.2 *Caki* as Anaphor

Contrary to the account of *casin* as a reflexive-marker, there is a strong position in the literature which assigns that function to *caki*. This begins with C. Lee (1973) and is continued by H.-B. Lee, who describes *caki* as obligatory in local contexts:

(200) a. Chelswu$_i$-nun caki$_i$-lul chwuchenha-yess-ta.
Chelswu-TOP self-ACC recommend-PST-DECL
'Chelswu recommended self.' (H.-B. Lee 1976, ex 13a)

b. * Chelswu$_i$-nun ku$_i$-lul chwuchenha-yess-ta.
Chelswu-TOP 3M-ACC recommend-PST-DECL
'Chelswu recommended him.' (H.-B. Lee 1976, ex 13b)

Because *Chelswu* and *ku* must be disjoint in reference in order to make (200b) grammatical, H.-B. Lee here concludes that in cases of clausemate reflexivity, *caki* is obligatory.

This is at odds with the claim that *caki* is anti-local. One possible explanation for the oddness of the example which Cole et al. and Jayaseelan use to illustrate the anti-locality of *caki* comes in the predicate. Recall that they use the predicate *miwe* "hate", a mental state. Chang (1977) claims that declarative sentences with non first-person subjects and predicates of mental states are actually quite rare in Korean, and that some researchers would mark (201) as ungrammatical on this basis:

(201) Joe-nun Sue-ka coh-ta.
Joe-TOP Sue-NOM good-DECL
'Joe is fond of Sue.' (Chang 1977, ex 12i)

If indeed speakers are reacting to the oddity of a speaker making declarative statements about the mental states of a third party, then this could be the reason for the ungrammaticality of some local uses of *caki*. However, this is countered by a claim made in C. Lee (1988), where it is reported that local uses of *caki* actually *improve* with predicates of mental states:

(202) Sue$_i$-nun caki$_i$-lul cwuk-i-ess-ta.
Sue-TOP self-ACC kill-be-PST-DECL
'Sue repressed herself.' (C. Lee 1988, ex 9)

C. Lee reports that in a literal sense, (202) would be ungrammatical, but in a figurative sense of emotional self-repression rather than physical immolation, the sentence is perfectly fine. A clear physical predicate is almost as bad:

(203) ? Joe$_i$-nun caki$_i$-lul chi-ess-ta.
Joe-TOP self-ACC hit-PST-DECL
'Joe hit self.' (C. Lee 1988, ex 8)

This, according to C. Lee, is just slightly better than the literal version of (202), and would be unambiguous if it appeared in an embedded context: *caki* would switch to long-distance binding. This mix of local and long-distance uses of *caki* is directly addressed in Park (1986), who uses *caki* to argue for a parametrization of Condition A, in which there may or may not be reference to a binding domain in Condition A for a given language. Later, Cho (1996) cites a lack of anti-locality constraints in his argument against the treatment of *caki* as a pronoun.

Still, while there is disagreement over whether or not *caki* can be locally-bound, there is universal agreement that it does indeed allow for some long-distance binding. It is perhaps for this reason that *caki* is referred to as an anaphor virtually interchangeably in the literature with "reflexive": while there is dispute over whether or not *caki* is felicitous with a co-argument antecedent (the logical definition of a reflexive), long-distance uses are universally agreed to be possible. For the sake of consistency in the use of "reflexive" to refer only to coargument contexts, I will, following Cho (1996), use the term "anaphor" as the principle analysis opposing the treatment of *caki* as a pronoun.

Further confounding the issue though is the fact that *caki* is not unconstrained in long-distance environments. Subject orientation of *caki* is expressed at least as far back as Lee (1973), who observes that *caki* is coreferential with sentential subjects or topics. This is echoed by H.-B. Lee:

(204) * John-i Tom$_i$-eykey [caki$_i$-ka ku-kes-ul hal-kes]-ul
John-NOM Tom-DAT self-NOM that-thing-ACC do-COMP-ACC
myenglyeng-ha-yess-ta.
order-do-PST-DECL
'John ordered Tom that he should do it.' (H.-B. Lee 1976, ex11b)

Earlier, it was shown that in this sort of object control structure, *casin* obligatorily took the non-subject antecedent. Here, however, H.-B. Lee reports the opposite for *caki*, claiming

that a reading which has *caki* taking the dative *Tom* as its antecedent is ungrammatical. Chang (1977) similarly repeats the subject-or-topic claim, despite providing some apparently contradictory data:

(205) Joe$_i$-eykey [Sue-ka caki$_i$-lul salang-ha-nun kes]-ka
Joe-DAT Sue-NOM self-ACC love-do-ADNOM COMP-NOM
punmyengha-yess-ta.
clear-PST-DECL
'To Joe it was clear that Sue loves him.' (Chang 1977, ex 3)

At first glance, this appears to be a case of *caki* taking a matrix dative as its antecedent, rather than a local subject. This could then be considered an argument for an anti-local account of *caki*. However, from the surrounding discussion, it is clear that despite the dative marking, Chang takes *Joe* to be the topic of (205), even if there is no explicit *-nun* topic marking, an echo of C. Lee's claim about dative subjects of psych predicates.

In later works, there is a shift in the rhetoric. By the late 1980's, it is fairly well-established that *caki* is *not* strictly subject oriented (Park, 1986; O'Grady, 1987; Lee, 1988; Yoon, 1989). Still, because subject-orientation is a consequence of the Cole et al. analysis of long distance anaphors in East-Asian languages, there is a sense from some authors (see S.-Y. Kim 2000) that *caki* should be subject-oriented, despite the fact Cole et al. explicitly state that *caki* does not fit their analysis. Sohng (2003) is a relatively recent example of a more moderate approach, borrowing a line from Moon (1995), calling *caki* "weakly subject oriented."

Similar to examples seen earlier with *casin*, there have been efforts to ascribe a logophoric analysis to *caki*:

(206) a. John$_i$-i Mary$_j$-eykey [caki$_{i/*j}$-ka am-i-la-ko]
John-NOM Mary-DAT self-NOM cancer-be-DECL-COMP
malha-yess-ta.
say-PST-DECL
'John told Mary that self has cancer.' (J.-M. Yoon 1989, ex 20a)

b. John$_i$-i Mary$_j$-lopwute [caki$_{i/j}$-ka am-i-la-ko]
John-NOM Mary-from self-NOM cancer-be-DECL-COMP
tul-ess-ta.
hear-PST-DECL
'John heard from Mary that self has cancer.' (J.-M. Yoon 1989, ex 20b)

J.-M. Yoon explains the contrast in judgements for (206) to be indicative of a logophoric use of *caki*, noting that while in the example with *malha* 'say', only the source of information is an acceptable antecedent, the logophoric effect is not strong enough with *tul* 'hear' to make *Mary* the only possible antecedent. Even when the source-hood of *Mary* makes non-subject binding possible, the subject is still the most likely interpretation.

J.-M. Yoon reports a more categorical contrast while trying to develop the notion of pivot:

(207) a. John$_i$-i Mary$_j$-eykey [Tom$_k$-i caki$_i$-lul
John-NOM Mary-DAT Tom-NOM self-ACC
po-le-o-ass-ta-ko] malha-yess-ta.
see-to-come-PST-DECL-COMP say-PST-DECL
'John told Mary that Tom came to see self.' (J.-M. Yoon 1989, ex 22a)

b. * John$_i$-i Mary$_j$-eykey [Tom$_k$-i caki$_i$-lul po-le-ka-ass-ta-ko]
John-NOM Mary-DAT Tom-NOM self-ACC see-to-go-PST-DECL-COMP
malha-yess-ta.
say-PST-DECL
'John told Mary that Tom went to see self.' (J.-M. Yoon 1989, ex 22b)

The contrast between these two sentences is in the embedded clause verbs, one with *o* 'come' and the other with *ka* 'go'. J.-M. Yoon claims that *o* 'come' requires speaker empathy with the goal/destination, in this case *John*, making *John* the pivot in the sentence, dictating that only he can be the antecedent for *caki*. Conversely, *ka* 'go' requires agent empathy, making *Tom* the pivot, thus blocking *caki* from taking the matrix subject as the antecedent. However, J.-M. Yoon further notes that this is not an iron-clad constraint, saying that (207b) could be possible in a scenario where Tom has gone to a place John can reasonably be expected to be found, but John is not actually there at the time of utterance.

A consistent feature of the literature on *caki* is the inconsistency of the analyses offered.[2] Similar or identical sentences in the hands of different researchers can yield conflicting judgements. The following two examples are taken from Park (1986), who uses these examples as part of an argument against the subject-orientation of *caki*:

[2]Just in case this was not already clear.

(208) a. John$_i$-un Mary$_j$-eykey [caki$_{i/j}$-uy cip-elo ka-la-ko]
John-TOP Mary-DAT self-GEN home-to go-FUT-COMP
myenglyeng-ha-yess-ta.
order-do-PST-DECL
'John ordered Mary to go to self's house.' (Park 1986, ex 19b)

b. John$_i$-un Tom$_j$-ul caki$_{i/j}$-uy cip-elo pona-ess-ta.
John-TOP Tom-DAT self-GEN home-to send-PST-DECL
'John sent Tom to self's house.' (Park 1986, ex 20a)

For both of these cases, Park reports complete ambiguity. While it is not a direct analogue to the earlier object control case where *caki* remained strictly subject-oriented in the eyes of H.-B. Lee, in this case the matrix dative is available as an antecedent for *caki*. More importantly, both of these sentences involve cases where the indirect object is being ordered into motion, which J.-M. Yoon might expect would make the indirect objects the pivots in (208), yet the subject remains a possible antecedent in both cases.

Similarly, Sohng (2003) has a different take on sentences with *malha* 'say':

(209) John$_i$-i Mary$_j$-eykey [Tom$_k$-i caki$_{i,?j,k}$-lul coaha-n-ta-ko]
John-NOM Mary-DAT Tom-NOM self-ACC like-PRES-DECL-COMP
malha-yess-ta.
say-PAST-DECL
'John told Mary that Tom likes self.' (Sohng 2003, ex 11a)

While J.-M. Yoon reported that for sentences where *malha* 'say' was the matrix verb only the matrix subject (source) could be the antecedent for *caki*, Sohng reports that any of the matrix subject, matrix dative, or embedded subject can antecede *caki* in the embedded clause, though he notes a slight degradation for the matrix dative. The flexibility of the judgement calls into question the logophoric treatment of *caki*, as taking this along with J.-M. Yoon's judgement for the *tul* 'hear' cases erases any sense of *caki's* antecedent varying with source-hood.

Still, perspective or point of view appears repeatedly in the literature. An early example of this is found in Chang (1977), where it is claimed that *caki* is used in contexts which make reference to the point of view of *caki*'s antecedent. This is justified with the following contrast:

(210) a. ?? Joe$_i$-ey kwanha-e malha-ca-myen, [caki$_i$-ka Sue-lul
Joe-DAT concerning-DAT say-at-if [self-NOM Sue-ACC
slangha-nun kes]-ka punmyengha-yess-ta.
love-ADNOM COMP-NOM clear-PST-DECL
'Talking about Joe, it was clear that self loves Sue.' (Chang 1977, ex 17)

b. Joe$_i$-uy kwancem-eyse po-ca-myen, [caki$_i$-ka Sue-lul slangha-nun
Joe-POSS viewpoint-from look-at-if [self-NOM Sue-ACC love-ADNOM
kes]-ka punmyengha-yess-ta.
COMP-NOM clear-PST-DECL
'From Joe's point of view, it was clear that self loves Sue.'
(Chang 1977, ex 19)

Chang ascribes the subtle difference between the two sentences in (210) to the fact that the first one is uttered from the speaker's own perspective, while the second one is explicitly framed in *Joe*'s point of view. Again, the judgement is not categorical; the sense is not that when a different point-of-view is invoked it would be ungrammatical to use *caki*, merely that the pronoun *ku* would be better.

This influence of point-of-view is not dissimilar from C. Lee's claims that *caki* is most acceptable with verbs reporting mental states. Such structures are also at the heart of C. Lee's early work on defining the structural relationships which need to hold between *caki* and its antecedent:

(211) a. [Sue$_j$-ka caki$_i$-lul palapo-nun-kes]-ka Joe$_i$-eykey
Sue-NOM self-ACC look at-ADNOM-COMP-NOM Joe-DAT
cilkep-ess-ta.
pleasant-PST-DECL
'Sue's looking at self was pleasing to Joe.' (C. Lee 1973, ex 79a)

b. * [Sue$_j$-ka Joe$_i$-lul palapo-nun-kes]-ka caki$_i$-eykey
Sue-NOM Joe-ACC look at-ADNOM-COMP-NOM self-DAT
cilkep-ess-ta.
pleasant-PST-DECL
'Sue's looking at Joe was pleasing to self.' (C. Lee 1973, ex 79b)

The grammatical example is presented as a challenge to the claim that *caki* does not allow for backward anaphora. Lee's ultimate analysis is that (211a) is derived from an underlying form in which *Joe* originates as a sentence-initial topic. The observed word order is a

result of a transformational operation available for psychological predicates which allows the dative-marked topic to be moved rightward, across the subordinate clause. (211b) is ungrammatical because in the proposed underlying form, *caki* would precede its antecedent (ruled out by Lee's analysis which pre-dates the definition of c-command), and in the surface structure there is no command relationship between *Joe* and *caki*. Using c-command, the ungrammaticality of (211b) is not surprising, as neither *Sue* nor *Joe* can c-command out of the embedded clause. (211a) becomes more of a challenge, requiring either the retention of Lee's analysis where there is an underlying form in which *Joe* c-commands the embedded clause, or the positing of a later movement of *Joe* into a similar position, with the caveat that the resolution of *caki*'s relationship with its antecedent be held over until LF.

This kind of backward anaphora embedded clause example is frequent in the early literature on *caki*. (211) shows a case where *caki* appears inside a subject clause, taking the direct object as its antecedent. Relative clauses which contain a *caki* whose antecedent is in the higher clause are also frequent:

(212) a. [Caki$_i$-ka sa-n] chayk-i John$_i$-ul kippukeyha-ess-ta.
self-NOM buy-ADNOM book-NOM John-ACC please-PST-DECL
'That book that self bought pleased John.' (O'Grady 1984, ex 50)

b. ?? [Caki$_i$-ka sa-n] chaek-i John$_i$-uy-chinkwu-ul
self-NOM buy-ADNOM book-NOM John-GEN-friend-ACC
culkepkeyhaecwu-ess-ta.
please-PST-DECL
'That book that self bought pleased John's friend.' (O'Grady 1984, ex 54)

O'Grady reports a diminished acceptance for the case where *caki*'s antecedent is itself embedded within a following genitive DP, but does not mark the reading as completely ungrammatical. Still, he reports no problem where the antecedent for *caki* is the following direct object. This is at odds with his later description of an almost identical example (*wrote* replaces *bought*; otherwise they are identical) in O'Grady (1987). There, the sentences in (212) are described as uncommon and sometimes not accepted by native speakers. Again, an account which required *caki* to at least have a c-commanding antecedent would require some analysis which had the antecedent in a structurally higher position than the subject. C. Lee's movement analysis for (211a) is linked to the fact that the example uses a psych

predicate. Looking at (212), this is also a psych predicate, so it could be possible to invoke the same kind of derivation in which the experiencer has a higher c-commanding position.

In addition to cases where *caki* is embedded in a subject clause, or a relative clause in the subject position, (210) provided an example in which the antecedent for *caki* was embedded within a conditional clause. Recall that this was fine, so long as the conditional set up a context which aligned the speaker's point-of-view to the embedded antecedent. This is problematic for any analysis in which *caki* should require a c-commanding antecedent: no movement, covert or otherwise, should ever allow a DP to c-command out of such a conditional clause. O'Grady (1987) presents a similar example with *caki* having potential antecedents embedded within a subject clause:

(213) [John$_i$-i [Sam$_j$-i ssu-n kisa-lul] swuciphayno-at-ta-nun]
John-NOM Sam-NOM write-ADNOM article-ACC keep-PST-DECL-ADNOM
sasil-i caki$_{i,j}$-uy tongsayng-ul kippukeyha-yess-ta.
fact-NOM self-GEN brother-ACC please-PST-DECL
'The fact that John kept the article that Sam wrote pleased self's brother.'
(O'Grady 1987, ex 54)

According to the presented judgement, either of the embedded subjects can act as an antecedent for matrix clause *caki*, though in a footnote O'Grady notes that the acceptability is degraded with each level of embedding. Furthermore, it is worth noting that in this case and the earlier case where the antecedent for *caki* was found inside a conditional clause, there is no viable human antecedent for *caki* in a c-commanding position. In light of these facts, the judgement for (211b) is almost unexpected, as the sentence is starred as ungrammatical with a similar structure. Either there is a real contrast between the different types of subordinate clauses inside which *caki* can find an antecedent, or this is another case of different sources citing more or less conservative judgements.

O'Grady's discussion of (213) forms the foundation of his analysis that *caki* is versatile in terms of the possible antecedents it can take, but that these are ordered in a rigid hierarchy. This is another common theme in the overall literature on *caki*: ranked antecedents. W.-C. M. Kim (1976) assigns numerical values to various structural configurations, with penalties for number of clause boundaries separating *caki* and the candidate antecedent, thus calculating the optimal antecedent for *caki*. Her algorithm assigns the highest value to immediate precedence, which suggests again that anti-locality is not a property of *caki*.

O'Grady (1987) defines a three-tier hierarchy based on grammatical function, considering subjects, VP complements, and "other" NPs, along with a Priority Principle which dictates (contra W.-C. M. Kim) that *caki* will take the highest possible antecedent. It is this principle which allows for non-c-commanding antecedents in O'Grady's system: if there is no higher potential antecedent, then any NP may serve as an antecedent. A similar judgement can be found in J.-M. Yoon (1989); while there is no formal statement of a hierarchy, some examples carry ranked indexation on *caki* (i.e. *i>j*) generally indicating that non-local antecedents are somehow preferable. Sohng (2003) makes use of a similar notational device to indicate a general preference for subject over non-subject antecedents. S.-Y. Kim (2000) also develops a hierarchy based on grammatical roles, though somewhat more articulated than O'Grady's, with six different levels:

(214) topic > subject > object of verb > object of preposition > genitive NP > object of comparative

Similar to previous efforts, S.-Y. Kim's system includes a scoring method: in cases where there are multiple possible antecedents, the preferred choice will be the one which is further away from *caki* on the hierarchy. Kim does not get into any structural detail though, and avoids the issue of considering *caki* with respect to non-c-commanding antecedents.

These ranked approaches often make use of examples in which a highly-ranked DP, for example a subject, is not a possible antecedent for *caki*. The following examples from O'Grady illustrate this:

(215) a. Nay-ga John$_i$-ul caki$_i$-eykey kewul-lo pichwuepoyecwu-ess-ta.
1SG-NOM John-ACC self-DAT mirror-in show-PST-DECL
'I showed John to self in the mirror.' (O'Grady 1987, ex 18)

b. Nay-ga John$_i$-eykey caki$_i$-lul kewul-lo pichwuepoyecwu-ess-ta.
1SG-NOM John-DAT self-ACC mirror-in show-PST-DECL
'I showed John to self in the mirror.' (O'Grady 1987, ex 19)

As noted earlier, *caki* is limited to third-person antecedents; this was seen in the Cole et al. example showing that *caki* was not subject to the blocking effect. Here, O'Grady (1987) has given two examples of *caki* in a simplex clause with a first-person subject. In both cases, *caki* resists binding from the subject, and takes the third-person non-subject argument as its antecedent. All other things being equal, if the antecedent were a third person human

(or at least animate) DP, then the subject would be the most natural choice according to O'Grady. Similar examples are shown in S.-Y. Kim (2000). These approaches are based on the idea that *caki* will only vary from long-distance subject orientation when the matrix subject is an incompatible antecedent. This has the appeal of being more straightforwardly predictive, but does not readily allow for the possibility that *caki* will ever be ambiguous; rather it mechanically selects one ideal antecedent for any given instance of *caki*. The only exception to this being in cases like O'Grady's (213), where ambiguity results when there is no ranked possible antecedent for *caki*.

Such examples are based on the assumption that *caki* can only take third-person antecedents. This too has been challenged. Firstly, there is at least one report of variation between speakers, with Li and Takahashi (1995) noting that for one consultant, *caki* was interchangeable with *casin*: both were fine with first and second person antecedents. However, this one case can be viewed as exceptional. Still, there are situations in which *caki* can be used to refer to second-person referents.

Firstly, there is the generic or arbitrary use of *caki* which appears in the literature from time to time:

(216) John$_i$-un [caki$_{arb/i}$-ka caki$_{arb/i}$-uy calmoss-ul kochi-eya
John-TOP self-NOM self-GEN faults-ACC correct-should
ha-n-ta-ko] sayngkakha-n-ta.
do-PRES-DECL-COMP think-PRES-DECL
'John thinks that one/she should correct one's/his faults.' (Sohng 2004, ex 13)

In (216), *caki* is ambiguous between a bound reading, with *John* as its antecedent, and a generic reading. Sohng also argues for an addressee-oriented second-person reading of some antecedentless instances of *caki*:

(217) Caki-ka chakhay.
self-NOM good
'You are good.' (Sohng 2003, ex 16a)

Sohng describes this as the inherent reference of *caki*, which would surface in contexts where there are no possible antecedents for *caki*. My own consultation with native speakers suggests that this can be used deictically, with two instances of *caki* in a single sentence referring to different people, provided enough explicit pointing:

(218) Caki$_i$-ka caki$_j$-lul piphan-ha-yess-ta.
self-NOM self-ACC criticise-do-PST-DECL
'You criticised you.' (Different people)

These are somewhat different from the discourse-binding cases which make up a part of the *caki*-as-pronoun argument. There, a clear referent is found in the prior discourse (usually the prior sentence). Here, there is either a generic non-specific reading for *caki*, or a situation where extra-linguistic factors (pointing/gesture) fix the reference for an antecedentless *caki*.

These last cases are at the periphery of the treatment of *caki* as an anaphor, and are generally regarded as exceptional. At the core of the *caki*-as-anaphor argument are those cases where *caki* is bound by a sentential subject, ideally from a higher clause. Generally speaking, there is agreement that *caki* should have a c-commanding antecedent, along the lines of Condition A, but without any domain restriction. As discussed, there is even variation along the dimension of whether or not *caki* more easily takes more or less local antecedents. There is discussion of the thematic role of the antecedent, and factors such as empathy or pivot, some of which may be able to be subsumed under a broader notion of topic. A connection between *caki* and topicality is in itself not surprising, as any case of forward anaphora with *caki* will dictate that the antecedent be at least a part of the conversational background by the time *caki* is introduced. This is nothing special about *caki* or Korean, just a logical observation that the antecedent would be uttered before the anaphor.

From the outset, the use of the term "reflexive" appears to be a misnomer. While it is true that *caki* is found in reflexive contexts, even those sources which use the term show *caki* in long-distance contexts, where *caki* and its antecedent are not coarguments. This, in my opinion, misuse of the term most likely stems from the contention that *caki* should fall under some looser version of Condition A. In terms of Reinhart and Reuland, *caki* presents something of a challenge. Taking a simple case of co-argument reflexivity with *caki*, one would assume that it would be *caki* which provides the reflexive marking, but such an analysis would lead to a Principle A (every reflexive-marked syntactic predicate is reflexive) violation every time *caki* is used with a non-local antecedent: *caki* would be providing reflexive marking on predicates which were not, in fact, reflexive. This suggests that *caki* itself would not be seen as a reflexive marker under the Reinhart and Reuland system then. To reconcile this with Principle B (every reflexive semantic predicate must

be reflexive marked), the only conclusion can be that the underlying semantic form of a mono-clausal sentence with *caki* is not true reflexivity, but rather some other form which mimics the effect. This will be explored later in the chapter. Updating the distributional chart from the previous subsection presents the following picture:

Table 3.2: Distribution of *Caki*: Second Summary

Environment	Status
Local Antecedent	Disputed
Non-Local Subject Antecedent	Grammatical
Non-Local Non-Subject Antecedent	Disputed
Non-C-Commanding Antecedent in Psych Predicate	Grammatical
Antecedent inside Embedded Clause	Grammatical
Split Antecedent in Plural	Grammatical
Discourse Antecedent	Grammatical
Logophoricity	Disputed
Generic/Arbitrary Reading	Grammatical
Inherent Reference/Deictic Use	Grammatical

Having laid out the general state of affairs in the *caki* as a reflexive/anaphor lines of research, I will now turn back to specifically deal with literature which has sought to reconcile the pronoun-versus-anaphor debate. These efforts generally take the shape of bringing the pronoun-like data in line with the anaphor analysis.

3.2.3 Refuting the Pronoun Arguments

Recalling the earlier discussion, there were four basic facts about *caki* which have been used to argue that it should be treated as a pronoun rather than an anaphor:

1. Anti-locality
2. No Blocking Effect
3. Split-antecedence

4. Discourse Binding

The first of these, the claim that *caki* exhibits anti-locality has already been addressed in the previous section, as it comes in direct conflict with the definition of *caki* as a reflexive. While I may argue against that particular definition, it cannot be denied that *caki* is felicitous in local contexts. While C. Lee does make a point of discussing this issue and noting that the local use of *caki* is contingent upon the type of predicate, it is worth noting that for both Cole et al. and Jayaseelan, two main proponents of anti-locality for *caki*, making the case for treating *caki* as a pronoun provides rhetorical support for their larger cross-linguistic claims. That is, their discussions of *caki* are in the context of broader analyses which predict *caki* should be a pronoun.

Specifically, for Cole et al., one of the main reasons for treating *caki* as a pronoun was that it did not exhibit blocking effects. Sohng (2003) provides a revised account of the covert head movement treatment of long-distance anaphors which accounts for both the lack of blocking effects, and the lack of strict subject orientation. First, unlike Cole et al., who show the long-distance anaphor moving only through Infl and C heads, Sohng has long-distance anaphors moving through all head positions up the clausal spine. Crucially, this leaves an A-chain through all V, Infl, and C heads. He further claims that *caki* already carries inherent *phi* features, and thus does not need to enter into a *phi*-checking relationship as it moves up the tree. In this way, he explains the lack of blocking effects for *caki*. To illustrate the different antecedent possibilities, (219) repeats the earlier (208):

(219) John$_i$-i Mary$_j$-eykey [Tom$_k$-i caki$_{i,?j,k}$-lul coaha-n-ta-ko]
John-NOM Mary-DAT Tom-NOM self-ACC like-PRES-DECL-COMP
malha-yess-ta.
say-PAST-DECL
'John told Mary that Tom likes self.' (Sohng 2003, ex 11a)

In this example, any of the c-commanding DPs can be an antecedent for *caki*. In the local antecedent case, Sohng proposes that *caki* will only raise as far as the embedded Infl head. For the case where the matrix subject is the antecedent, *caki* raises all the way to the matrix Infl, and only checks features once at the end of the derivation; there is no feature checking at the intermediate Infl head. For the non-subject antecedent, Sohng proposes a mechanism of chain binding, whereby the dative argument would locally c-command a link in the

movement chain of *caki* as it moves up to the matrix Infl. Movement to the matrix Infl must still take place even if the matrix subject is not the ultimate antecedent:

(220) Na$_i$-nun Mary$_j$-eykey [sensayngnim$_k$-i caki$_{*i,j,k}$-lul coaha-n-ta-ko]
John-NOM Mary-DAT Tom-NOM self-ACC like-PRES-DECL-COMP
malha-yess-ta.
say-PAST-DECL
'I told Mary that the teacher likes self.' (Sohng 2003, ex 11b)

In (220), with a first person subject, *caki* must still raise to the matrix Infl position so that there is a trace of *caki* in the matrix clause V head which the matrix dative can locally c-command. With this modified analysis, Sohng is able to bring *caki* (and *casin*) in line with Chinese *ziji* and Japanese *zibun*, capturing the lack of a blocking effect within Cole et al.'s general analysis.

The next major piece of evidence leading toward a pronominal analysis of *caki* is that when pluralised, *caki* takes split antecedents, as shown in this repetition of (197):

(221) Pierre$_i$-ka Marie$_j$-eykey caki-tul$_{i+j}$-uy sacin-ul poyocwu-ess-ta.
Pierre-NOM Marie-DAT self-PL photo-ACC show-PST-DECL
'Pierre showed Marie selves' photograph.' (Jayaseelan 1997, ex 46a)

This is a case of split antecedence in that the reference for *caki-tul* is taken to be Pierre and Marie. However, a different picture emerges in a parallel example from Cho (1996):

(222) John$_i$-i Mary$_j$-eykey caki-tul$_{i+j/i+k}$-eykwanhay iyaki-ha-yess-ta.
John-NOM Mary-DAT self-PL-about tell-do-PST-DECL
'John told Mary about selves.' (Cho 1996, ex 6a)

Here, Cho reports an ambiguity; treating the contribution of *-tul* as 'and others' which may include Mary, but does not necessarily have to.[3] Cho further goes on to show that *caki-tul* does not require a plural antecedent, in a structure which Madigan and Yamada (2006) label as "inclusive reference":

[3]From my own consultations, there is some speaker variance on this issue; while all would accept the *i+j* reading, for some speakers the *i+k* reading is not readily available. After some debate, even those for whom the reading is not immediately available will concede that it is possible.

(223) John$_i$-i caki-tul$_{i+k}$-i ik-yess-ta-ko malha-yess-ta.
John-NOM self-PL-NOM win-PST-DECL-COMP say-PST-DECL
'John said that selves won.' (Cho 1996, ex 7b)

Cho gives this example to show that *caki-tul* is acceptable with a singular antecedent. He argues that the data which seem to show split antecedence are merely cases parallel to (223) in which the "and others" contributed by *-tul* happens to contain the dative argument.

The last piece of evidence cited in favour of a pronominal analysis of *caki* is that it demonstrates discourse binding:

(224) Pierre$_i$-nun Marie-lul hosangha-ess-ta. kunye-nun caki$_i$-ka akki-ko
Pierre-TOP Marie-ACC think of-PST-DECL 3F-TOP self-NOM cherish-and
salang-ton sonye-i-ess-ta.
love-RELAT young girl-be-PST-DECL
'Pierre thought of Marie. She was the young girl self cherished and loved the most.' (Jayaseelan 1997, ex 47)

Again, repeating the example from earlier, this shows *caki* taking an antecedent from a previous sentence. A canonical example of this phenomenon originates in Yang (1982):

(225) Speaker A:

John$_i$-i salam-ul ponay-ess-ni?
John-NOM man-ACC send-PST-INT
'Did John send a man?'

Speaker B:

Ani, caki$_i$-ka cikcep o-ess-ta.
no self-NOM in person come-PST-DECL
'No, self came in person.' (Yang 1982, ex 68)

Here, the antecedent for *caki* in Speaker B's answer comes from the original question posed by Speaker A. The immediately preceding example in Yang's original paper shows what is ostensibly the same phenomenon:

(226) Speaker A:

John$_i$-i mues-ul kaciko-o-ess-ni?
John-NOM what-ACC bring-come-PST-INT

'What did John bring?'

Speaker B:

pro caki$_i$-uy chayk-man kaciko-o-ess-ta.
self-GEN book-only bring-come-PST-DECL
'*pro* brought only self's book.' (Yang 1982, ex 67)

In this example, *caki* is not in the subject position, and so Yang posits a *pro* subject which acts as a covert antecedent for *caki*, making (226) *not* a case of discourse binding. Possibly taking inspiration from this analysis, Park (1986) provides the following re-casting for Speaker B's portion of (225):

(227) Ani, [[$_T e_i$] caki$_i$-ka cikcep o-ess-ta].
no self-NOM in person come-PST-DECL
'No, self came in person.' (Park 1986, ex 48)

Park's addition is to posit a covert topic in the left periphery of Speaker B's answer. This topic fills the same role as Yang's *pro* from (226): providing a sentence-local antecedent for *caki*. Park motivates this by pointing out that topic marking *caki* in Speaker B's response is ungrammatical:

(228) Speaker A:

John$_i$-i salam-ul ponay-ess-ni?
John-NOM man-ACC send-PST-INT
'Did John send a man?'

Speaker B:

* Ani, caki$_i$-nun cikcep o-ess-ta.
no self-TOP in person come-PST-DECL
'No, self came in person.' (Park 1986, ex 50)

Park argues that the answer in (228) is ungrammatical because *caki* would be occupying the position which should be left open for the empty element which serves as the antecedent for *caki*. Thus, Park determines that Yang's example can be treated just as any other case of *caki* with a c-commanding antecedent. Gill (1999) later replicates this same analysis, also building from examples originating in Yang (1982).

Park goes on with this covert topic analysis to cover two more types of problematic example. The first of these is another case of what looks to be discourse binding:

(229) Speaker A:

John$_i$-un eti-lo ka-ni?
John-TOP where-DAT go-INT
'Where is John going?'

Speaker B:

caki$_i$-nun Chicago-lo ka-n-tae.
self-TOP Chicago-DAT go-PRES-RPT
'*e* says that self is going to Chicago.' (Park 1986, ex 51)

Here again, Park posits an *e* topic in the left periphery. He considers this to be an argument of the reportative particle *-tae*, which he treats as a higher predicate. This bi-clausal analysis is necessary, as *caki* here carries the *-nun* marking; Park's analysis is that this *-nun* is marking contrastive focus rather than topicality, and that it marks the left periphery of the embedded clause.

Park (unlike Gill) carries this analysis through to cases where it appears that *caki* has a non-c-commanding antecedent:

(230) [e_i caki$_i$-ka Mary-lul ttaeli-ess-ta-nun] sasil-i John$_i$-ul
self-NOM Mary-ACC hit-PST-DECL-ADMON fact-NOM John-ACC
koelophi-ess-ta.
worry-PST-DECL
'The fact that self hit Mary worried John.' (Park 1986, ex 44)

Recall that when discussed earlier, this sort of example could only be brought in line with an anaphor analysis of *caki* if either an underlying structure could be posited which had the antecedent c-commanding *caki* or an LF movement could be motivated. This is not that far-fetched, as this is a psych predicate example, the exact environment for which C. Lee originally proposed the analysis in which the experiencer originated in a higher position. Here, Park provides a simpler solution, a topic within the embedded clause, which he describes as being controlled by *John*, though no control mechanism is described. Furthermore, it is worth noting that Park is working under the assumption that all instances of *caki* must be A-bound, and thus describes the Korean topic position as an A-position which is present for all clauses, explaining the proliferation of topics. Park's analysis dovetails nicely with that of S.-Y. Kim (2000), whose hierarchy predicted that topics are the ideal antecedents for *caki*. Another update yields the picture in Table 3.3.

Table 3.3: Distribution of *Caki*: Third Summary

Environment	Status
Local Antecedent	Disputed
Non-Local Subject Antecedent	Grammatical
Non-Local Non-Subject Antecedent	Disputed
Non-C-Commanding Antecedent in Psych Predicate	Grammatical
No Antecedent in Reportative Context	Grammatical
Antecedent inside Embedded Clause	Grammatical
Split Antecedent in Plural	Ambiguous
Discourse Antecedent	Grammatical
Logophoricity	Disputed
Generic/Arbitrary Reading	Grammatical
Inherent Reference/Deictic Use	Grammatical

As a result the analyses of C. Lee and Park, there are not many environments left which challenge an analysis of *caki* as an anaphor. The only real challenges to this analysis are the cases where the antecedent of *caki* is in an embedded clause (not in the context of a psych predicate), and the generic and quasi-referential uses introduced by Sohng. If, as with English *self*-pronouns, there is allowance for exempt referential uses of *caki*, then Sohng's examples are also not a problem. Still, there is the matter which led into the discussion of *caki* in the first place: *caki-casin*. Why is it then that *caki* can be grouped among the pronouns in taking *-casin* as a suffix to yield an anaphoric form? A possible answer to this can be found in the examination of a more subtle thread which runs through the discussion of *caki*: its use as a bound variable.

3.2.4 Bound Variable *caki*

In this subsection, I begin with literature which hints at a bound variable analysis for *caki*, then move into argumentation drawn from Storoshenko (2007). One of the earliest references to bound variables in connection with Korean anaphors arises in I.-H. Lee (1978):

(231) a. Yengswu$_i$-man casin$_i$-eykey thwuphyo-ha-yess-ta.
Yengswu-only self-DAT vote-do-PST-DECL
'Only Yengswu voted for himself.' (I.-H. Lee 1978, ex 68b)

b. Nwukuna$_i$ [casin$_i$-ka pwucatoy ki]-lul wenha-n-ta.
everyone self-NOM become rich-ACC want-PRES-DECL
'Everyone wants that self becomes rich.' (I.-H. Lee 1978, ex 69b)

While these examples contain *casin*, they are equally acceptable with *caki*. Lee describes these as problematic for his transformational analysis, and suggests that instead of a system based on replacing co-referential forms, a bound variable analysis might be the best way to treat these data. Unfortunately, he does not elaborate at all on exactly how these examples are problematic, coming as they do toward the end of his paper.

Y.-S. Kang brings forward an account which treats *caki* as a resumptive pronoun in relative clause and topic contexts:

(232) a. [kangto$_j$-ka caki$_i$-uy namphyen-ul khal-lo ccillu-e cwuki-n] ku
robber-NOM self-GEN husband-ACC knife-with stabbing kill-REL that
yeca$_i$
woman
'That woman who a robber killed self's husband by stabbing with a knife.'
(Y.-S. Kang 1986, ex 1a)

b. Ku yeca$_i$-nun kangto$_j$-ka caki$_i$-uy namphyen-ul khal-lo ccill-e
that woman-TOP robber-NOM self-GEN husband-ACC knife-with stabbing
cwuki-ess-ta.
kill-PST-DECL
'As for that woman, a robber killed self's husband by stabbing with a knife.'
(Y.-S. Kang 1986, ex 1b)

First, recalling Chapter One, Déchaine and Wiltschko report appearance as a resumptive pronoun as a test for bound variable anaphora. In light of this, it is not surprising that Y.-S. Kang treats both of these as a result of an A′ movement: *ku yeca* is shown in a displaced position for both examples in (232), and, contra Park, Y.-S. Kang treats topics as being derived non-arguments. However, because what he takes to be the originating positions of *ku yeca* are occupied by *caki*, Y.-S. Kang concludes that *caki* in these cases is a resumptive pronoun. He further specifies this to mean that such instances of *caki* should be treated as base-generated bound variables, rather than as anaphors.

Furthermore, data is presented which indicate that there are contexts in which only *caki* can serve as a bound variable:

(233) a. * [ku$_i$-ka sihem-ey silphayha-n] John
3M-NOM test-DAT fail-REL John
'John who he failed in the exam' (Y.-S. Kang 1986, ex 27)

b. [caki$_i$-ka sihem-ey silphayha-n] John
3M-NOM test-DAT fail-REL John
'John who self failed in the exam' (Y.-S. Kang 1986, ex 4)

Again, working under the assumption that these relative clauses represent a bound variable structure, Y.-S. Kang observes that the pronoun *ku* cannot be used in these contexts, only *caki*.

Similar claims are found in Tajima (1987), who reports a complete lack of bound variable interpretations for pronouns in Korean:

(234) a. Nwu$_i$-ga caki$_i$ emeni-lul salang-ha-n-ya?
who-NOM self mother-ACC love-do-PRES-INT
'Who loves his mother?' (Tajima 1987, ex 2b)

b. Motun sonye$_i$-ka [caki$_i$-nun yeph-i-ta-ko] mit-nun-ta.
every girl-NOM self-TOP pretty-be-DECL-COMP believe-PRES-DECL
'Every girl believes that she is pretty.' (Tajima 1987, ex 3b)

According to Tajima, *caki* in (234) cannot be replaced with a pronoun if the bound variable reading is to be maintained. For further evidence that *caki* is a variable rather than a pronoun, Tajima claims that pronouns in Korean do not induce a weak crossover effect. Finally, paralleling Park and Gill, Tajima also claims that discourse binding of *caki* can be reduced to local A′ binding of an empty topic operator.

Only slightly different is the account of B.-M. Kang, who limits his comment to saying that *ku* does not allow for a bound variable reading as readily as *caki*:

(235) Kak salam$_i$-i [caki$_i$/?ku$_i$-ka chencay-i-ta-ko] saygkak-ha-n-ta.
each person-NOM self/3M-NOM genius-be-DECL-COMP think-do-PRES-DECL
'Each person thinks that self/he is a genius.' (B.-M. Kang 1988, ex 14)

This is problematic for B.-M. Kang, trying to construct an argument that in Korean (and Japanese) it is the anaphors which do "double duty" acting as bound variables in addition to

their canonical usage, rather than the set of referential pronouns, as is the case in English. He goes on to note that in some contexts a pronoun is the best choice for a bound variable reading:

(236) Etten kkoch$_i$-itunci kukes$_i$-uy alumtawum ttaymwune salangpat-n-ta.
any flower-PART it-GEN beauty beacuse of be loved-PRES-DECL
'Any flower is loved because of its beauty.' (B.-M. Kang 1988, ex 16)

Kang notes that *caki* cannot be used in this context, as the intended antecedent is inanimate. Instead, he gives the pronominal *kekes* as the best potential bound variable. He makes no mention of *cache*, which is cited as an inanimate equivalent to *caki* (Moon, 1995).[4] Kang provides no solution for this apparent contradiction which allows both anaphors and pronominals to act as bound variables; in actuality this is not all that different from English, in which reflexives must be used in lieu of bound variable pronouns in co-argument contexts.

Moon (1995) is also among those who note that *caki* can function as a bound variable, with examples similar to those given by Tajima. The same ground is also covered by Li and Takahashi (1995), who adopt terminology from Katada (1991), describing *caki* as an "operator anaphor". As originally defined by Katada with reference to Japanese *zibun*, this analysis would have *caki* undergoing LF A′ movement into a left-peripheral operator position. This is in some sense the exact opposite analysis (treating *caki* as an operator rather than a variable), but I mention it here because it falls in line with the general theme of connecting *caki* to the A′ operator-variable system.

An final contribution to this line of inquiry comes from Cho (1996). The main thrust of Cho's work is to argue against the *caki*-as-pronoun analysis. He takes on the issues of purported anti-locality and split antecedence, along the lines previously discussed. He also mobilises some standard binding tests including VP ellipsis:

(237) John$_i$-i caki$_i$-lul kwasinha-yess-ko, Mary-to kuleha-yess-ta.
John-NOM self-ACC overtrust-PST-CONJ Mary-also do-PST-DECL
'John overtrusted self, and Mary did too.'
(= Mary overtrusted herself)
(≠ Mary overtrusted John) (Cho 1996, ex 19a)

[4]Because there is very little material on *cache*, and it all aligns with uses of *caki* (except for the animacy constraint), I will not be providing a separate discussion of that form.

Cho reports that only the sloppy reading is available in the second conjunct, indicating that *caki* is functioning as a bound rather than a referential form. However, he concludes by arguing exactly against the treatment of *caki* as a bound variable, on the basis of very simple cases such as (238):

(238) Nwukuna$_i$-ka caki$_i$-lul sinloy-ha-n-ta.
everyone-NOM caki-ACC trust-do-PRES-DECL
'Everyone trusts himself.' (Cho 1996, ex 24a)

This is from the data which Cho uses to argue that *caki* is not subject to anti-locality. He chooses a quantifier example specifically because *ku* cannot have a bound variable interpretation within its own clause; this is a situation in which pronouns are not admissible. On the basis of the fact that *caki* has a bound variable interpretation in this local environment, he concludes that *caki* is not a bound variable. His reasoning is that in most languages, as in English, bound variable forms are inadmissible in reflexive environments.

Whether this is a sound basis for rejecting the notion that *caki* is a bound variable is uncertain. First of all, Cho does not mention whether or not *ku-casin* or *caki-casin* would be preferable to *caki* in (238). Cho suggests this is possible, mentioning the existence of a short-distance reflexive which can be used in such contexts, but he does not specify which one. Given that he makes this reference for Korean after mentioning the Japanese *zibun-zisin*, one can speculate that he means *caki-casin*, but that is never made clear. If *caki* is the Korean bound variable, it could simply be the case that the expected anti-locality constraint is weaker than that for English or Japanese. This notion that *caki* is anti-local has definitely persisted through the literature, but according to Cho, it seems that an anti-locality constraint on *caki* can be used to *support* the notion that *caki* should be treated as a bound variable.

Based on the picture emerging here, there is general agreement that the referential pronouns in Korean can function as bound variables non-locally. As indeed can *caki*. The difference is that *caki* is also acceptable in local contexts, though it is not the only option for local variable binding, as *caki-casin* is also available. In light of this, a bound-variable option appears to be the best fit for *caki*. What separates *caki* from the pronouns is that *caki* does not readily allow for unbound, referential uses. Separating *caki* from *casin* is the fact that *caki* cannot be employed as a reflexive marker, suffixed onto pronouns or proper names. In the Déchaine and Wiltschko typology, this places *caki* as a ϕ^0 head.

As hinted by Cho, a bound variable analysis of *caki* is not incompatible with an anti-locality constraint. The other arguments used to classify *caki* as a pronoun can also be easily brought in line with a bound variable analysis. Cho's own discussion of the split antecedence cases works equally well for bound variable treatments of *caki*, as the apparent split antecedence reading is a result of the interpretation of the plural morpheme *-tul*; *caki* would still have one singular antecedent. As separately proposed by Park, Tajima, and Gill, discourse binding cases can be reduced to local binding if an empty operator is posited; Park further goes on to suggest that *caki* in reportative contexts can fall under the same analysis. Finally, the lack of a blocking effect for *caki* could be accounted for in that operator-variable binding structures can function across clause boundaries without the need for any LF movement mechanism of the variable. Under this analysis, *caki* would never enter into the long-distance head-movement operation. As such, Sohng's adaptation of the Cole et al. analysis is not necessary for *caki*, though it may still be needed to account for non-subject antecedents of long-distance monomorphemic *casin*.

The examples showing that *caki* does not show any sign of the blocking phenomenon were constructed using simple proper name antecedents, not quantifiers or *Wh*-elements, to which the bound variable arguments have been generally restricted. To bring these in line with a bound variable analysis of *caki*, is simply a matter of employing a generalised quantifier analysis of proper names along the lines of Barwise and Cooper (1981). With this change in place, all uses of *caki* can be brought under the umbrella of a bound-variable analysis:

(239) a. Motwu$_i$-ka [John$_j$-i caki$_i$-lul salang-ha-n-tako]
everyone-NOM John-NOM self-ACC love-do-PRES-COMP
sayngkak-ha-n-ta.
think-do-PRES-DECL
'Everyone thinks John loves self.'

b. Everyone λx[x thinks [John loves x]]
$= \lambda P.\forall y$[person (y)] [$P(y)$](λx[x thinks [John loves x]])
$= \forall y$[person (y)][y thinks [John loves y]]

(240) a. John$_i$-i [caki$_i$-ka iki-ess-tako] mal-ha-yess-ta.
John-NOM self-NOM win-PAST-COMP say-do-PAST-DECL
'John said self won.'

b. John λx [x said [x won]]
= $\lambda P.[P(j)](\lambda x$ [x said [x won]])
= j said [j won]

(239) presents a semantic representation of long-distance binding of *caki* by a quantifier. In (240), the same binding mechanism is illustrated with *John* represented as a generalised quantifier. These examples show that variable binding semantics can be applied to cases where *caki* has a proper name antecedent.

Under this bound variable analysis, a c-command constraint still obtains though: the antecedent for *caki* must c-command the variable. Possible counterexamples to this are presented in (241) and (242):

(241) Motwu$_i$-uy sinpal-un caki$_i$-uy pal-pota hwelssin kuta.
everyone-GEN shoes-TOP self-GEN foot-than a lot big
'Everyone's shoes are a lot bigger than self's feet.'

(242) Suni$_i$-uy sinpal-un caki$_i$-uy pal-pota hwelssin kuta.
Suni-GEN shoes-TOP self-GEN foot-than a lot big
'Suni's shoes are a lot bigger than self's feet.'

In (241), a bound variable reading of *caki* obtains. This suggests that there is a mechanism of quantifier raising which allows *motwu* to escape its position embedded within the subject DP and provide an antecedent for *caki*. Again, treating proper names as quantifiers, the example in (242) can be similarly treated. The more problematic examples such as (213), repeated below, where *caki* was shown to be able to take an antecedent remain a challenge:

(243) [John$_i$-i [Sam$_j$-i ssu-n kisa-lul] swuciphay-no-atta-nun]
John-NOM Sam-NOM write-ADNOM article-ACC keep-?-PST-PST-ADNOM
sasil-i caki$_{i,j}$-uy tongsayng-ul kippukeyha-yess-ta.
fact-NOM self-GEN brother-ACC please-PST-DECL
'The fact that John kept the article that Sam wrote pleased self's brother.'
(O'Grady 1987, ex 54)

As previously mentioned, O'Grady notes that the acceptability of such examples deteriorates with the depth of embedding, but he says that either of the embedded clause nominals may serve as the antecedent for *caki*. Quantifier Raising will not help here, as any raising of *John* or *Sam* will be confined to the embedded clause and should not escape high

enough to a position from which *caki* can be bound. Furthermore, an empty topic operator in a position which could bind *caki* would also c-command the embedded clause, and co-indexation with *John* or *Sam* would yield a Condition C violation. According to my consultants, judgements on sentences such as (243) are mixed; for most speakers there is no available reading where *caki* takes either of the embedded clause DPs as an antecedent; at best the inherent reference (second person) reading of *caki* is available.

Caki is subject to other constraints on bound variables, crossover for example:

(244) * caki$_i$-uy emma-ka Mary$_i$-lul piphan-ha-yess-ta.
self-GEN mother-NOM Mary-ACC criticise-do-PST-DECL
'Self's mother criticised Mary.'

Under a generalised quantifier analysis, it should be possible for *Mary* to undergo quantifier raising to a position where it can bind *caki* in (244). Such a movement would violate weak crossover, in that *Mary* does not originate in a position where it binds *caki*. Park's treatment of backward anaphora where *caki* is embedded in a subordinate clause preceding the antecedent can avoid crossover; an empty topic operator within the embedded clause provides the necessary antecedent for *caki* without triggering any Condition C effect (it doesn't c-command the co-indexed antecedent), and without any need to posit a crossover-inducing movement.

Along the same lines at the VP-ellipsis cases, *caki* also demonstrates sloppy readings in focus particle contexts:

(245) Na-nun Minswu-eykey-man caki-uy-chimsil-ey issu-lako malhanta.
I-TOP Minswu-DAT-only self-GEN-room-DAT stay-COMP said
'I said that only Minswu must stay in self's room.'

Here, the first person subject forces *caki* to take the dative *Minswu* as its antecedent. The strict interpretation, where alternatives to Minswu were not told to stay in Minswu's room, is not reported to be available in (245). Rather, the primary reading that emerges is that Minswu was the only one confined to his own room. A secondary meaning where *caki* has the inherent reference second person usage is also reported, but this is expected; while it is true that there are contexts in which this reading is the only one available, it is not the case that the inherent reference reading is limited to last-resort contexts. Rather, that reading should always be available, but heavily context-dependent.

Table 3.4: Distribution of *Caki*: Fourth Summary

Environment	Status
Local Antecedent	Disputed
Non-Local Subject Antecedent	Grammatical
Non-Local Non-Subject Antecedent	Disputed
Non-C-Commanding Antecedent in Psych Predicate	Grammatical
No Antecedent in Reportative Context	Grammatical
Antecedent inside Embedded Clause	Disputed
Antecedent inside Genitive DP	Grammatical
Split Antecedent in Plural	Ambiguous
Discourse Antecedent	Grammatical
Logophoricity	Disputed
Generic/Arbitrary Reading	Grammatical
Inherent Reference/Deictic Reading	Grammatical
Resumptive Pronoun Reading	Grammatical
Local Bound Variable Reading	Grammatical
Non-Local Bound Variable Reading	Grammatical
Strict Reading in Ellipsis/Focus	Ungrammatical
Crossover Structures	Ungrammatical

The bound variable data substantially changes the overall picture for the distribution of *caki*, shown in Table 3.4. The distribution of *caki* now includes most standard diagnostics for bound variable anaphora, and most of the prior facts can be reconciled with this analysis. C. Lee's treatment for psych predicates is not incompatible with this analysis, as he treats the derived word order, where the antecedent does not c-command *caki* as derived; it is not the case that he posits a crossover-inducing movement. Similarly, Park's covert topics can be re-analysed as A′ binders, and a covert binder in reportatives is also compatible with a bound variable analysis. The issue of embedded antecedents must be addressed though. *Caki* is shown to have quantificational antecedents embedded within DPs, suggesting that quantifier raising may be an important component of the *caki* resolution system. QR will

not rescue the embedded clause antecedents though, but there is a lack of consensus on the judgements. In the next section, I turn to a corpus analysis of *caki* which was undertaken to determine whether a large sampling of texts will reflect the described usage.

3.3 *Caki* in Corpus

Previous corpus research on *caki* has been limited to determining its status as local or long-distance, its frequency relative to *kucasin* and *cakicasin*, and the grammatical roles of those forms. This study, from B.-M. Kang (2001), makes use of the 10 million word KOREA-1 corpus. The corpus covers a range of fiction and non-fiction, across literary and commercial genres. Interestingly, only 12% of the corpus is from spoken sources. Out of the whole corpus, Kang finds relatively similar counts for *caki* and *casin*, 10,005 occurrences versus 10,601 occurrences, respectively, and a much lower count for *caki-casin* at 508 occurrences. Furthermore, both *caki* and *casin* showed the same rankings for their top three most frequent uses: genitive > nominative > accusative. In terms of these raw numbers, *caki* and *casin* are quite similar.

This similarity disappears in a comparison of the accusative forms though. In a more focused look at half of the corpus, Kang finds 316 instances of *caki-lul*. Interestingly, only just over half of these turn out to be cases where *caki*'s antecedent is in a higher clause. The results of this part of B.-M. Kang's study are presented in Table 3.5.

Table 3.5: Antecedents of accusative-marked forms in KOREA-1 Corpus (B.-M. Kang 2001)

	caki-lul	*casin-ul*	*caki-casin-ul*
local	151	311	66
long-distance	165	123	5
total	316	434	71

Caki-lul was the only one of the three forms to occur in a majority of long-distance contexts, but based on these numbers, Kang quite reasonably states that *caki-lul* can take local and long-distance antecedents equally-well. Conversely, *caki-casin-ul* is almost ex-

clusively local. It is *casin-ul* that stands out the most in this set of data. While the number of long-distance antecedents is too high to be a result of limited contexts or an alternate reading (like the few long-distance instances of *caki-casin-ul* might well be), *casin-ul* is used locally much more often than in long-distance cases. Kang does not speculate on the reason for this difference, but he notes that it counters any claim that *caki-lul* and *casin-ul* can be used interchangeably. One possibility might arise from the fact that *casin-ul* is subject to the blocking effect for long distance anaphors; long distance *casin-ul* can essentially only arise when all clausal subjects in a sentence match for person features. For the local cases only, Kang compares the theta roles of the antecedents. Across the board, agent antecedents are more frequent than experiencer antecedents, but as a proportion of the total local usage, *casin* is used with experiencer antecedents more than *caki*. This seems to run counter to C. Lee's claim that local *caki* is less acceptable with agentive predicates rather than predicates of mental states.

In terms of adding to the discussion on the nature of *caki*, probably the most interesting conclusion from this study would be the observation that accusative *caki-lul* is almost equally distributed between local and long distance cases. This quite effectively shows that any claim that *caki* is strictly anti-local is shaky at best. Combining these results with those for *casin* suggests that while their overall distributions are similar, there does appear to be a difference in the way they function across long distances. Furthermore, the difference is exactly what one would expect of a situation where *caki* has a relatively unrestrained method of binding long distance while *casin* requires agreement between multiple elements. Assuming Cole et al. are right in that *casin* patterns as a "typical" long distance anaphor, these numbers suggest that *caki* is indeed something else. While Cho found local instances of *caki* to be equally problematic for a bound variable analysis, that analysis still seems to be the one which best fits the facts.

Next, I report on a corpus study presented in Han and Storoshenko (2009) which specifically examined *caki* in terms of this bound variable analysis. Specifically, we are interested in what kinds of antecedents *caki* can take, and whether they can be brought in line with the claim that *caki* is a bound variable.

The corpus used was the Sejong Colloquial Corpus, published by the National Korean Language Institute and the Department of Tourism and Culture in Korea. In all, this corpus amounts to 550,000 words, collected from transcribed recordings of radio and TV inter-

views, as well as various news and entertainment programs. As with the English corpus, this began with an automatic extraction process, which identified 675 instances of *caki*. Of these, 20 were dismissed as unanalysable because they were contained in either incomplete or garbled contexts, or contained so many speech errors as to make it impossible to reconstruct the intended sentence. This leaves 655 tokens. The first major breakdown involves the ϕ features of the antecedent: third person *caki* is of particular interest, as these should be the cases of binding, second person uses being instances of inherent reference. The breakdown is shown in Table 3.6. As shown, the vast majority of cases contained a third-person *caki*. Second person instances are infrequent, but not unexpected; what is surprising is that first-person antecedents are reported at all, as these are generally considered impossible, except in the focus structure of (245).

Table 3.6: Person Features of *caki* in Corpus

	Number	Percent
1^{st} Person	9	1.4%
2^{nd} Person	23	3.5%
3^{rd} Person	623	95.1%

Looking solely at the third person cases, the next major consideration is to determine what proportion of these can be treated under a bound variable analysis. The first grouping is to break the instances of *caki* down into cases where there is a c-commanding antecedent, a non-c-commanding antecedent, or no antecedent. The figures are shown in Table 3.7.

Table 3.7: Antecedents of 3^{rd} Person *caki* in Corpus

	Number	Percent
C-command	497	79.7%
No c-command	26	4.2%
No antecedent	100	16.1%

Taken together, these findings clearly indicate that the majority of cases fall directly

into the bound-variable analysis. Third person uses represent 95% of the total uses of *caki*, and close to 80% of those have a c-commanding antecedent. This already bodes well for the bound variable analysis, but it still remains to be seen whether the remaining 20% can be accounted for.

3.3.1 Non C-Commanding Antecedents

The examples with a non-c-commanding antecedent for *caki* can be broken down into three different categories, shown in Table 3.8.

Table 3.8: *caki* with Non-C-Commanding Antecedents

	Number
Topic antecedent	22
Genitive antecedent	2
Displaced *caki*	2

As shown, the majority of these are cases where the antecedent for *caki* is topical, represented by either a *pro* or a topic-marked DP. Often, these would occur in situations where the antecedent appears in a topic-introducing phrase. A first example of this is presented in (246):

(246) Yoshida$_i$ yayki-incuk, caki$_i$-ka ilcey ttay cwungkwuk
Yoshida tell-according to self-NOM Japanese occupation time China
pongchen-uy chongyengsa-lo iss-ess-nuntey
Pongchen-GEN consulate general-as exist-PAST-and
'According to Yoshida$_i$, during the Japanese occupation self$_i$ was the consulate general of Pongchen in China and' [23;;013.txt]

This example can be seen as a parallel to the original Yang example of discourse-bound *caki*, in that *caki* appears in the subject position with nominative case marking, suggesting that there is an open position for a topicalised *nun*-marked element. Here, the familiar empty operator can be posited to provide a c-commanding antecedent for *caki*. To avoid

a potential Condition C violation though, this operator would need to be lower than the phrase containing the antecedent *Yoshida*.

A more straightforward case is shown in (247):

(247) Kyengsam$_i$-un mayil pap-ul cie talak-ey nehe cwuko-n caki$_i$-nun
Kyengsam-TOP everyday meal-ACC make attic-at put give-and self-TOP
cwung-ul capule tany-ess-ta.
monk-ACC catch go-PAST-DECL
'Kyengsam$_i$ made a meal and put it in the attic every day and self$_i$ went around to catch the monk.' [19;;008.txt]

This is a straightforward coordination structure, with *caki*'s antecedent found in a prior conjunct. The topic operator analysis more easily avoids the potential Condition C issue, as the empty operator would only have scope over its local conjunct. Problematic here though is the fact that *caki* already carries the *-nun* marking, suggesting there is no room here to posit a c-commanding topic operator. This can be resolved though, under the assumption that operators exist for more than just topics. Recalling the empathy discussion from J.-M. Yoon, in the context of a verb like 'go', speaker empathy should be with the agent, here making *Kyengsam* the pivot. If there is an empty operator corresponding to speaker empathy or to this notion of pivot, then it would be that operator which could antecede *caki*, not necessarily the topic.

There were also two cases where *caki*'s antecedent was embedded as the possessor in a genitive DP. These are analogous to the cases in (241) and (242) where it was mentioned that quantifiers can bind *caki* from such an environment, suggesting that an LF movement operation is available:

(248) Ku$_i$-uy chilyopep-un kuce caki$_i$-ka sangkakna-nun tay-lo
he-GEN treatment-TOP always self-NOM think-ADNOM in accordance to
chilyoha-nun saylo-wun pangpep-i-ci.
treat-ADNOM new-ADNOM method-COP-DECL
'His$_i$ treatment method is a new method where self$_i$ always treats in accordance to his thoughts.' [18;;008.txt]

Because this is exactly parallel to the cases already discussed, there is nothing new to add here.

The final two cases involve instances where a larger constituent containing *caki* has been displaced:

(249) [Emma-ka caki$_i$ nwui tongsayng-ul com aniolsita ha-kwu sayngkakha-n
mom-NOM self sister younger-ACC a little disapprove do-COMP think-ADNOM
kel]$_j$ yay$_i$-ka elin nai-ey t$_j$ nwunchi-lul chay-ss-nunci
fact-ACC kid-NOM young age-at t$_j$ aware of become-PAST-DECL
'The kid$_i$ became aware of the fact that his Mom was thinking a little disapprovingly of self's$_i$ younger sister' [514.3;;180.txt]

Here, *caki* is embedded within the complex DP direct object which has been scrambled to the sentence initial position. In its canonical position, the clause containing *caki* would be c-commanded by the antecedent *yay* 'kid'.

In sum then, all the cases of *caki* with a non-c-commanding antecedent can be made compatible with the overall bound variable analysis. Furthermore, there were no instances similar to O'Grady's example, where the antecedent for *caki* was deeply embedded within a subject clause; the examples found all dealt with the antecedent for *caki* originating in a coordinated clause or a topic-introducing phrase, giving more room for the topic operator analysis than in O'Grady's example.

3.3.2 Antecedentless *Caki*

The breakdown of Antecedentless *caki* is given in Table 3.9. Taking each type in turn, it will be shown that these too are compatible with the bound-variable analysis of *caki*.

Table 3.9: *caki* with No Antecedent

	Number
Discourse topic	34
Generic	25
Sentence fragment	21
Reportative particle	13
Compound noun	7

The most frequent use of antecedentless *caki* was in contexts where prior discourse easily provides an antecedent. These can be treated exactly according to the empty topic analysis proposed by Park:

(250) Speaker A:

Park Tongsil sensayng <Yelsaka>-nun nwuka ...?
Park Tongsil teacher Yelsaka-Top who ...
'Who (composed) Yelsaka that teacher Park Tongsil (sang)?' Speaker B:

Caki-ka mantul-ess-ciyo.
self-NOM make-PAST-DECL
'Self made it.' [348;;118.txt]

Exactly as in the original example from Yang, this is a case where the antecedent for *caki* (Park Tongsil) is contained within a prior *Wh*-question. Again, *caki* carries nominative case marking, leaving open the position for the topic operator.

The next most frequent use of antecedentless *caki* is in generic contexts, again paralleling previously-discussed examples:

(251) a. Caki swukcey-nun caki-ka ha-nun ke-ya.
self homework-TOP self-NOM do-ADNOM Fut-DECL
'In general, self should do self's homework.' [111;;053.txt]

b. Caki mom-un caki-ka cikye-ya toy-nuntey.
self body-TOP self-NOM keep-COMP must-DECL
'In general, self should keep self's body.' [219;;089.txt]

These cases have a similar flavour to the proverb examples from Zubin et al. With these, there are two possible ways to proceed. One would be to actually classify them as second person, interpreting them as direct instructions to the addressee, but this would of course strip them of their generic interpretation. To bring these under a binding analysis would require positing yet another operator, this one representing a generic referent which could serve as the antecedent for *caki*. It's worth noting though that this is not an issue limited to *caki* or Korean. English *oneself* shows up with this same generic reading and no overt antecedent:

(252) a. I always pass on good advice. It is the only thing to do with it. It is never of any use to oneself. -Oscar Wilde

b. Invention, in my opinion, arises directly from idleness, possibly also from laziness - to save oneself trouble. -Agatha Christie

c. I don't watch television, I think it destroys the art of talking about oneself. -Stephen Fry

For these cases as well as the generic uses of *caki*, a covert operator representing a generic individual, can provide the necessary antecedent.

The next two types of antecedentless *caki* deal with reported speech. In (253), the speaker is talking about a letter he had received from a student:

(253) Caki-nun yeksa sikan-i cham silh-ess-ta.
self-TOP history time-NOM very dislike-PAST-DECL
'(The student said) self disliked history very much.' [210;;088.txt]

In this sentence fragment, there is a missing matrix clause in which the antecedent student would be introduced. If this elided material is recovered, an antecedent for *caki* is readily apparent.

As previously discussed, *caki* also emerges in cases where the reportative evidential is used:

(254) a. Enceyna caki-ka mac-tay.
always self-NOM correct-RPT
'(Steve said) self is always correct.' [8;;007.txt]

b. Caki-nun PD-ka toy-ko sip-tay-yo.
self-TOP PD-NOM become-AUX want-RPT-HONOR
'(Swuyen said) self wants to become a PD.' [198;;086.txt]

For both of these cases, the surrounding context provides the antecedent. Following Park's analysis of such cases, the reportative can introduce an implicit speaker argument, binding *caki*. Thus, in cases where the act of speaking is either totally elided, or indicated only through a reportative, mechanisms can be found to provide antecedents for antecedentless *caki*.

The final cases of antecedentless *caki* deal with cases where *caki* is incorporated in a compound noun:

(255) a. Kuken caki pyenmyeng-i-ko wiysen-i-ya.
that-TOP self excuse-COP-and hypocrisy-COP-DECL
'That is self-excuse and hypocrisy.' [78;;044.txt]

b. Yocum-un caki phial sitay-eyyo.
these days-TOP self promotion age-DECL
'These days, it is the age of self-promotion.' [425;;133.txt]

These cases can either be treated similarly to the generics, or they can be discounted from the bound-variable analysis by being treated as frozen expressions which are exempt from any need for an antecedent.

The end result of the corpus study is that all of the instances of third-person *caki* can be accommodated under a bound variable analysis for *caki*. What is stunning in taking the corpus results as a whole is that they a fairly representative cross section of the kinds of data present in Table 3.4. While the vast majority of cases involved simple binding from a c-commanding antecedent, discourse binding, binding from within a genitive, generic readings, and reportative contexts all emerged. One exception is that in the non-c-commanding cases, there was no evidence of *caki* being bound from elements embedded within a subject clause. Given the diversity of data which was found in the corpus, the absence of such examples combined with the conflicting judgements in the original source and among consultants all suggest that such examples may not pose a threat to the bound variable analysis. Having found this further support for the bound variable analysis of *caki*, in the last section of this chapter, I turn my attention to an experiment which comes at the *caki*-as-pronoun analysis from a new direction.

3.4 Testing the Context-Sensitivity of *Caki*

If indeed *caki* is a pronoun, then it should show one more pronoun-like behaviour in addition to the ones which have been discussed so far: it should be context-sensitive. True referential pronouns are sensitive to their surrounding discourse context, and their reference in a potentially ambiguous context can be influenced by the relative discourse salience of the potential antecedents. Thus, an experiment was designed which compared *caki* with the referential pronouns *ku* and *kunye*. This experiment, reported in Han et al. (in press), sets out to do two things: first, to test whether or not context can have an impact on *ku* and

kunye, establishing that contextual manipulation can systematically impact pronominal reference. Secondly, testing whether or not this same manipulation can have an impact upon *caki*.

3.4.1 Methods and Materials

Inspired by the success of the Runner and Kaiser experiments of binding phenomena in English, it was decided that an eye-tracking experiment would be the ideal way to proceed in Korean. Because there is such a variety in the reported judgements in the literature, eye-tracking would be an ideal method for determining whether there is a difference between initial processing reactions to the experimental stimuli and the considered judgements of participants. If, as one might expect given the state of the literature, there would be considerable variation in the judgements provided by native speakers, then eye-tracking would provide a second channel of information which would not be filtered by the conscious considerations of the speakers.

In this experiment, participants see a static visual display while a recorded Korean narrative is played. The audio-visual component, along with all input from the participants is managed using E-Prime experiment design software. A sample visual display is given in Figure 3.1. The display shows two characters, always a male and a female, with the images counterbalanced through the experiment in terms of the position (left vs right) of the male and female characters. In the centre of each display is an item which serves to anchor the setting; in the sample image, a blackboard sets the context for two characters in school uniforms. For the experimental trials, sixteen different displays were produced, each with different characters, in various costumes and settings. Other images included scenes such as two characters in athletic clothes with a treadmill signifying a gym, two characters in outerwear near a tree signifying a walk in the park, and two characters in casual wear beside a stove, setting the stage as at home in the kitchen. Each of the sixteen images was used in multiple trials, details of which are described below.

In total, each trial consists of a five sentence script presented aurally while the image is on screen. The first two sentences consist of an introduction. For the school scenario, the introduction is as in (256):

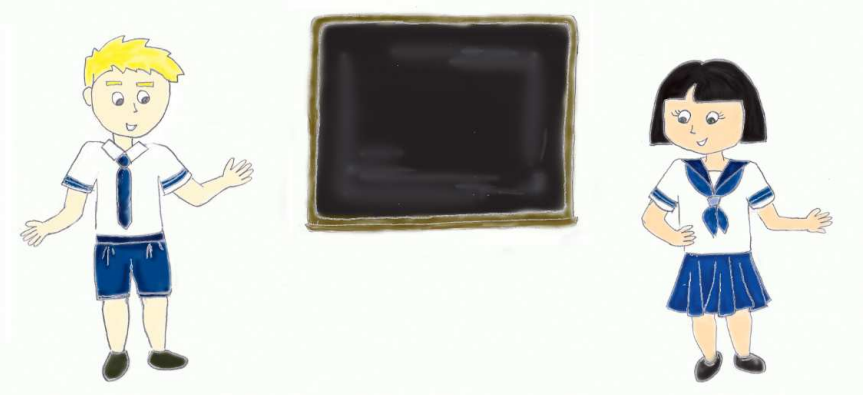


Figure 3.1: Korean Eyetracking Experiment Sample Display

(256) Jongwu-wa Yuli-ka kyosil-ey iss-ta. Jongwu-wa Yuli-nun
Jongwu-and Yuli-NOM classroom-DAT be-DECL Jongwu-and Yuli-TOP
pangkum shiem-ul chi-ess-ta.
just test-ACC take-PST-DECL
'Jongwu and Yuli are in their classroom. Jongwu and Yuli just took a test.'

First and most importantly, these sentences establish the names of the two characters, stating them twice, always in left-right order with respect to the image. The setting is also stated, reinforcing what is already shown on screen. Finally, a situation is introduced. These introductory sentences are kept constant across all trials using a given image.

Following this two sentence introduction comes a second set of two sentences which comprise the first independent variable, the contextual bias. In a given trial, only one of the three sentence pairs in (257) would be played. These are about either the male character, the female character, or the item in the centre of the screen:

(257) a. Jongwu-nun mayil pam yele sikan tongan kongpwuha-yess-ta. Kuliko
Jongwu-TOP every night several hour while study-PAST-DECL And
Jongwu-nun cinan sihem-eyse iltung-ul ha-yess-ta.
Jongwu-TOP last test-at first-ACC do-PAST-DECL
'Jongwu studied for many hours every night. And Jongwu was the top student on the last test.'

b. Yuli-nun wutungsayng-i-ta. Yuli-nun sihem-eyse 90cem iha
Yuli-TOP honour student-COP-DECL Yuli-TOP test-at 90 point below

mat-un cek-i eps-ta.
score-ADNOM experience-NOM non exist-DECL
'Yuli is an honour student. Yuli has never scored below 90 on a test.'

c. Kyosil-ey chilphan-i iss-ta. Chilphan-ey-nun amwukesto ssuye iss-ci anh-ta.
classroom-at blackboard-NOM be-DECL blackboard-at-TOP anything written be-CONNECT not-DECL
'There is a blackboard in the classroom. The blackboard doesn't have anything written on it.'

These sentence pairs are designed to make one or the other, or neither, character more discourse salient.

The final sentence is the target sentence, which introduces the second independent variable:

(258) Jongwu-ka Yuli-eykey chilphan yep-eyse **caki/ku/kunye**-ka sihem-ul cal chi-ess-tako malha-n-ta.
Jongwu-NOM Yuli-DAT blackboard beside-at self/he/she-NOM test-ACC well take-PAST-COMP tell-PRES-DECL
'Jongwu tells Yuli beside the blackboard that self/he/she did well on the test.'

For each image, two different versions of the target sentence were produced; one containing *caki*, and one containing one of *ku* or *kunye*. The first two sentence pairs in (257) were named according to the grammatical function of their topic in the target sentence: Subject Bias and (Indirect) Object Bias. The third, which placed additional emphasis on the locative adjunct, was labelled as the Neutral Bias. While the characters and settings would change with each display image, the choice of matrix clause verb remained constant. Across the different scenarios, target sentences were counter-balanced for the gender of the subject, and whether the subject was the left or right character. All sentences were recorded by a native speaker of Korean; each participant heard the same set of recordings, adjusted to a comfortable volume.

In all, there were two independent variables with three levels each:

- Contextual bias: Three levels

1. Emphasis on the target sentence subject. (Subject bias)

2. Emphasis on the target sentence indirect object. (Object bias)

3. Emphasis on the target sentence adjunct. (Neutral bias)

- Anaphor type: Three levels

1. *ku*

2. *kunye*

3. *caki*

This gives a total of nine different conditions. Each participant was given 96 trials in total, presented in two blocks with a five minute break in the middle. The trials were presented in a randomised order for each participant.

The experiment has two dependent variables. The first of these is the Behavioural Data, in the form of a response to a forced-choice comprehension question. After the end of the target sentence has played out, the display is replaced with a question presented on screen in Korean. For the sample classroom scenario, the question would be as shown in (259):

(259) Jongwu-nun nwu-ka sihem-ul cal chi-ess-tako malha-yess-supnikka?
Jongwu-TOP who-NOM test-ACC well take-PAST-COMP tell-PAST-INT
'Who did Jongwu say did well on the test?'

Clickable text boxes containing the names of the two characters would be displayed below the question; participants would click on their answer, indicating their judgement of the antecedent for the target sentence embedded clause subject. Responses selecting the target sentence subject were scored as 1, and those selecting the target sentence object were scored as 0.

The second dependent variable comes in the form of the eye-tracking data. During the playback of the target sentence, participants' eye movements were recorded using a free-standing eye-tracker (Tobii X120), sampling at a rate of 60 Hz. Eye-tracking results were coded according to fixations upon the target sentence subject, object, or the centre item. During the playback of the target sentence, there were three areas of particular interest. The first two were the utterances of the two proper names. Examining the participants' gaze during these periods provides an important control, as it ensures that their eyes to

track to the relevant referents. The most important piece of eye-tracking data is the window of time starting approximately 200ms after the onset of the embedded clause subject. The 200ms delay is to allow for the necessary response time between hearing the onset and the execution of a corresponding eye movement. Eye movements during this time would provide an insight into the initial response upon hearing *caki*, *ku*, or *kunye*.

In all, 27 participants were recruited for this experiment. All were native speakers of Korean residing in the Vancouver area, with no formal education outside Korea before the age of 12. After arriving at the lab and reading the briefing documents, the participants were directed to an experiment booth where they were introduced to the eye-tracking equipment and run through a calibration routine. Once calibration was complete, participants received two practice trials with unambiguous comprehension questions and different images from those used in the experiment trials. These practice trials ensure the smooth function of all the equipment, and get the participants oriented to the interface. After the experiment was completed, the participants were given an optional written debriefing questionnaire, as well as a verbal debriefing with the experimenter.

3.4.2 Results

First, I present the eye-tracking results. Data for only 14 out of the 27 participants are reported here; for 13 of the participants, more than 25% of their eye-tracking data was lost, so their entire data sets were excluded from the analysis. The eye-tracking results for all proper names are presented in Figure 3.2. The arrows in the lower left of the graph indicate the averaged duration of the proper names. The line of circles represents fixations on the correct name, while the line of vertical dashes represents fixations to the incorrect person. The third line represents looks to the item in the centre of the screen. As expected, approximately 400ms after the onset of a proper name, the proportion of fixations to the correct person goes up, while fixations on the incorrect person and the centre of the screen both fall off. This suggests that the participants' eye-movements are tracking along with the audio presentation.

Figure 3.3 shows the proportions of fixations after the onset of *caki* across all three bias contexts. In interpreting these graphs, it is important to remember that immediately before the utterance of *caki* comes the locative adjunct. This explains the relatively high proportion

of looks to the item in the centre of the display, and was included as a deliberate control to try as much as possible to direct the participants' gaze off the two characters before the onset of the embedded clause subject. Across all three biases, the general pattern is the same: after the onset of *caki*, looks to the subject character track upward. Similar results were obtained for *ku* and *kunye*; more than anything else, participants were fixating on the target sentence subject.

Figure 3.4 combines the proportions of fixations for all nine conditions, averaging over the 300ms to 1000ms time duration. While there is some variation in the numbers, the general trend is clear: regardless of the context or type of embedded clause subject, participants were looking at the matrix clause subject more than anything else.

To confirm this observation, a three-way ANOVA was conducted, comparing the variables of Contextual Bias, Anaphor Type, and Target of Fixation. The ANOVA revealed a main effect of Target of Fixation ($F(1,13) = 27.610$, $p = .000$) confirming that participants looked significantly more often at the image corresponding to the subject than anything else, regardless of Anaphor Type and Contextual Bias. There was no main effect of Anaphor Type or Contextual Bias. The ANOVA did further reveal two significant interactions, the first between Anaphor Type and Target of Fixation ($F(2,26) = 4.179$, $p = .027$). This can be attributed to the increase in looks to the subject in the *caki* conditions, possibly a result of the gender-neutrality of *caki* versus *ku* and *kunye*. While in theory only half of the matrix subjects should be an antecedent for *ku* or *kunye*, all are possible antecedents for *caki*. A second interaction between Contextual Bias and Target of Fixation ($F(2,26) =$

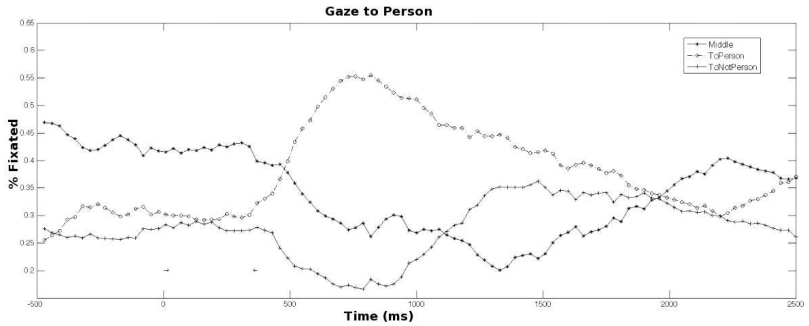


Figure 3.2: Proportion of Fixations over Time to Proper Name Referents

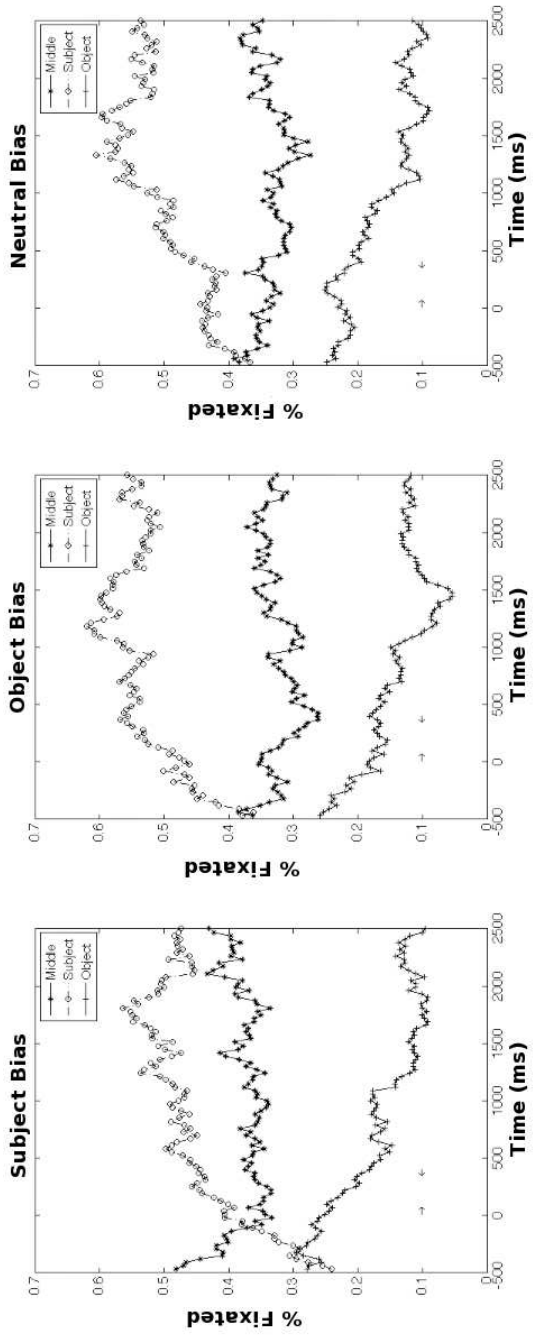


Figure 3.3: Proportions of Fixations in Image after *caki* for each Bias

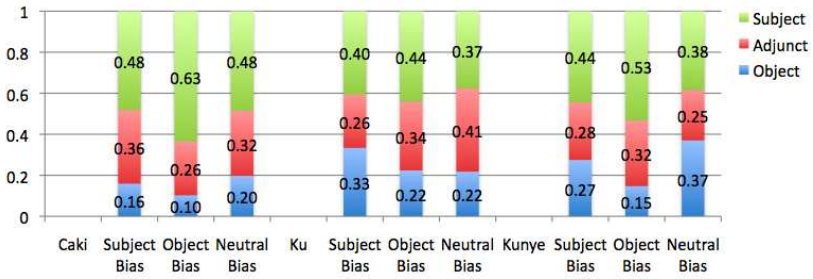


Figure 3.4: Proportions of Fixations in all Conditions

4.317, p = .024) is harder to explain. For all three embedded clause subjects, there was an increase in looks to the target sentence subject in the Object Bias condition. This is an unexpected result, and no credible speculation can be offered.

The behavioural results are presented in Figure 3.5. Recall that selections of the target sentence subject were scored as 1, and the indirect object as 0. For *caki* then, the behavioural results correspond to the eye-tracking results; subjects were selected almost universally. The results for *ku* and *kunye* are more surprising. Recalling that there was an even split between the genders of the target sentence subjects, if *ku* and *kunye* were going to pattern purely according to gender, then scores should be 0.5 across the board. Alternatively, because there was no significant main effect of Anaphor Type in the eye-tracking data, one might expect all three of *caki*, *ku* and *kunye* to come out the same in the behavioural data as well. As shown in Figure 3.5 neither of these results came out.

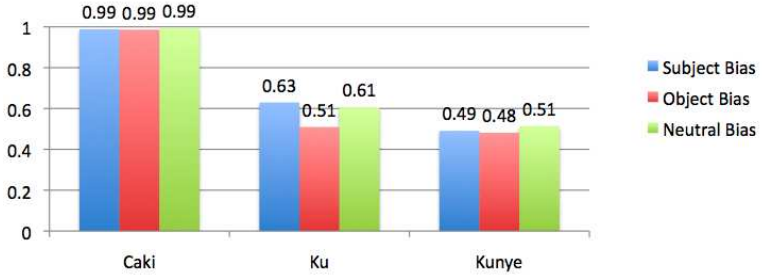


Figure 3.5: Average Means for each Anahpor/Bias Combination

To fully interpret these results, a 2-way ANOVA was conducted, comparing the vari-

ables of Contextual Bias and Anaphor Type. A main effect of Anaphor Type was observed ($F(2,52) = 305.180$, $p = .000$), with all three levels significantly different on pairwise comparisons. This indicates that all three of the embedded clause subjects did something different. There was a main effect of Contextual Bias ($F(2,52) = 7.788$, $p = .001$), though on pairwise comparisons, it was only the case that the Object Bias was significantly different from the other two. There was no significant difference between the Subject and Neutral Bias conditions. Finally, a significant interaction between Anaphor Type and Contextual Bias was found ($F(4,104) = 5.809$, $p = .000$), with *ku* showing the greatest sensitivity to Contextual Bias, less for *kunye*, and virtually none at all for *caki*.

3.4.3 Conclusions

Looking back to the original goals of this experiment, the first was to confirm that in Korean, contextual manipulation would have an impact upon the interpretation of pronouns. Based on the eye-tracking data, this would not seem to be the case. What the eye-tracking data suggest is that regardless of context, there is a "default setting" in the reference resolution system which will always consider the sentential subject first, regardless of the Anaphor Type. The behavioural data tells a different story though, in that context has such an effect on the third person pronouns that it can even trump gender agreement. While the means did remain centred around the 0.5 level expected for full adherence to agreement, *ku* showed a selection of female antecedents in all three Bias conditions with a significant increase in those female antecedents in the Subject and Neutral Bias conditions. The effect was much more subtle for *kunye*, though this too crossed the 0.5 mark in the Neutral Bias condition, indicating that some male antecedents were selected. In debriefing, the participants were very conscious of this effect, reporting it was possible for *ku* and *kunye* to take gender-mismatched antecedents.[5] Taking the two sets of results together suggests that reference resolution in Korean is a two-stage process. There is an initial default to considering the subject as the antecedent, represented by the results for *caki*, but *ku* and *kunye* are subject to further processing in which context can even override gender features, both of which may conflict with the default subject orientation. Going back to the second goal of this experiment, it is clear that *caki* does not show the same sensitivity to context as

[5]Quite a few even asked if this had been the point of the experiment.

do *ku* and *kunye*.

This notion of a default subject setting is bolstered by the finding in the behavioural data of no significant difference between the Subject and Neutral Bias conditions. Assuming there is a default consideration of the subject, then a context which emphasises that subject should have no effect on judgements. The Neutral Bias condition provides a picture of what happens when there is no external factor superimposed upon the default. The result was not significantly different from when the Subject Bias was present; where neither potential referent was emphasised, the result was the same as when the default itself was emphasised.

Beyond the initially-stated goals, the results of this experiment should also be considered within the larger discussion on the nature of *caki*. The proposed bound variable analysis for *caki* does not predict subject orientation in the same way as the LF head movement approach does for *casin*. This experiment, however, provides an external account for the observed subject orientation. Subject orientation emerges for *caki* because subjects appear to be the default antecedents for Korean. *Ku* and *kunye* reflected this bias in the eye-tracking, but showed a further context sensitivity which is not present for *caki*. It could be argued that some factor such as logophoricity was at play here, but it should be noted again that even with the verb *malha*, *caki* is not strictly logophoric. Some sources in the literature would judge the target sentence in this experiment to be ambiguous, with either of the subject or indirect object being possible antecedents for *caki*. Furthermore, the literature shows that *caki* can easily take the indirect object of *malha* as its antecedent if the subject is first-person. Even in this experiment, subjects were not chosen as the antecedent for *caki* 100% of the time, suggesting that non-subject antecedents were possible. Finally, the eye-tracking results from *ku* and *kunye* cannot be discounted: there still appears to be a strong subject orientation, but for these pronouns that is overridden by context. This default is just as likely, if not more, to be the explanation for the behavioural results for *caki* than any logophoric effect.

3.5 Summarising Korean

In summary, the reflexive system of Korean is more complex than what has been seen so far for English. Like English, Korean has a set of bi-morphemic pronouns which generally serve as clausemate reflexives, though the *pronoun-casin* forms lead a double life acting

as emphatics with a similar function as the English exclusive adjacent. The difference in Korean is that here the emphatic stands on its own, whereas in English it attaches either to the nominal or predicate it modifies. Further differentiating the emphatic use between the two languages is the observation from G. Lee and the experiment by J.-H. Kim and J.H. Yoon demonstrating that in these emphatic usages, the bimorphemic reflexives do not require local antecedents. Their different binding mechanism was brought out by the lack of sloppy readings in the J.-H. Kim and J.H. Yoon experiment for *caki-casin*. Again, unlike English, Korean also allows for this exclusive emphatic to attach directly to proper nouns.

Further differentiating Korean from English is the existence of two monomorphemic forms which can be used in co-argument reflexive contexts, as well as long distance contexts. The lesser-discussed of these, *casin*, appears to fit the general description of a long-distance anaphor, a key fact in this analysis being the presence of the blocking effect as observed for Mandarin *ziji*. Kang's corpus analysis supports the position that *caki* is something different. The exact nature of *caki* though has been a matter of some debate for at least the last 30 years. While generally split into two camps, long-distance anaphor versus pronoun, a middle road has been consistently hinted at in the literature: bound variable anaphora, which derives the long-distance readings *in situ*, as opposed to positing the type of movement proposed by Cole et al.

The final summary on the distribution of *caki*, updated to reflect the results of the new corpus and experimental work, is presented in Table 3.10. The new corpus study of *caki* presented shows that all instances of *caki* can be brought under the umbrella of such an analysis. As such, *caki* can be considered a reflexive in that its semantics can create a reflexive-like relation between coarguments, and Kang's corpus work shows that locally-bound *caki* is productive. In light of the corpus findings presented here, along with the reactions of my own consultants, I treat the cases of *caki* having an antecedent within an embedded clause with no matrix psych predicate to be ungrammatical. At best, these can be seen as exempt uses, which are rare, though reported.

Differentiating *caki* from the English *self* pronouns is the fact *caki*'s semantics equally allows for non-reflexive (long-distance) uses. The eye-tracking experiment provides one more piece of evidence for the camp that *caki* should not be considered a pronoun in that *caki* does not show the same contextual sensitivity as *ku* and *kunye*. The experiment also provides evidence that subject orientation is feature of Korean anaphor resolution in gen-

Table 3.10: Distribution of *Caki*: Final Summary

Environment	Status
Local Antecedent	Grammatical
Non-Local Subject Antecedent	Grammatical
Non-Local Non-Subject Antecedent	Disputed
Non-C-Commanding Antecedent in Psych Predicate	Grammatical
No Antecedent in Reportative Context	Grammatical
Antecedent inside Embedded Clause	Ungrammatical
Antecedent inside Genitive DP	Grammatical
Split Antecedent in Plural	Ambiguous
Discourse Antecedent	Grammatical
Logophoricity	Disputed
Generic/Arbitrary Reading	Grammatical
Inherent Reference/Deictic Reading	Grammatical
Resumptive Pronoun Reading	Grammatical
Local Bound Variable Reading	Grammatical
Non-Local Bound Variable Reading	Grammatical
Strict Reading in Ellipsis/Focus	Ungrammatical
Crossover Structures	Ungrammatical

eral, not just *caki*. This was demonstrated in the eyetracking results, which showed a default effect of always considering subjects over other antecedents, regardless of context and type of anaphor. The last open disputes with respect to *caki* call for future experimental work. While the reported eye-tracking study allowed participants to choose their preferred antecedent for *caki*, the experiment can be redesigned to use the same target items in more explicit contexts where only one antecedent is possible. The behavioural part of the experiment would then become a truth value judgement task, where participants would evaluate an ambiguous *caki* against a context where the referent (subject or indirect object) is fixed. This manipulation would allow for more targeted testing of the non-subject antecedent uses, as well as more direct testing for logophoricity.

Like English, Korean will be revisited in Chapter Six, where an STAG analysis of bound variable anaphora as exemplified by *caki* will be presented. In the next chapter, I shift from East Asia to Southern Africa, with a discussion of reflexivity in Shona. There too, I will propose that the best account for the observed data lies in bound variable anaphora.

Chapter 4

Keeping it Local
Bound Variable Reflexivity in Shona

Eight is enough!
-Mary Bradford. *Family Guy*.

This chapter presents an examination of the reflexive *zvi* in the Bantu language Shona. It begins with an overview of the literature on the subject, illustrating that there is a debate in the literature on the issue of whether or not the reflexive morpheme *zvi* is a detransitivising operator on the verb, or an anaphoric element. Further clouding the issue is that there is a second *zvi* form, which occupies the same position in the verb morphology, marking object agreement. To address these issues, I discuss related languages, as well as a set of defining characteristics of object markers and detransitivising reflexives. I argue that *zvi* is not a detransitiviser, and instead shows characteristics of a locally-bound variable. Based on corpus evidence indicating non-reflexive contexts in which *zvi* appears, I conclude that Shona reflexivity is instantiated by a covert bound variable, restricted to local binding. The *zvi* morpheme is not taken to be inherently reflexive, but rather a marker of agreement triggered by the presence of this bound variable. Homophony between the reflexive *zvi* and the object agreement marker *zvi* thus reduces to a single account of object agreement under this analysis. The chapter begins with a basic overview of Shona syntax.

4.1 Introducing Shona

Shona is an SVO language with an extensive series of verbal affixes, indicating argument agreement(s), tense, and various alterations to the verb's valence. This section is by no means a comprehensive presentation of Shona morphosyntax; rather it is intended to introduce the forms which are relevant to the later discussion.

A minimal transitive sentence in Shona is given in (260):

(260) A-ka-bik-a gudo.
SUBJ.1-REM.PST-cook-FV baboon.5
'He cooked the baboon.'

In (260), the verb-final suffix is standardly glossed as 'final vowel', despite intuitions that there is some function beyond simply maintaining a CV syllable structure. While in most contexts *-a* is used, this vowel changes in imperatives and negated sentences, suggesting that this is actually a clause-typing morpheme. The inner prefix *ka-* is a tense marker, and the outer prefix *a-* marks subject agreement. While some sources describe this as a pronominal, the generally-accepted position in the formal literature, working from Bresnan and Mchombo (1987), is that the subject marker is a reflex of syntactic agreement, rather than an incorporated pronoun. (260) would thus be better re-cast as (261):

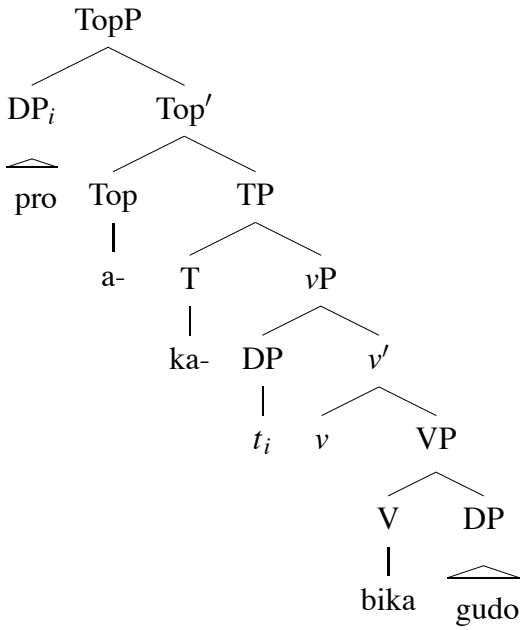
(261) *pro* A-ka-bik-a gudo.
pro.1 SUBJ.1-REM.PST-cook-FV baboon.5
'He cooked the baboon.'

Here, *pro* takes the place of the subject, carrying the noun class (ϕ features) which control agreement on the verb. Agreement is with one of a set of 21 different noun class markers, class one being the marker of a singular person.[1] The prefix ordering is derivable through cyclic head movement from V^0 up the verbal spine. Making an uncontroversial assumption that the tense marker would be hosted at T^0 suggests an analysis in which the subject agreement is at a higher head, and sentence-initial subjects appear in a specifier position of this higher head.

[1] Classes 1-2, 3-4, 5-6, 7-8, and 9-10 all form singular-plural pairings. Other noteworthy class markings include 15 which appears on nominalised verb stems, and the 16-18 range which mark locatives. Additionally, there are distinct forms for first and second person, singular and plural, making for 25 possible agreements (Fortune, 1955).

This dovetails nicely with the observation that the sort of pro-drop shown in (261) is contextually-limited. This sort of *pro* subject is only admissible when the referent is recoverable from previous context (in this case with reference to a folk tale involving an anthropomorphised hare and baboon). The same restriction is observed with overt subjects: sentential subjects in Shona must be topical. Following work on Swahili and Kirundi presented in Henderson (2006), combined with Shona data demonstrating the inadmissibility of indefinites and focused elements in subject positions, Bliss and Storoshenko (2008) propose that sentential subjects in Shona are derived through A′ movement to a topic position in the C-domain, rather than A-movement to [Spec, TP]. The basic clause structure of Shona is given in (262):

(262)



Two observations are in order concerning the tree in (262). The first of these is that the subject does not pass through [Spec, TP], the standard position for nominative case assignment. This is proposed in conjunction with the analyses of Harford-Perez (1985) and Diercks (in press b) which claim that there is no case marking in Bantu languages. The second issue is that no separate head is given for the final vowel. This treatment is "industry standard" (Simango, 2006; Diercks, in press a), though it does ignore the possibility that this morpheme does have a separate head; I will have more to say on this shortly.

In addition to the core clausal elements, Shona makes use of a set of verbal suffixes

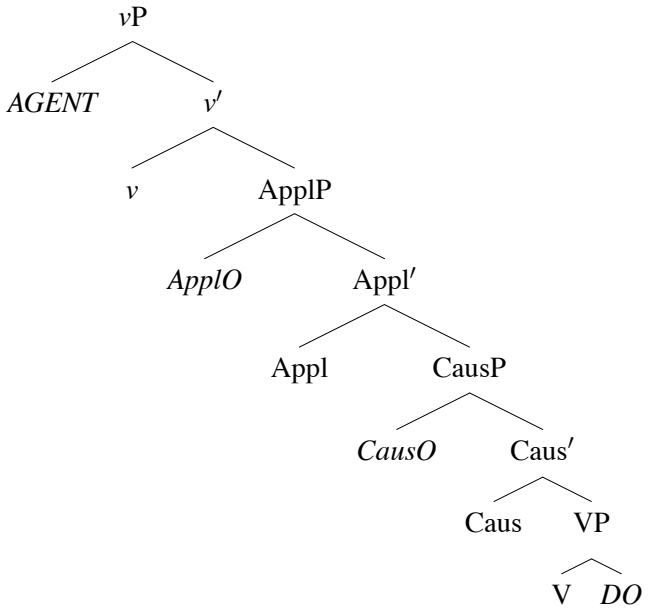
which alter the valence of the predicate. I limit my discussion to an illustration of three of these: the applicative, the causative, and the passive. It is possible to get all three marked on one predicate:

(263) Mu-sango ma-ka-don-edz-(er)-w-a Mufaro na Shingi.
CL.18-forest SUBJ.18-REM.PST-fall-CAUS-APPL-PASS-FV Mufaro by Shingi
'In the forest was caused to fall Mufaro by Shingi.'

In (263), the unaccusative predicate *don* 'fall' has taken the causative *-edz*, optionally takes the applicative *-er*, and finally is marked with the passive *-w*. The causative adds an external causer to a given event, in this case *Shingi*. According to Bliss (2009), in addition to introducing benefactives, the applicative *-er* may be used to introduce locative elements as it does here with *mu-sango* 'in the forest'. In this example, the locative straddles the line between argument and adjunct in that it does not necessarily need to be licensed by the applicative in order to become the sentential subject under passivisation.

The relative ordering of these suffixes is fixed, leading Bliss to propose the following structure for the lower verbal domain, illustrated for an active transitive with applicative and causative marking:

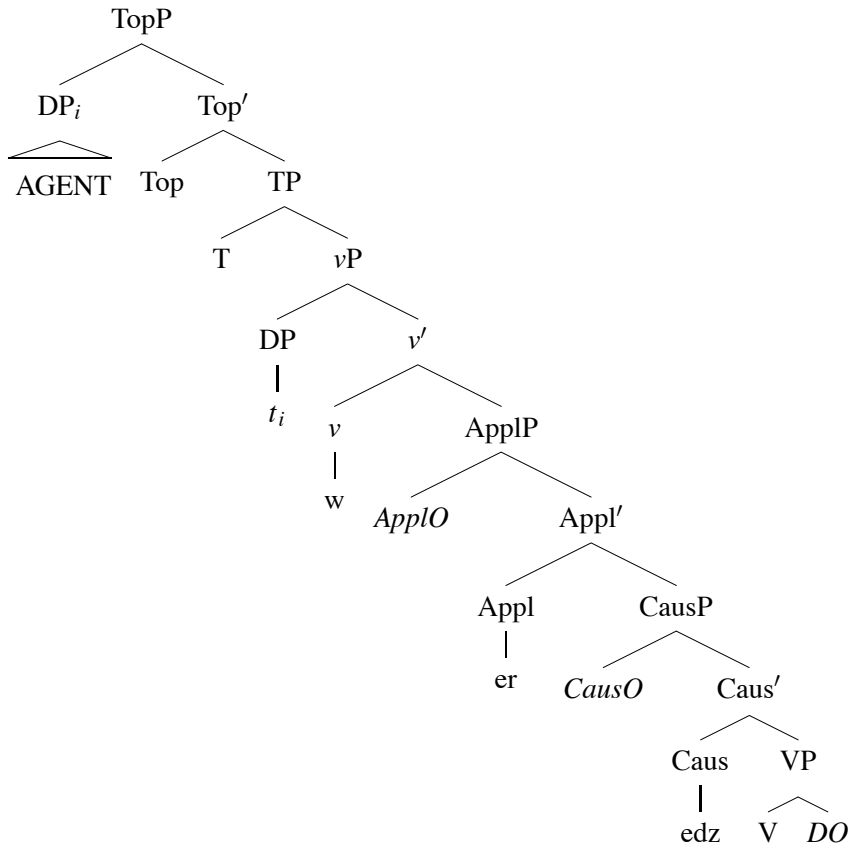
(264)



In this structure, the CausO position corresponds to the causee, while the external causer appears in the AGENT position. The analysis in Bliss and Storoshenko treats the passive

morpheme as a variant of the *v* head. The relative hierarchy of the applicative and causative is supported by tests using relative quantifier scope judgements, and the head-order is supported by Baker's mirror principle. That is, the syntactic hierarchy of heads mirrors the morpheme order as attached to the verb. Combining the trees in (262) and (264) yields the following configuration:

(265)



In examining the head order, it appears that all heads from *v* downwards are right adjoined as suffixes, while everything above *v* adjoins to the left as a prefix. It is for this reason that the final vowel stands out as unusual: if indeed it is a clause-typing morpheme, then one would expect it to be the outermost prefix, reflecting the highest head, rather than the outermost suffix. One possible explanation would be to propose an analysis whereby the direction of affix adjoining changes when a phase boundary is encountered. All lower domain heads adjoin to the right, up until *v*. Then, everything up to the topic position adjoins to the left. Somewhere above the Topic in the C domain would be another phase boundary

which switches the direction of adjoining back to the right, explaining the position of the final verb at the highest functional head.[2]

The only thing missing from the picture at this point is object agreement, described as optional:

(266) a. Ndi-no-tem-a huni.
SUBJ.1*st*-HAB-chop-FV firewood.10
'I chop firewood.'

b. Ndi-no-**dzi**-tem-a huni.
SUBJ.1*st*-HAB-OBJ.10-chop-FV firewood.10
'I chop firewood.'

As shown in (266), optional object agreement appears to the immediate left of the verb root, placing it between T^0 and v^0 in the structure from (265). In this example, the object marker appears in conjunction with the direct object. Used in this fashion, native speaker judgements have described this as marking definiteness or just redundant mention of the object. The connection between object markers and definiteness has also been cited for Swahili (Creissels, 2001), along with a reading of topicality. Kunene (1975) notes for Zulu that object markers in that language are licensed only when the object being marked has been mentioned in previous discourse. This observations has been confirmed by one of my Shona consultants, who systematically uses object marking only when the object has been mentioned in a prior sentence.

Like subject agreement, object markers can be employed without an overt object:

(267) Mufaro a-ka-**mu**-don-edz-er-w-a na Shingi.
Mufaro SUBJ.1-REM.PST-OBJ.18-fall-CAUS-APPL-PASS-FV by Shingi
'Mufaro was caused to fall there by Shingi.'

In this variant of (263), the internal argument has become the subject of the passive, and the locative has been dropped, replaced by an object marker displaying locative class agreement. This usage of object markers without an overt object appears to be more common than (266b), though it is still constrained to contexts where the object is found in prior discourse.

[2]This analysis is speculative and peripheral to the issue at hand, thus held over for future work.

There are additional constraints on the use of object markers. Recall that in (263) applicative marking is optional; this is not the case in (267):

(268) * Mufaro a-ka-**mu**-don-edz-w-a na Shingi.
Mufaro SUBJ.1-REM.PST-OBJ.18-fall-CAUS-PASS-FV by Shingi.
'Mufaro was caused to fall there by Shingi.'

With the applicative removed, the sentence becomes ungrammatical. This suggests that unlike the subject position, which is open to arguments and some adjuncts, the object marking position is restricted to arguments which have been introduced by the predicate or one of the verbal extensions. Thus, either the object marking position is more sensitive to the argument structure of predicates, or the difference is purely structural, arising from the fact the topic projection has a wider domain in which to probe for a goal. A sensitivity to argument structure is somewhat at odds with the information-structure sensitive nature of the object position. In a sense, the object marker appears to have the function of a secondary topic position, identifying arguments from the lower verbal domain which are discourse-old. Furthermore, one might expect that if the object markers were strictly connected to the argument structure of the predicate, in a situation where there are multiple arguments in the lower verbal domain, it would only be possible to mark one of the two objects (most likely the highest). However, this is not the case:

(269) a. Shingi a-ka-**yi**-bik-ir-a i-ye.
Shingi SUBJ.1-PST-OBJ.9-cook-APPL-FV PRN-CL1
'Shingi cooked it for him.'

b. Shingi a-ka-**mu**-bik-ir-a i-yo.
Shingi SUBJ.1-PST-OBJ.1-cook-APPL-FV PRN-CL9
'Shingi cooked it for him.'

The context in (269) is that Shingi is cooking goat (a class 9 object) for her son. Either the class 9 direct object or the class 1 applicative object may be picked up by the object marker, depending on context. In this example, the unmarked objects are replaced by pronouns, a secondary strategy for discourse-old referents. If it were really an argument-adjunct distinction which was responsible for the contrast between (267) and (268), then one might expect that the object marker position should be keyed to a specific argument. Rather, what (269) shows is that either of the direct object or applicative object is a candidate for object

marking. This suggests that object marking has a broader search domain than just a single argument position. If the non-argument use of the locative in (268) is outside that structural domain, then this would explain the impossibility of object marking.

In the Bantu literature, there are two competing analyses for the nature of the object markers. The first of these, exemplified by Bresnan and Moshi (1990), treats the object markers as incorporated pronouns. This is based on data from Kichaga and Chicheŵa in which object markers are claimed to be in complementary distribution with full nominal objects. Shona, as has already been shown, does not reflect this same complementary distribution. A second distinction is that Kichaga allows for multiple object markers, while Shona is restricted to just one. Bresnan and Moshi do not comment on whether there are contextual factors which determine how many arguments are object marked, but multiple object-marking languages have a fixed order for the object markers, generally a mirror image of the post-verbal order for the corresponding nominal objects (Creissels, 2002). This suggests that object marking in these languages is tied to the syntactic positions of the arguments.

Restrictions on the co-occurrence of object markers in passivised ditransitives are also noted for some languages. According to Woolford (1995), only languages which allow multiple object markers in the active should allow object marking in the passive. She notes one exception to this, SiSwati, which allows for only one object marker in the active, and yet still allows for an object marker in the passive, but only in the case where the higher of the two objects has been moved to the subject position. Object marking of an applicative object is blocked in a passive where the direct object has moved to the subject position in that language.

Shona does not fit any of these patterns for passives and object marking. Firstly, Shona does not allow multiple object markers in the active:

(270) a. * Mufaro a-ka-**ya**-**mu**-bik-ir-a.
Mufaro SUBJ.1-PST-OBJ.9-OBJ.1-cook-APPL-FV
'Mufaro cooked it for her.'

b. * Mufaro a-ka-**mu**-**ya**-bik-ir-a.
Mufaro SUBJ.1-PST-OBJ.1-OBJ.9-cook-APPL-FV
'Mufaro cooked it for her.'

These are the same object markers which were present in (269). Regardless of the order-

ing, they cannot be used simultaneously. However, Shona does tolerate object marking in the passive. Unlike SiSwati though, there is no restriction on which object can be object marked and which can be passivised:

(271) a. Ma-nhanga a-ka-**yi**-bik-ir-w-a.
CL6-pumpkin SUBJ.6-REM.PST-OBJ.9-cook-APPL-PASS-FV
'Pumpkin was cooked for it.'

b. Mbudzi ya-ka-**a**-bik-ir-w-a.
goat.9 SUBJ.9-REM.PST-OBJ.6-cook-APPL-PASS-FV
'For the goat was cooked it.'

In (271a), the direct object *manhanga* 'pumpkin' has moved into the subject position, controlling subject agreement. In (271b), the applied object *mbudzi* 'mbudzi' has moved up, and the direct object appears as a class 6 object marker. All of this points to an analysis in which the Shona object marker is definitely a unique position (there can only be one), but it is not concerned at all with the structural position of the object within a given domain.

In some sense, the object marker functions rather like the subject agreement of a passive sentence. Under the Bliss and Storoshenko analysis, Shona passive sentences can take any DP element as subjects to check an EPP feature on a Topic head. In the active, external arguments are treated as having privileged status, but when the external argument is removed in the passive, this EPP can target any topical element in its c-command domain. A similar phenomenon is observed for the object markers: they are restricted to marking only one topical (i.e. discourse old) element, but there does not appear to be any privileged element based on syntactic structure. Object marking can be seen as an optional agreement with a second topic distinct from the "primary" topic in the subject position, so long as that second topic is in an argument position.

The alternative treatment for object marking in Bantu may be more compatible with this analysis. Rather than treating subject markers and object markers as distinct phenomena, some analyses make use of a parallel treatment for both. Visser (2008), in her treatment of isiXhosa, makes use of parallel AgrS and AgrO heads for subject and object agreement. She describes a situation similar to Shona in which an optional object marker carries an additional meaning, though she characterises this as "specificity" rather than topicality or definiteness. For cases where the object marker appears without an overt object, she posits a ϕ-featured *pro* in the original object position: while subject agreement is strictly Spec-Head

in her analysis, object marking does not involve movement into a Spec-Head configuration, merely c-command. This approach may be a better fit for Shona, which does not show the same degree of sensitivity to the relative positions of the object arguments as other Bantu languages. While there is only ever one object marker, there is no observed constraint on which argument may be marked in the active or the passive. This discussion of the syntactic character of the object markers is an important first step in the discussion on reflexives in Shona, as I will argue in the next section that the Shona reflexive is a member of the set of object markers.

4.2 Shona Reflexives

Reflexivity in Shona is expressed using the morpheme *zvi*:

(272) a. *pro* Nda-ka-**zvi**-pis-a.
pro.1st.SG 1st.SG.SUBJ-PST-REFL-burn-FV
'I burned myself.'

b. Mwana a-ka-**zvi**-pis-a.
NC1.baby SUBJ.1-PST-REFL-burn-FV
'The baby burned itself.'

c. Mbudzi dza-ka-**zvi**-pis-a.
goats.10 SUBJ.10-PST-REFL-burn-FV
'The goats burned themselves.'

As shown in (272), *zvi* appears immediately preceding the verb root: the position canonically reserved for object markers. However, unlike the object markers, *zvi* does not show any agreement, suggesting that this morpheme could be a detransitiviser, operating upon the argument structure of the verb in the same way as the passive or other valence-changing operations. For first person, second person, and all third person uses (regardless of noun class) *zvi* is used for singulars and plurals. As discussed in the previous section, such valence operators in Shona generally appear after the verb root, rather than before. Thus, the first question addressed in this section will be the issue of "object marker versus valence operator". After concluding that *zvi* belongs to the set of object markers in the language, the second issue to address is whether *zvi* has a unique status among the set of object markers.

Ultimately, I argue that reflexivity in Shona is best-treated under a bound-variable analysis, with *zvi* being the marker of agreement with a covert variable.

4.2.1 The Literature on Shona and Beyond

In the Shona literature, the *zvi* morpheme is usually glossed as a reflexive, with little to no argumentation to support this analysis, nor a full discussion of its distribution. Fortune (1955), says very little about reflexive *zvi*, which he groups among object markers, noting only that it is identical to the class 8 object marker. Fortune does not expand on this point, but this connection between reflexivity and class 8 is something to which I return later in the chapter. Brauner (1995) devotes all of one sentence to the subject, stating that reflexives are marked by an infix *zvi*, which is treated as an object. Interestingly, Brauner still lists the reflexive among affixes such as the applicative or the passive, valence changers, rather than the object markers. Similarly, where Bellusci (1991) notes that Shona verbs take a number of suffixes which alter the valence of the predicate, *zvi* is counted among them. Thus, the question of whether *zvi* is an object marker or a valence operator is open for debate. To get further insight into this issue, it is necessary to look at how reflexives are treated in other Bantu languages. Here, I will discuss the state of affairs in four of these languages: Zulu, Xhosa, Tswana, and Kikamba.

In Zulu, the reflexive morpheme is *zi*. Like Shona, this morpheme occupies the same position as the object markers in the language. Kunene (1975) observes that in Zulu, the reflexive may or may not co-occur with the co-referential object of the sentence:

(273) Zulu

a. Umfana u-ya-**zi**-shaya.
boy he-ASP-REFL-hit
'The boy hits himself.'

b. U-ya-**zi**-shaya.
he-ASP-REFL-hit
'He hits himself.'

c. U-**zi**-shaya yena umfana.
he-REFL-hit him boy
'He hits himsélf, the boy.'

(273a) shows the basic reflexive structure of Zulu, which is parallel to that of Shona. In (273b), the subject is only indicated by the agreement on the verb, which Kunene claims can only occur when the identity of the subject is discourse-old information. Finally, (273c) shows the same discourse-old subject, but with a redundant mention of the object, in the form of *zi* and *yena umfana*. This is claimed to be a contrastive focus construction, indicated by the accent in the gloss, where the speaker is making the implicit claim that the boy hits himself, not some other person.

The reflexive *zi* is also attested in Xhosa (Dalyedwa, 2002), even appearing in complex sentences involving multiple verb extensions:

(274) Xhosa

Unomyayi u-**zi**-leq-is-el-a intshonsho ebaleni.
black crow SM-REFL-chase-CAUS-APPL-FV chick in field
'The black crow causes himself to chase the chick for himself in the field.'

Dalyedwa does not elaborate on how exactly this sentence is doubly-reflexive, though she notes elsewhere that the object marker position can stand for any one of the objects of the verb, similar to Shona. So, while it is not surprising that the reflexive could either be the causative object (causee) or the applied object, the fact that it emerges as both in the same sentence is unexpected based on the fact that there has so-far been a one-to-one relationship between object markers and arguments, as well as valence operators and the arguments they add or subtract.

The situation in Tswana is somewhat more complex. Firstly, Tswana is one of the Bantu languages which allows multiple object marking (Creissels, 2002):

(275) Tswana

a. Ki-f-ets-e bomalome dikgomo letswai.
SUBJ.1st-give-APPL-FV NC2.uncle NC8.cow NC5.salt
'I gave salt to the cows for my uncles.'

b. Ki-**li**-**di**-**ba**s-f-ets-e.
SUBJ.1st-OBJ.5-OBJ.8-OBJ.2-give-APPL-FV
'I gave it to them for them.'

As shown, the ditransitive verb root *f* 'give' can be extended with the applicative, yielding a sentence with three objects. In (275b), all three are replaced with object markers in the expected mirrored order.

The ordering of the object markers becomes relevant under passivisation. When a ditransitive is passivised, either of the two objects may raise to the subject position, but like SiSwati (and unlike Shona) only in one configuration can the other remain as an object marker:

(276) Tswana

a. Ki-f-il-e bana dikwalo.
SUBJ.1st-give-PERF-FV NC2-.baby NC8.book
'I gave the books to the children.'

b. Bana ba-**di**-f-il-w-e.
NC2.baby SUBJ.2-OBJ.8-give-PERF-PASS-FV
'The children were given them.'

c. Dikwalo di-f-il-w-e bana.
NC8.book SUBJ.8-give-PERF-PASS-FV NC2.baby
'The books were given to the children.

d. * Di-**ba**-f-il-w-e.
SUBJ.8-OBJ.2-give-PERF-PASS-FV
Intended: 'They were given to them.'

Starting with the active ditransitive in (276a), under passivisation, either of the two objects can be raised into the subject position, as shown in (276b) and (276c). However, while an object marker is permitted when the first object has raised (276b), it is not permitted when the second object has been promoted to the subject position (276d). This further suggests an interaction between passivisation and the licensing of object markers, and if the reflexive is to be considered an object marker, then reflexives may be equally sensitive to passivisation. However, given that Shona has not so far shown this kind of object marker sensitivity to the passive, what applies for Tswana need not be expected to apply in Shona.

Reflexive meanings in Tswana are realised through the *i* morpheme. As in the previous cases, this morpheme appears in the same position as object markers, and can co-occur with another object marker if the verb's valence permits multiple objects. However, Creissels argues that this is not an object marker in Tswana, but rather a middle voice operator which blurs the distinction between agent and patient. He notes that there are uses of *i* in Tswana which are not prototypically reflexive, and that there are a number of verb roots in the language which only occur with *i*. To explain the positional and morphological similarity

with the object markers, Creissels proposes that *i* is derived from what was historically a reflexive object marker, but has undergone lexicalisation over time and become a valence operator in Tswana.

Kioko (2005) also gives a diachronic account of the reflexive marker, but to the opposite conclusion. In Kikamba, she notes that the reflexive marker *i* is unusual in that it does not reflect any agreement with the noun class of its antecedent. Kioko uses three tests to determine whether or not *i* should be considered an object marker, or something more akin to a valence operator. Based upon the location of the morpheme, co-occurrence with full nominal objects, and final vowel alternations in imperatives, Kioko concludes that *i* in Kikamba is an object marker. The fact that there are apparent cognates for this marker in various Bantu languages suggests that there is a historical change underlying this morpheme. Kioko speculates that the Kikamba reflexive could be the only surviving member of a full set of reflexive markers which once existed in Bantu, and that grammaticalisation has led to this general form which does not observe noun class agreement conventions. I will not present Kioko's tests for Kikamba here, but I will replicate those tests in Shona. Before moving one, one final cross-linguistic observation is that in addition to their phonological similarity, these reflexives all have one thing in common with that of Shona: for Zulu, Xhosa, and Kikamba, the reflexive form is identical to the class 8 object agreement. Of the languages surveyed, the only language which differentiates class 8 agreement and reflexivity is Tswana, and it is also the only one given a detransitivising analysis.

4.2.2 Object Marker versus Detransitivisation

In her discussion of the reflexive marker in Kikamba, Kioko formulates three morphological tests to determine whether *i* in that language is an object marker or a valence operator. Here, I present the results of those tests, as applied to Shona.

Kioko's first test is distributional: the reflexive in Kikamba occupies the object marker "slot" between the tense morpheme and the verb root. This has already been shown to be the case for Shona, and is one of the primary reasons to treat *zvi* as an object marker. More telling are the following ungrammatical sentences:

(277) a. * A-ka-**yi**-**zvi**-bik-ir-a.
SUBJ.1-REM.PST-OBJ.9-REFL-cook-APPL-FV

Intended: 'He cooked it for himself.'

b. * A-ka-**zvi**-**yi**-bik-ir-a.
SUBJ.1-REM.PST-REFL-OBJ.9-cook-APPL-FV
Intended: 'He cooked it for himself.'

The pair in (277), echoing (270) where two object markers on the same verb are ruled out, shows that *zvi* and an object marker cannot co-occur, suggesting that they compete for the same position.

The second test is based on the fact that in Kikamba, it is possible to have a verb's object indicated both by an object marker, and a full NP after the verb. The same pattern holds with the reflexive *i*, though Kioko does not detail the same discourse effect as Kunene does in Zulu. In her discussion of object marking, Kioko claims that in Shona it is not possible to have both the object marker and a full object nominal referring to the same entity. However, Fortune makes the opposite claim about object markers and full nominal objects in Shona, with numerous examples. While this co-occurrence may have an information-structural function, it is not ungrammatical. Thus, as with Zulu, it should be possible to have *zvi* occur with the full object nominal, if *zvi* is an object marker:

(278) ? Shingi a-ka-**zvi**-bik-a Shingi.
Shingi SUBJ.1-PST-REFL-cook-FV Shingi
'Shingi cooked herself, Shingi.'

This sentence has been judged as "very marginal", but not categorically ungrammatical, and it is reported to be possible that embedded in a suitable discourse context, the judgement might improve. That the sentence in (278) is not categorically ungrammatical may be seen as an argument for treating *zvi* as an object marker for another reason. If it really were decreasing the number of argument positions for the predicate, then a full nominal object should not be licensed whatsoever. Given the marginal status of the sentence and its heavy dependence on context though, it would be better to have more to go on.

The final test employed by Kioko is based on the observation that when an imperative in Kikamba takes an object marker, the final vowel on the verb is *-e*, rather than the *-a* found with a full nominal object. Because imperative reflexives in Kikamba have a final *-e* vowel, Kioko takes this as further proof that *i* is an object marker.

Brauner (1995) notes the same alternation in Shona; when object markers are used in the imperative, the final vowel is inflected as *-e*, marking subjunctive mood. The same pattern emerges with *zvi*:

(279) a. Gez-a mwana!
wash-FV baby.1
'Wash the baby!'

b. **Mu**-gez-e!
OBJ.1-wash-FV
'Wash him/her!'

c. **Zvi**-gez-e!
REFL-wash-FV
'Wash yourself!'

(279a) shows a transitive imperative sentence with the full nominal object and a final vowel *-a* on the verb. When an object marker is used to refer to the baby (279b), the final vowel does indeed change to *-e*. The reflexive imperative (279c) also has the *-e* final vowel. As such, this is further evidence that *zvi* patterns with the object markers.

Having tested *zvi* against diagnostics for object markers, a next logical step would be to test it against diagnostics for valence operators. In his discussion of reflexivity, Lidz (1996) makes a set of observations identifying universal behaviours of verbal reflexives (i.e. predicates in which reflexivity is expressed through a valence-changing operation). These centre around a cross-linguistic generalisation that verbal reflexives tend to be broader in function than simple reflexivity. Two non-reflexive functions in particular, he claims, are present in all languages having a verbal reflexive. Testing *zvi* against these universals would thus point to whether or not it should be included in the set of verbal reflexives, essentially testing whether or not *zvi* is a detransitiviser.

The first observation is that verbal reflexives are also used in decausative constructions:

(280) a. Imbabura Quechua

pungu-kuna-ka paska-**ri**-rka.
door-PL-TOP open-REFL-PST.3
'The doors opened.'

b. Kannada

baagil-u mučč-i-**koND**-itu.
door-NOM close-PP-REFL.PST-3.SM
'The door closed.'

In these sentences, a transitive verb is being used with only the theme present. In Shona, the reflexive is not licit in these contexts:

(281) a. Mu-siwo wa-ka-vhar-a.
CL3-door SUBJ.3-REM.PST-close-FV
'The door closed.'

b. Whindo ra-ka-puts-ik-a.
window.5 SUBJ.5-REM.PST-break-STAT-FV
'The window broke.'

In the first sentence, there is no marking on the verb indicating that only one of the arguments is present. The equivalent transitive verb has exactly the same form. In the second case, the stative morpheme *-ik* is added to the verb stem. Thus, while Shona appears to have multiple means of expressing this decausative function, the reflexive *zvi* is not among them.

The second observation made by Lidz is that a verbal reflexive universally shows up on a transitive predicate where the object is possessed by the subject:

(282) a. Fula

O hett-**ike** fedenndu.
he cut-REFL.PERF finger
'He cut his finger.'

b. Kannada

hari-yu tann-a angi-yannu hari-du-**koND**-a.
Hari-NOM self-GEN shirt-ACC tear-PP-REFL.PST-3.SM
'Hari tore his shirt.'

Again, the evidence is that *zvi* does not have this function:

(283) a. Shingi a-ka-won-a ruoko wa Mufaro.
Shingi SUBJ.1-PST-see-FV hand.3 POSSD.3 Mufaro
'Shingi saw Mufaro's hand.'

b. Mufaro a-ka-won-a ruoko w-ake.
Mufaro SUBJ.1-PST-see-FV hand.3 POSSD.3-POSSR.1
'Mufaro saw his hand.'

c. * Mufaro a-ka-**zvi**-won-a ruoko w-ake.
Mufaro SUBJ.1-PST-REFL-see-FV hand POSSD.3-POSSR.1
'Mufaro saw his hand.'

The first sentence in (283) shows the basic structure for a possessed object. When the object is possessed by the subject, as in the second sentence, the reflexive does not emerge, and is shown to be ungrammatical in the third sentence. Still, given the awkwardness of (278), it could simply be that *zvi* is reacting to the repeated presence of the possessor. As shown in (284), even a simple first person sentence is not permissible in this form:

(284) * Nda-ka-**zvi**-won-a ruoko.
SUBJ.1ST-PST-REFL-see-FV hand.3
'I saw myself the hand.'

Here, no possessor appears, and the sentence is still ruled out. The conclusion of this test is backed up by data from Fortune (1955): the reflexive does not emerge in this possessive context.

Based upon these two tests, it appears that *zvi* does not conform to the two universals for verbal reflexives put forth by Lidz. At this point, there is evidence showing that *zvi*- patterns like an object marker in Bantu, and evidence that it does not pattern with valence-reducing reflexives across languages. As such, it can be tentatively concluded that *zvi* is not a valence operator. Still, all the tests so far have been searching for positive evidence that *zvi* is an object marker, either directly, or by proving that it cannot be a valence operator. The question can also be approached from the other direction; the next section will look for behaviour which would not be expected of an object marker in Shona.

4.2.3 *Zvi* as a unique member of the set of Object Markers

Based on everything presented thus far, *zvi* looks as though it can be treated as a member of the set of object markers. It should then show another trait that appears to be unique to the object markers in Shona: an unrestricted use in the passive. Recalling from before, Shona

allows object markers in the passive, with no constraints based on the relative positioning of the arguments:

(285) a. Mu-riyo wa-ka-**mu**-bik-ir-w-a.
CL3-vegetable SUBJ.3-PST-OBJ.1-cook-APPL-PASS-FV
'Vegetables were cooked for her.'

b. Mufaro a-ka-**u**-pis-is-w-a.
Mufaro SUBJ.1-PST-OBJ.3-burn-CAUS-PASS-FV
'Mufaro was caused to burn them.'

In (285), two different object marking configurations are given. (285a) shows a situation in which the direct object has moved to the subject position, and the applied object is represented by an object marker. It was already shown in (271) that the opposite marking was possible, with the applied object going up and the direct object as an object marker. (285b) shows a slightly different configuration, where the causee, considered to originate at the specifier of the Caus^0 head, has moved up into the subject position, with the theme emerging as an object marker. Both of these are fine with object markers, but something changes with the reflexive:

(286) a. * Shingi a-ka-**zvi**-bik-ir-w-a.
Shingi SUBJ.1-PST-REFL-cook-APPL-PASS-FV
'Shingi was cooked for herself.'

b. ? Mufaro a-ka-**zvi**-pis-is-w-a.
Mufaro SUBJ.1-PST-REFL-burn-CAUS-PASS-FV
'Mufaro was caused to burn himself.'

(286a) attempts to parallel (285a); the subject of this passive is once again the direct object, and the reflexive replaces the applied object. Similarly, (286b) shows a causee advancing to the subject position with the direct object position being the reflexive. Unlike the object marker equivalents though, there is a change in the grammaticality judgements. The case with the applicative is ungrammatical, while the case with the causative is questionable.[3] An even sharper judgement emerges when the sentences are re-cast as questions:

[3]This causative form was tested three different times with mixed reactions. In a statue context where Mufaro was forced to burn an ugly doll of himself, the sentence was fine. Where there was direct bodily harm, judgements vacillated from ? to *. The question form in the next example was always accepted though.

(287) a. * Ndi-yani a-ka-**zvi**-pis-ir-w-a?
it was-who SUBJ-1-REM.PST-REFL-burn-APPL-PASS-FV
'Who was burnt for himself?'

b. Ndi-yani a-ka-**zvi**-pis-is-w-a?
it was-who SUBJ-1-REM.PST-REFL-burn-CAUS-PASS-FV
'Who was caused to burn himself?'

With these, the patten becomes clearer. A reflexive is acceptable in a passive question where the causative object has become the subject, and the reflexive refers to the direct object, but a question in which a direct object has passivised over a reflexive applicative object remains unacceptable.

While this satisfies the earlier desire for finding a situation where object markers could appear but *zvi* could not, this pattern is not enough to conclude that *zvi* is not a member of the set of object markers. A treatment of *zvi* as a detransitiviser would need to find separate accounts for the morphological evidence that *zvi* is an object marker, not the least of which being that *zvi* is also a class 8 object marker. Rather, (286) and (287) point to the uniqueness of *zvi* within that set: there is some constraint on the usage of *zvi* at play here, making sentences ungrammatical which are fine with other object markers. This constraint against *zvi* is not only present in the passive. Where an unaccusative verb has been extended with the applicative, *zvi* can not replace the applied object:

(288) a. Nda-ka-don-er-a Shingi.
SUBJ.1st-PST-fall-APPL-FV Shingi
'I fell for Shingi.'

b. * Nda-ka-**zvi**-don-er-a.
SUBJ.1st-REM.PST-fall-APPL-FV
Intended: 'I fell for myself.'

Though it may seem odd, the sentence in (288a) can have the meaning that the speaker's falling was somehow in Shingi's best interest, for example in the context of purposefully losing a footrace to let Shingi win. However, (288b) shows that the speaker can not use a parallel expression when the falling is in his or her own best interest, in this case having been offered a bribe to throw a race. Again, this ungrammaticality should reduce to *zvi* as there is independent evidence that this verb can take the applicative, and there is independent evidence that *zvi* can replace an applicative object.

There is another reason that the contrast in (288) cannot simply be written off as a matter of the reflexive being incompatible with an unaccusative. One could argue that (288b) is ruled out because it's not possible to detransitivise an unaccusative. If there was a constraint here on the combination of certain valence operators with the unaccusative, then one would equally expect a passive unaccusative to be impossible. This is not the case:

(289) Kwa-ka-don-w-a.
SUBJ.17-REM.PST-fall-PASS-FV
'There was falling.'

Passivising an unaccusative yields an impersonal passive, describing a situation where falling took place. Bliss and Storoshenko (in prep) treat the subject of (289) as a locative *pro* controlling the subject agreement as a stage topic. Given that an unaccusative can be passivised, (288b) cannot be dismissed as a case of the reflexive applying to a predicate with too few arguments; the language does allow for such structures.

What then is constraining *zvi* in (288b)? Upon closer inspection, this is not so dissimilar from the passive examples above. Because *don* is an unaccusative verb, its subject is not typically agentive. Indeed, the underlying syntax of unaccusatives has been argued to be quite similar to passives, wherein an underlying verbal complement moves into the subject position. Thus, whatever the constraint on *zvi* is, it is not specifically tied to the passive. Instead, this appears to be the first evidence in Shona for a sensitivity to the movement of objects. What ties in this unaccusative analysis with the ungrammatical passive is that in both cases, a lower argument is moving into a position where it would be the antecedent for the reflexive. Recalling the Bliss and Storoshenko analysis of subject positions in Shona, this movement of arguments is an A′ movement, and these ungrammatical cases have the appearance of a crossover phenomenon. A lower element is undergoing an A′ movement into a position where it would be the antecedent for the reflexive which it would not normally c-command. Like *caki* then, the Shona reflexive may be considered a bound variable which requires a c-commanding antecedent.

Crucial to this argument is that the reflexive be interpreted in the original argument position, not the surface location of *zvi*. This can be seen in the schematic representations of (286):

(290) a. Shingi$_i$ a-ka-**zvi**-bik-ir-w-a ApplO$_i$ t$_i$

b. Mufaro$_i$ a-ka-**zvi**-pis-is-w-a t$_i$ DO$_i$

In both cases, the subject derived through passivisation has crossed over *zvi*, but it is only in the first case that the derived subject has also crossed over the original argument position of the argument being replaced by *zvi*. The unaccusative (288b) would have a similar structure to (290a), with the internal argument moving over the applied object. This corresponds best with a Visser-style treatment where instead of treating the object markers as incorporated pronouns, *zvi* would be a manifestation of agreement between a lower topic head (Visser's AgrO), and a covert element in the relevant argument position. Much as the object markers are treated as reflecting agreement with a ϕ-featured *pro*, I propose that *zvi* is a reflection of agreement with a ϕ-featureless covert bound variable. For discursive ease, I continue to refer to this variable as *zvi*, distinguishing the covert variable and the overt realisation of agreement with that variable when relevant. I will return to the question of why the agreement manifests as class 8 *zvi*.

To support a bound variable treatment of *zvi*, it can be shown that *zvi* can be bound under *wh* and quantifiers:

(291) a. Ndi-yani a-ka-**zvi**-bik-ir-a mu-riyo.
it was-who SUBJ.1-PST-REFL-cook-APPL-FV NC3-vegetables
'Who cooked vegetables for himself?'

b. Imbwa y-oga-yoga ya-ka-**zvi**-rum-a.
dog.9 CL9-every-REDUP SUBJ.9-PST-REFL-bite-FV
'Every dog bit itself.'

Sentences such as those in (291) point to an operator-variable structure in which the *Wh* or quantificational expression binds the reflexive.

It should be noted that *zvi* is limited to object positions; quantificational binding of a genitive makes use of free pronouns:

(292) a. Mu-rume w-oga-woga a-ka-bik-a nhanga
CL1-man CL1-each-REDUP SUBJ.1-REM.PST-cook-FV pumpkin.5
r-ake.
POSSD.5-POSSR.1
'Each man cooked his pumpkin.'

b. Nhanga r-oga-roga ra-ka-bik-ir-w-a
pumpkin.5 CL5-each-REDUP SUBJ.5-REM.PST-cook-APPL-PASS-FV
mu-ridzi wa-ro.
CL1-owner POSSD.1-POSSR.5
'Each pumpkin was cooked for its owner.'

For these cases, possession is indicated not *zvi*, but rather by a complex pronoun whose root agrees with the possessed item, and shows agreement with the binding possessor. As shown in the tests determining whether *zvi* could be counted as a valence-changing reflexive, the reflexive is also not used in cases of inalienable possession.

Finally, *zvi* is not the exclusive choice for co-argument reflexivity:

(293) a. Shingi a-ka-bik-a i-ye.
Shingi SUBJ.1-REM.PST-cook-FV PRN-CL1
'Shingi cooked her.'

b. Shingi a-ka-**mu**-bik-a i-ye.
Shingi SUBJ.1-REM.PST-OBJ.1-cook-FV PRN-CL1
'Shingi cooked her.'

For both cases in (293), an ambiguity is reported. While the most natural interpretation is that the pronoun in the object position refers to someone else, a reflexive reading is also possible. That is, it was reported that the pronoun could refer to the subject Shingi without any sense of a Condition B effect. While this reading was judged possible for both sentences, the judgement was that this configuration is improved with object marking (293b), not surprising given that object marking seems to be reserved for topical or discourse-old entities. In light of this data, which shows that a co-referential pronoun can be used in a reflexive context, the role of *zvi* in Shona does not even appear to be one of strict reflexive, but rather one of disambiguation. As an obligatorily bound form, *zvi* is not open to the same ambiguity as a straightforward pronoun would be.

Aside from this limitation to argument positions, *zvi* has one important distinction from Korean *caki*: it is limited strictly to local binding. This can be shown using a simple sentence pair:

(294) a. Mu-rume w-oga-woga a-ka-t-i [Shingi
CL1-man every.1 SUBJ.1-REM.PST-say-FV Shingi
a-ka-**mu**-won-a.]
SUBJ.1-REM.PST-REFL-see-FV

'Every man$_i$ said that Shingi saw him$_{i/j}$.'

b. Mu-rume w-oga-woga a-ka-t-i [Shingi
CL1-man every.1 SUBJ.1-REM.PST-say-FV Shingi
a-ka-**zvi**-won-a.]
SUBJ.1-REM.PST-REFL-see-FV
'Every man said that Shingi$_i$ saw herself$_i$.'

In (294a), the object marker *mu-* in the embedded clause can have a bound reading, with the matrix clause subject as its antecedent, or it can have a free reading, referring to some other person. Conversely, *zvi* in (294b) can only have the clause-bound reading that every man is making a statement about Shingi seeing herself.

Because *zvi* lacks ϕ features, it is not possible to construct sentences in Shona where a local antecedent will be incompatible, as was possible for *caki* in Korean. Rather, this determination that long-distance binding is impossible relies on more subtle judgements from native speakers. For example, the sentence in (295) only has a single reported reading:

(295) Mu-rume wogawoga a-ka-t-i [mu-rume
CL1-man every.1 SUBJ.1-REM.PST-say-FV CL1-man
a-ka-**zvi**-won-a.]
SUBJ.1-REM.PST-REFL-see-FV
'Every man said that the man saw himself.'

This sentence cannot have the bound variable reading where *zvi* is bound from the matrix clause (i.e., that for every matrix clause man, he said that the embedded clause man saw him); the only reading available is one where every man is reporting a state of affairs about one particular man who sees himself. Because of a general tendency for subjects to be referential, it is not possible that each man from the matrix clause is talking about a different man seeing himself. Furthermore, there is no obligatory relationship between the matrix and embedded subjects; all the men could be reporting about some man who is not a part of their group.

The nearest thing to long-distance binding shows up in (296):

(296) Mu-rume wogawoga a-ka-t-i [a-ka-**zvi**-won-a.]
CL1-man every.1 SUBJ.1-REM.PST-say-FV SUBJ.1-REM.PST-REFL-see-FV
'Every man said that every man saw himself.'

While there is no overt subject in the embedded clause, agreement on the embedded verb still indicates the presence of a class 1 *pro* in this subject position of the embedded predicate, providing a local antecedent for *zvi*. If there is any relationship between the two clauses, it would appear to be between the two subject positions, as indicated by the translation offered. There is less a sense that the quantifier binds into the embedded clause than it is somehow reinterpreted in the lower clause. Again, the reading is not that for every man who spoke, every seer saw him.

An attempt to force something akin to non-local binding yielded the sentence in (297):

(297) Mufaro a-ka-**zvi**-won-a [a-chi-dy-a mu-chero.]
Mufaro SUBJ.1-REM.PST-REFL-see-FV SUBJ.1-PROG-eat-FV CL3-fruit.
'Mufaro saw himself eating fruit.'

In this situation, elicited in the context of Mufaro watching home movies of himself, Mufaro is the subject of both predicates, *won* 'see' and *dy* 'eat'. To get this reflexive-like reading, where Mufaro specifically saw himself eating fruit, no special marking appears on the *dy* 'eat' predicate. Rather, *zvi* appears on *won* 'see', as if to make it clear that Mufaro saw himself, and not someone else. This marking disappears when Mufaro is the object of the second predicate:

(298) Mufaro a-ka-won-a [imbwa i-chi-**mu**-rum-a.]
Mufaro SUBJ.1-REM.PST-see-FV dog.9 SUBJ.9-PROG-OBJ.1-bite-FV
'Mufaro saw the dog biting him.'

Here, there is no reflexive on *won* 'see', but the co-reference is carried by the object marker on *rum* 'bite'. Again, knowing that object markers are restricted to topical items, it is not surprising that this would be the preferred method for expressing the intended co-reference. Because of this, *zvi* is not needed in the matrix clause, given that the subject of *rum* 'bite' is specified as an entity distinct from Mufaro. The reflexive returns to the matrix clause (and the object marker on *rum* 'bite' disappears) when the embedded clause of (298) is passivised:

(299) Mufaro a-ka-**zvi**-won-a [a-chi-rum-w-a ne imbwa.]
Mufaro SUBJ.1-REM.PST-REFL-see-FV SUBJ.1-PROG-bite-PASS-FV by dog.9
'Mufaro saw himself being bitten by the dog.'

Again, it seems that *zvi* is compatible with an interpretation where Mufaro saw himself. Note that this is done through a local binding mechanism. There is no strong evidence that *zvi* tolerates anything but local binding, which is unsurprising given that it is restricted to non-subject positions.

One fact about *zvi* which can serve as a potential challenge to the proposed bound-variable account comes from nominalised infinitives, indicated by class 15 marking and agreement. First, such structures can freely take full objects (300a) or object markers (300b) given the correct context:

(300) a. [Ku-remek-edz-a Eric] kwa-kanak-a.
CL15-respect-CAUS-FV Eric SUBJ.15-be good-FV
'To respect Eric is good.'

b. [Ku-**mu**-remek-edz-a] kwa-kanak-a.
CL15-OBJ.1-respect-CAUS-FV SUBJ.15-be good-FV
'To respect him is good.'

These can be uttered as a general statement without reference to a specific person who should take heed. The original context was in a class where Eric was the name of the instructor, thus the statement would cover all students in that class. Given the parallels observed so far between the object markers and *zvi*, it is not surprising that *zvi* is also seen here, but it challenges the binding analysis in that there is no antecedent:

(301) Ku-**zvi**-remek-edz-a kwa-kanak-a.
CL15-REFL-respect-CAUS-FV SUBJ.15-be good-FV
'Self-respect is good.'

Again, there is no sense that the statement need be directed at a particular individual. Rather, this is a generic statement, similar to what was seen in the previous chapter for certain cases of antecedentless *caki*. The same analysis used there can be carried over: a covert generic operator provides the antecedent for *zvi* in situations like this where there is no topic on a nominalised verb stem.

The arguments for the treatment of Shona reflexivity reduce to a covert ϕ-featureless bound variable in an argument position, reflected by *zvi* in the object marker position. One last question needs to be addressed: why *zvi*? Recall that for Shona, along with some of the other Bantu languages discussed, the reflexive marking was identical to class 8 agreement.

In the literature, class 8 is cited as having two distinct functions. First, and unremarkably, it is used for plurals of class 7 *chi-*:

(302) a. **Chi**-garo **cha**-ka-teng-w-a na Peter.
CL7-chair SUBJ.7-REM.PST-buy-PASS-FV by Peter
'The chair was bought by Peter.'

b. **Zvi**-garo **zva**-ka-teng-w-a na Peter.
CL8-chair SUBJ.8-REM.PST-buy-PASS-FV by Peter
'The chairs were bought by Peter.'

Class 7-8 is described by Brauner (1995) as being for "things" including tools, objects of small size, and also for verbal derivations:

(303) *da* "love" (verb stem) → *chido* "love" (noun)

The same basic character for class 7-8 is described in Fortune (1955), though he lists 7-8 as being among the indefinites. He also lists a number of class 8 agreeing deadjectival forms (including one demonstrative) which have an adverbial use:

(304) a. *zvikuru* "great things" → "greatly"

b. *zvakanaka* "good things" → "well"

c. *zvakadayi* "things which did this" → "in this way"

d. *zvino* "these things" → "now"

This is not an exhaustive list of the forms presented by Fortune, but it is representative of the types of adverbial meanings which derive from forms carrying class 8 marking. However, this is not limited to the plural; some of these adverbial forms show up as having derived from class 7 singulars:

(305) a. *choga* "sole thing" → "differently"

b. *chinyoro* "soft thing" → "softly"

c. *chimbishi* "raw thing" → "raw"

Again, this is not an exhaustive list, but note that this is a derivation which spans the class 7-8 spectrum, and is not limited to those forms which are homophonous to the reflexive.

Rather, Fortune describes these as references to abstract items or sets of items having these characteristics. Thus, the frequent antecedentless or generic class 8 forms is not evidence for a widespread generic or impersonal use of the reflexive: rather, this is a trait of the class 7-8 nominals as a whole.

The second use of class 8 has a more direct bearing on the current analysis for the reflexive use of *zvi*: class 8 agreement is employed as a default when there is a clash of ϕ features. For example, class 8 agreement emerges when two entities of different classes are conjoined:

(306) Mufaro na imbwa **zva**-ka-famb-a.
Mufaro and dog.9 CL8-REM.PST-walk-FV
'Mufaro and the dog walked.'

For this conjunction of class 1 and class 9, the resulting agreement is class 8.[4] Casting this back to the analysis of the reflexive use of *zvi*, the overlap between class 8 and the reflexive is no longer surprising. Because it can be bound by any antecedent, regardless of ϕ features, the logical conclusion is that the bound variable underlying the Shona reflexive lacks ϕ features. Showing up in a non-subject argument position, and necessarily being a second reference to some entity, this element is ripe for object marking. Given that when there is a clash of ϕ features on the subject, class 8 emerges, it is therefore not surprising that class 8 object agreement marks entities lacking in ϕ features.

In the next section, a corpus study of Shona will be presented to determine whether the proposed analysis can account for all the data, or whether refinements will be needed.

4.3 Corpus Study

In this section, I detail a corpus study carried out to test the proposed analysis of the Shona reflexive. Compared to English and Korean, relatively few resources are available. A substantial corpus of Shona (over 2.2 million words) does exist. Originally collected by the African Languages Lexical Project (ALLEX), this has now been transferred to the African

[4]The observed *zva* form of the agreement merely indicates past tense; more properly, this should be represented as *zvi-a*, with phonology eliminating the *i* vowel of the class 8 agreement.

Languages Research Institute (ALRI) at the University of Zimbabwe. Early on in considering the approach to this corpus work, I did investigate the ALLEX (now ALRI) corpus, which is available as a web-based resource.

However, the web interface[5] suffers a number of problems making it an impractical tool for this work. With some guessing and testing at the regular expression syntax used in the search command, it is possible to extract from this corpus examples containing the string "zvi" word-medially, but the public interface is limited to 1000 results. Because the results are sorted alphabetically according to the first letter of the word containing the search string, it looks as though out of the entire corpus, there would not be much more than 1000 hits, but still an exact count would be desirable. The more serious limitation though is the display format of the search results. In addition to the word containing the search string, the context returned can be set to the surrounding 30, 40, 50, or 100 characters:

(307) Sample ALLEX output:
m akanga asingazivi kuti kune muporofita anogona kupa mhinduro asi muranda wake aimushandira akange **achizviziva** Areruya hamani, vamwe mune nyembe hameni, hameni munogona kupihwa nyembe muchechi, unogona kuva

Most troubling about this display is that even with the maximum allowable context, the beginning of the sentence containing *zvi* is not recoverable; there is not even any control to make sure the display does not cut off the beginnings or ends of words. Also, there is no translation or morphological tagging available, making a search specifically for reflexive forms impossible. Because of this inability to reconstruct whole sentences 100% of the time, I decided that this corpus would be no more useful than a collection of texts to which I had complete access. Using texts from recited folk-tales and from internet-based news sources, I constructed my own (small) corpus. While it does not match the size of the ALLEX/ALRI corpus, having access to the entire text gives better control over the search and display options.

[5]Still available at http://www.edd.uio.no/allex/corpus/africanlang.html as of July 19, 2010.

4.3.1 Building the Corpus

The corpus derives from two main sources: folk texts and contemporary news stories. The folk tales are from a collection in two volumes entitled *Ngano* (Fortune, 1974). The first volume contains written stories, while the second volume is made up of transcripts of spoken narrative. It is this second volume from which I have drawn a sample of spoken data. Published in 1974 from what looks like a type-written original, these transcripts are not in the best of shape. The first step in their conversion to digital texts began with a PDF scan, then submitted through optical character recognition processing using PDF-OCR-X, a Mac-based program which directly converts PDF files into Unicode-encoded text files suitable for further processing.

However, optical character recognition is by no means a perfect technology even at the best of times, and with these degraded originals the first raw text files contained a number of errors. After examining the files, certain predictable errors began to emerge, such as a common scan of "l<" for "k". Identifying a number of these substitutions allowed for a first automated pass through the text, essentially a complex Find and Replace python script. Still, to ensure accuracy of the digital file, all 78 pages of the second volume of *Ngano* were manually corrected against the original PDF scan. The end result was 30,699 words of text.

By contrast, assembling the news reports was a simple matter. For these, I made use of the news website Voice of America (http://www1.voanews.com/shona/news/), which publishes a new set of stories in Shona four days a week. These were downloaded and saved as text files directly from the site with no inspection. To keep the corpus balanced, a sample roughly equal in size to the spoken component of the *Ngano* collection was used, making for a total corpus size of just over 61,000 words. Admittedly not as impressive as the ALLEX corpus, but this made it far easier to control the output, and with the availability of a steady web resource, this corpus can be built over time.

4.3.2 Corpus Results

Because of the small size of the corpus, it is impossible to make any strong generalisations about the nature of the Shona reflexives, but even with this limited sampling of texts, some important data comes to light. Thus, rather than a full-scale analysis along the lines of the previous English and Korean studies, this work is more exploratory and impressionistic,

looking for patterns in the data which may not emerge in elicitation contexts, simply because one does not know to look for them. In this section, I first discuss the findings related to the reflexive, moving on into a broader discussion of class 8 agreement as it appears in the texts analysed.

Lacking the resources to perform a full morphological tagging, the identification of reflexive forms in this mini-corpus is just as challenging as with the full ALLEX corpus. Using a python script matching against a regular expression, I was able to extract full sentences from the collected texts which contained forms likely to contain reflexive markers. Specifically, the python script flagged any sentence containing a word in which the string "zvi" appeared medially, with at least one preceding and one following vowel. Being in the object marker position, even with the most minimally extended verb form, there should always be at least one syllable preceding and one following *zvi*, so this seemed to be the best way to narrow down the search. The result was that out of the 61,000 words, there were 354 which matched to this regular expression, breaking down with 232 in the news corpus, and 122 in the folk tale corpus. In my further investigation, I concentrated on the folk tale corpus, as these were the texts with which I was most familiar, having read through them all during the text cleanup process.

Of these 122 potential reflexive forms, many were identifiable as not being reflexive. Some of these were class 8 nominal prefixes, as seen earlier on the plural for "chair" *zvi-garo*, which had been further nested within a locative class marker or an associative marker which carries preposition-like functions. Additionally, some of these were non-verbal forms which just happen to contain the *zvi* string. Finally, there were some cases where a verb with the *zvi* object marker was followed by an agreeing class 8 argument, detectable from the class 8 prefixes present on the object, along with agreeing quantifiers and demonstratives.

More likely candidates for reflexivity were those cases where a verb carried *zvi* in the object marking position, but had no indication of a separate class 8 object. Most frequent among these were the verb roots *on* 'see', *nzw* 'hear', and *ziv* 'know'. In many cases, these were followed by complement clauses introduced by *kuti*, which Brauner treats as a complementiser:

(308) a. Dzi-no-**zvi**-ziv-a [kuti...]
SUBJ.10-PRES-OBJ.8-know-FV COMP

'They know that...'

b. Nd-a-**zvi**-on-a [kuti...]
SUBJ.1st.SG-PST-OBJ.8-see-FV COMP
'I (just) heard that..."

Interpreting these as reflexives seems unnatural, as the complement clause should be the verb's direct object; a reflexive interpretation does not make sense in such contexts. One explanation which may be raised is that this could be some sort of emphatic usage of the reflexive (they themselves know), but this is not supported by data from my consultants, where the independent pronouns in Shona fulfill a role similar to the English emphatics:

(309) a. I-ye Shingi a-ka-**zvi**-won-a.
PRN-CL1 Shingi SUBJ.1-REM.PST-REFL-cook-FV
'Shingi herself saw herself.'

b. Shingi a-ka-**zvi**-won-a, i-ye.
Shingi SUBJ.1-REM.PST-REFL-cook-FV PRN-CL1
'Shingi herself saw herself.'

The two sentences in (309) are reported to be synonymous, with the independent pronoun providing an emphatic meaning in both positions. Incidentally, there was significantly less resistance to (309b) than there was with a sentence repeating the proper name sentence finally. I take this to be evidence that this pronoun is not in an argument position, but is rather a post-posed adjunct, similar to what is seen in English. Fortune (1955) further notes that with these independent pronouns, extra emphasis can be indicated through reduplication of the second syllable. Clearly, Shona has a distinct mechanism for expressing emphasis.

What then to make of (308)? If these sentences are neither reflexives nor emphatics, then the most plausible analysis is that the *zvi* is marking agreement with the direct object: the embedded clause. Testing in elicitation shows that subject clauses with the same *kuti* complementiser also trigger class 8 agreement:

(310) [Kuti Obama a-ka-bodir-ir-a]
that Obama SUBJ.1-REM.PST-win-APPL-FV
zva-ka-sham-is-a Shingi.
SUBJ.8-REM.PST-shock-CAUS-FV Shingi
'That Obama won shocked Shingi.'

With no nominalisation present on the subject clause, the resulting subject agreement is class 8. This is something of a surprise, as subordinate clauses are not listed in either the Fortune or Brauner grammars as showing class 8 agreement, yet evidence from both subject and object marking points to class 8 being the agreement for this type of argument. Recalling that class 8 agreement is employed when there is a ϕ feature clash within a coordination, the observed agreement with clauses makes intuitive sense. Because these clauses are not nominalised, they lack ϕ features. Thus, in the same way that mis-matched coordinations agree as class 8 because of conflicting ϕ features, these clauses agree as class 8 due to a lack of ϕ features. The class 7-8 pairing is used for miscellanea, and so it is not surprising that one member of this pairing be employed as a default agreement. Furthermore, if the "core case" of such default agreement lies with the mismatched coordinations, then it is not surprising that the plural class 8 is used rather than the singular class 7.

If this analysis is on the right track, then the homophony between the reflexive *zvi* and the class 8 object marker *zvi* ceases to be a mystery, and is a prediction of the bound variable analysis. As observed at the outset of this chapter, the reflexive *zvi* can take antecedents of any noun class or ϕ-featured personal pronoun. To be bound by all these various forms, the simplest analysis is to assume that the bound variable which contributes the reflexive meaning does not have any ϕ feature valuation itself. Lacking ϕ features like these subordinate clauses, the default agreement for featureless elements emerges: class 8. Re-cast in this light, there is no difference between "reflexive" *zvi* and "object marker" *zvi*, something which is hinted at in Fortune's grammar. This is in line with all of the morphological evidence derived from Kioko's tests that the reflexive *zvi* should be considered an object marker. Further, this is a more satisfying analysis than the one Kioko proposes for the same situation in Kikamba. She speculates that in Proto-Bantu there may have been a full set of reflexives for all noun classes, but that only the class 8 survived. The present analysis makes the exact opposite claim, saying that there is no evidence for a dedicated reflexive object marker, and that *zvi* is, in all cases, class 8 agreement. The reflexive meaning comes from the binding relation between the sentential subject and the covert bound variable which triggers class 8 object agreement by default.

One consequence of this analysis is the possibility for subject reflexives, as there is no formal mechanism blocking the bound variable from appearing in a subject position. While the cases discussed earlier in the chapter all appeared restricted to local binding, and there

was no evidence of *zvi* in a subject position, it is certainly not the case that class 8 subject agreement is a rarity in Shona. If indeed there is nothing special about *zvi* as a reflexive in the object marking position, then it must at least be entertained as possible that some instances of class 8 subject agreement *zvi* or *zva* could be "reflexive-like". To address this question, I again turn back to the corpus.

Searching the corpus for class 8 subject agreement is a relatively easy task, as a first step is to extract any sentence containing a word beginning with either *zvi*, or its past tense equivalent *zva*. The results are that over the 61,000 word corpus there are 3280 such words, breaking down into 1204 from the collected news articles, and 2076 in the folk tales. Again, I will concentrate on the folk tales.

This search again turns up a number of false positives, as there are a number of adverbs in the language which Fortune claims to have been derived from adjectives and demonstratives taking class 8 agreement with some covert element standing for "things". Out of the 2076 instances of a word starting with *zvi/zva* in the folk tale collection, 720 of those are the word *zvino* 'now', a common connective device in the stories. Regarding these adverbs to be frozen expressions, and recalling that Fortune noted some of them to be derived from singular forms, I do not count them among potential reflexives. More interesting are those cases where class 8 agreement is on a verb. Again, as with the search for object marker *zvi*, there are some cases where a clear class 8 subject is present, detectable by the class marker on the noun, and agreement with demonstratives.

There were, however, some cases where class 8 agreement appears to be marked on a verb with no overt subject. One worth particular mention, making up a further 459 of the words beginning with *zvi*, is the form *zvikanzi*.[6] In both the folk tales and news stories, it introduces quoted speech. This somewhat matches the translation provided by Kahari (1981): 'you are told.' However, this translation cannot be a literal translation of the form, based on the morphology. First of all, there is no evidence of a passive, and secondly the final vowel here is *-i*, which generally occurs only in negated contexts in conjunction with the negative prefix *ha-*, not present in this case. Setting aside the final vowel for the moment, it is possible to entertain the idea that this could be better translated as 'they tell you', but this then raised the question of why class 8 agreement is used rather than class

[6]Just looking at *zvino* and *zvikanzi* covers over half of the word initial instances of *zvi-/zva* in the folk tales.

2, used for groups of people. A *caki*-like story of a discourse binder cannot be used here, as the topic position in *zvikanzi* would already be occupied by whatever triggers the class 8 agreement; there is no room for a higher antecedent. Furthermore, the general pattern described in the literature is that continuing topics agree with the noun class of that referent, rather than a generic. Given the sheer frequency of use, the conventionalised function, and unexpected morphology, it may be just safe to classify this as another frozen form.

This is not to say that all instances of class 8 subject agreement can be so easily dismissed. There are some cases in the corpus data where it appears that an infinitive is showing up in conjunction with *zvi* or *zva*. I have been able to replicate this in elicitation using parallels of the "respect" examples from (300):

(311) [Ku-tsva-ir-a mu-mba ma-zuva e-se] **zva**-kakosh-a.
CL15-sweep-APPL-FV CL18-house CL6-day CL6-every SUBJ.8-important-FV
'Sweeping the house every day is important.'

Structurally, this is parallel to (300) in that some property is being ascribed to a nominalised verbal predicate. There though, the agreement was directly with the class 15 deverbal form, whereas here the agreement which comes out is class 8. However, there is a difference between (300) and (311). In (311), there is a sense of repeated action contributed by *mazuva ese*, which suggests multiple events of sweeping. In this light, a plural class 8 form *zvi* is motivated. This can be teased out further with a slight paraphrase of (311):

(312) [Ku-tsva-ir-a mu-mba ma-zuva e-se] **chi**-nhu
CL15-sweep-APPL-FV CL18-house CL6-day CL6-every CL7-thing
cha-kakosh-a.
SUBJ.7-important-FV
'Sweeping the house every day is something important.'

Where "sweeping the house" has been specified as a singular concept, the class 8 marking disappears, and singular class 7 emerges. What this indicates is that there is an active singular/plural dimension to these uses of the class 7-8 subject indicating reference to abstract events or actions, rather than to a generic entity or entities.

Turning back to the question of *zvikanzi*, one more variant of (311) is worth mentioning:

(313) **Zva**-kakosh-a [ku-tsva-ir-a mu-mba ma-zuva e-se.]
SUBJ.8-important-FV CL15-sweep-APPL-FV CL18-house CL6-day CL6-every
'Sweeping the house every day is important.'

(313) is reported to be synonymous with (311), the difference being that the relative ordering of the infinitival and the agreeing predicate has been swapped, possibly in response to a constraint keeping "heavy" elements toward the end of the sentence. This same structure, with class 8 agreement on the predicate, is reported in Chichewa sentences where a subject *kuti* clause appears to the right of the verb (Bresnan and Kanerva, 1989).

Re-examining the *zvikanzi* cases relative to this, a new possibility emerges, and it is the quoted text itself which contributes the class 8 agreement, in the same way as a subordinate clause:

(314) **Zvi**-ka-nz-i, "I-we u-ri ku-famb-a..."
SUBJ.8-NARR.PST-tell-FV PRN-2SG SUBJ.2SG-AUX CL15-walk-FV
"'You were walking..." is told.'

This is a typical example of the use of *zvikanzi*, opening a sentence which contains nothing else but quoted speech. The combination of *zvi* (rather than *zva*) and the tense marker *ka* may be somewhat unexpected, but this is described by Brauner as a narrative past, which is natural in an oral narrative. (314) shows the opening fragment of a relatively long quotation, but often *zvikanzi* is used to introduce single word quotations, often just "yes" or "no". Used with such simple quotations, there is no room for any potential analysis that the class 8 agreement originates within the quotation. As with other clausal elements, the agreement is with the clause itself. *Zvikanzi* can thus be brought in line with the general analysis that *zvi* is used when ϕ features are absent, and the variation between (311) and (312) shows that there can be a sense of a covert singular/plural distinction triggering an alternation between classes 7 and 8 even in the use with an infinitival form.

So, while the corpus may be small in size, some useful data has been brought out. Firstly, there is evidence from the corpus, mirrored in elicitations, which suggests that class 8 agreement is a default used not just for mismatched conjunctions, but for clausal arguments which lack their own ϕ feature specification. This suggests that class 8 has two broad uses: one for direct agreement with a class 8 marked nominal, and another more general use as a default when no other agreement is available. It is under this usage which the reflexive falls; class 8 object agreement is used with the covert bound variable which contributes the reflexive meaning. Finally, corpus data on subject uses of class 8 agreement, again re-inforced with elicited data, suggest that seemingly unexplained occurrences of class 8 subjects can be treated as either agreement with a clausal argument, or as reflecting

agreement with some covert plural, detectable in a productive alternation with a singular form.

4.4 Concluding Shona

In concluding this chapter on reflexivity in Shona, which has developed an argument based on a locally-bound variable, it still remains to compare this form with the original occupant of this slot in the Déchaine and Wiltschko typology, the Romance reflexive clitic. Furthermore, I conclude with some remarks on the state of affairs seen in Shona compared with some other Bantu languages.

4.4.1 Comparison with Romance Reflexives

In looking at the literature on Romance reflexives, it becomes immediately clear that those forms are quite different from the Shona reflexive. First of all, the Romance languages have a full set of ϕ-feature valued reflexives, shown with these samples from Italian:

(315) a. Mi lavo.
1stSG.REFL wash.1SG
'I washed myself.'

b. Ci laviamo.
1stPL.REFL wash.1PL
'We washed ourselves/each other.'

c. Si lava.
3rd.REFL wash.3rdSG
'She washed herself.'

In addition to this distinction between the agreement systems, Russi (2006) reports that in Italian, the plural forms are ambiguous between reflexive and reciprocal readings. This ambiguity is not present in the Shona system:

(316) a. Ta-ka-**zvi**-nzw-a.
SUBJ.1stPL-REM.PST-REFL-hear-FV
'We heard ourselves.'

b. Ta-ka-nzw-**an**-a.
SUBJ.1*st* PL-REM.PST-hear-RECIP-FV
'We heard each other.'

As shown, in Shona the reciprocal is marked with the *-an* suffix, which appears to the right of the verb root in the domain of the valence-changing operators. In theory, there is a potential ambiguity in (316a) in that the *zvi* could be indicating agreement with some discourse-old class 8 referent, but this is not the same ambiguity present in (315b).

The Romance reflexives also show some overlap with the set of characteristics Lidz ascribed to detransitivising reflexives. Among these is the appearance of a reflexive in the derivation of an ergative form from an underlying transitive:

(317) a. Spensi la luce.
turn off.1*st* SG the light
I turned off the light.

b. La luce si spense.
the light 3*rd* .REFL turn off
'The light went off.'

While Manzini (1986) considers this to not be a productive form in Italian, it was already shown earlier in the chapter that *zvi* does not surface with this type of construction in Shona.

Reflexives in contexts of inalienable possession are also attested in Romance languages. For Italian, Russi singles out grooming verbs as having this trait where the reflexive marking appears when the direct object of a grooming verb is a part of the agent's body:

(318) a. Mi lavo le mani.
1*st* SG.REFL wash.1*st* SG the hands
'I wash my hands.'

b. Si spazzola i capelli.
3*rd* .REFL brush.3*rd* SG the hair
'S/he brushes her/his hair.'

Attempts at replicating this kind of structure in Shona have failed to naturally produce a reflexive. When attempting to use the *zvi* in the object marking position on a sentence involving shaving, one of the Shona consultants remarked that it sounded as though the agent was trying to shave himself out of existence, being rased off the face of the Earth. The nearest pattern to the Romance case comes in the following pair:

(319) a. Dennis a-ka-pis-a mu-romo.
Dennis SUBJ.1-REM.PST-burn-FV CL18-lips.10
'Dennis burnt on the lips.'

b. Dennis a-ka-**zvi**-pis-a mu-romo.
Dennis SUBJ.1-REM.PST-REFL-burn-FV CL18-lips.10
'Dennis burnt himself on the lips.'

In (319a), the initial interpretation from the consultant was that it sounded as though someone else's lips were burned. To get the sense that I had burned myself on the lips, the reflexive was obligatory. However, this is partially due to the lack of possessive marking on the phrase which should be acting as the direct object. Instead, *mu-romo* is acting as a locative, indicating where the burning took place, rather than what was actually burnt. Keeping this in mind, the data in (319) does not directly challenge the claim that Shona reflexives do not appear in contexts of inalienable possession.

One trait of the Romance reflexives which does at first glance appear to align with the Shona reflexive is that the Romance reflexives can be used in impersonal constructions:

(320) a. Si dice che piovera.
si says that will-rain
'It is said/ they say/ somebody says that it will rain.'
(D'Alessandro 2007, ex 1)

b. In Italia si mangiano gli spaghetti.
in Italy si eat.3^{rd}.PL the.MASC.PL spaghetti.MASC.PL
'In Italy they eat Spaghetti.' (D'Alessandro 2007, ex 3)

In these Italian examples, the reflexive *si* has an impersonal reading, referring to a generic individual or population. These look somewhat similar to the uses of class 8 subject agreement which arose out of the corpus work. As discussed above though, those cases can fall under a general analysis of class 8 agreement being used as a default, with nothing specifically reflexive featuring in the analysis. Furthermore, Shona makes use of a different mechanism involving the passive for impersonal constructions where an agent is backgrounded:

(321) Kwa-ka-bik-**w**-a (na Mufaro).
CL17-REM.PST-cook-PASS-FV by Mufaro
'There was cooking (by Mufaro).'

The sentence in (321) uses the passive morphology and shows the corresponding demotion of the agent to an optional oblique phrase, but no other argument has moved into the surface position. Rather, locative agreement is used, which can be taken to show agreement with a covert locative *pro*. This type of structure is only licensed when a specific setting is already active in the context, the example here being given in the context of talking about a wedding, and mentioning who did the catering. This agreement pattern, along with the class 7-8 pair are what Fortune describes as the only indefinite forms in the language, and any overlap with apparent reflexive cases is merely a result of the reflexive showing class 8 agreement by default.

So, while the Shona reflexive and the Romance reflexives share some common ground in the Déchaine and Wiltschko typology in that they are both captured under the semantics of bound variable anaphora, the Shona reflexive itself is much more limited in its scope of application.

4.4.2 The Wider Bantu Picture

As remarked earlier in the chapter, Shona is not alone in having this overlap between reflexive forms and class 8 agreement. In addition to the languages previously mentioned, consulting the African Anaphora Project (www.africananaphora.rutgers.edu) shows that a cognate reflexive form also exists in CiNsenga, Ikalanga, KiNande and Kirundi. Further, where data is available, they show a similar pattern to Shona with respect to a lack of reflexive marking in cases of inalienable possession.

Gast and Siemund (2006) make note of another member of the Bantu family, Kinyarwanda, again listing a cognate reflexive form. Interestingly, They note a slightly different pattern from Shona when it comes to emphatic marking:

(322) Nda-shaka ku-vuga-na na [direkteri ub-we].
1st.SG.PRES-want CL15-speak-ASSOC with director self-POSS.CL1
'I want to talk to the director himself.' (Gast and Siemund 2006, ex 54a)

Gast and Siemund report this as a dedicated self-form which acts as the base of this emphatic appositive, not mentioning the possible emphatic use of pronouns seen in Shona. Still, they do list an *ii-* reflexive in the same object marking position as Shona; the *ub* self form does not enter into the co-argument reflexive system.

Again, this is like Shona, as Fortune lists two self forms, neither of which appears to be used for co-argument reflexives:

(323) a. mambo o-mene
chief.CL1 CL1-self
'the chief himself'

b. pa-hu-zima
CL16-CL14-self
'on my own'

The first of these two examples shows a *self* root, similar to Kinyarwanda, but it patterns structurally as a quantifier rather than as a pronominal. As yet, this has not been spontaneously offered in elicitation, and the form does not appear at all in the assembled corpus.

The second example, in light of all that has been said thus far about *zvi* is somewhat more alarming. Here, a nominal root *zima* is given, which Fortune claims to have evolved from an UrBantu form γima. It cannot be overlooked that this does bear a slight similarity to *zvi*, but there is a difference in initial consonants which one would not necessarily expect if these were related forms within the same language. As with *mene* though, there are no instances of *zima* in the corpus to examine, and it has never been offered in elicitation. Based on the strength of the analysis developed in this chapter, and the lack of available data on this *zima* form, I am confident that the similar consonants can be regarded as a coincidence rather than anything that seriously challenges the bound variable analysis developed here.

Chapter 5

Losing an Argument
Valence Reduction in Plains Cree

Almost there...almost there.
-Red Leader. *Star Wars: A New Hope.*

In this chapter, I take a closer look at the data from the language Déchaine and Wiltschko use to exemplify the NP reflexive: Plains Cree. Limited to little more than a review of established literature, this chapter will draw its analysis largely from the account of Plains Cree reflexivity presented in Hirose (2003). Still, this chapter serves three key functions. First, it provides the only example of a language which instantiates reflexivisation through a valence-changing operation. Secondly, the data discussed provides a valuable contrast to the Shona data from the previous chapter, further supporting the argument that reflexivity in Shona is indeed a function of binding, while something quite different is at work in Plains Cree. Thirdly, it will emerge in the forthcoming STAG discussion that Plains Cree presents unique challenges to the formalism. After an examination of the three types of reflexivity Hirose describes, I move on to a brief discussion of the emphatic system in Plains Cree.

5.1 Plains Cree Reflexives

As mentioned in Chapter One, Plains Cree allows for normally transitive predicates to be rendered intransitive through a process of noun incorporation:

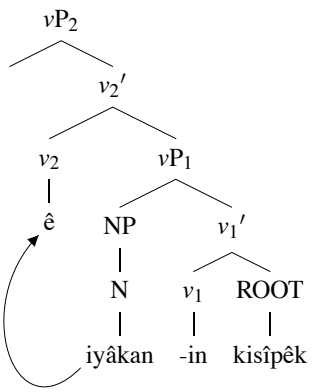
(324) a. kisîpêk-in-**am**(-**w**) wiyâkan.
wash-by hand-I.TH-3SG dish
'S/he washes a/the dish.'

b. kisîpêk-in-iyâkan-ê-**w**.
wash-by hand-dish-INTR-3SG
'S/he does the dishes.' or 'S/he washes a/the dish.' (Hirose 2003, ex 4.1)

In (324a), the content is expressed by way of a transitive predicate, detectible by the presence of two agreement markers on the verb, *-am* and *-w*. Looking to (324b), the *-an* marker for the internal argument is absent, and an intransitive marker has appeared in its place. More importantly though, the internal argument *wiyâkan* is now directly incorporated into the verb stem.

First, in terms of defining the syntactic nature of the incorporated element, Baker (1988) argues that only bare nominals can undergo incorporation. By extension then, the incorporated *(w)iyâkan* should be treated as an NP rather than DP. In conjunction with this, Hirose develops an analysis in which all θ-role assignment is handled either via this noun-incorporation or through *pro* elements which remain within the *v*P. Full DP arguments are treated as adjoined elements along the lines of Jelinek's pronominal argument hypothesis (Jelinek, 1984). Hirose's account for Plains Cree noun incorporation is that the internal argument has moved from its base-generated position, where it is assigned a θ-role, into the immediately dominating v^0 position, occupied by the intransitivising suffix *ê*, where it left adjoins, fusing into the head:

(325)



(325) shows the *v*P domain at the stage of noun incorporation in Plains Cree, before merging of the external argument; the final morpheme order is derived through cyclic head movement with right adjunction of heads. Before this happens though, the internal argument is targeted by the intransitivising head, merging with it in the same type of right-adjunction.

While not a reflexive structure, this is indicative of the type of analysis Hirose develops, with functional heads in the *v*P domain manipulating the argument structure via movement of *v*P-internal NP arguments. Hirose defines three different configurations as reflexive, in that they involve the loss of an argument position without a loss of a θ-role. One of these corresponds to co-argument reflexivity, while another has a reading closer to the exclusive emphatic or EAA of English. A third form straddles these two uses. Each of these will be discussed in turn.

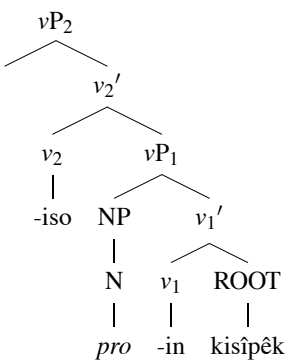
5.1.1 Co-argument Reflexivity

Co-argument reflexivity is signalled primarily by the suffix *-iso*:

(326) pîko-n-**iso**-w.
break-by hand-REFL-3SG
'S/he breaks her/himself (i.e. financially).' (Hirose 2003 ex 7.19a)

The underlying syntax for (326) is given below in (327). While Hirose is not explicit in this regard, the previous account of noun incorporation which showed that the *v*P internal arguments can be considered NPs is retained here.

(327)



The reflexive head is defined as having a raising function; that is, it triggers movement of the *pro* argument from [Spec, vP$_1$] to [Spec, vP$_2$]. In so doing, this argument checks two θ-roles, admissible in Hirose's analysis under a relativised θ-criterion which allows for an argument to receive two θ-roles so long as those roles are assigned in different positions. Contra Déchaine and Wiltschko's 2002 account, this analysis does not present the Plains Cree reflexive as being explicitly one of noun incorporation, but there are nonetheless marked similarities in that both noun incorporation and the reflexive analysis presented here are characterised in terms of NP-movement. Under Hirose's account though, the fact that *-iso* occurs in the same position as an incorporated noun is not indicative of *-iso* acting as an incorporated noun, but rather as a head position licensing a movement of a covert element in the immediately adjacent internal argument position.

Semantically, Hirose makes use of an argument structure reduction operation from Reinhart (1997) which has the effect of reducing the valence of a predicate:

(328) Transitive → Unergative
Argument Structure: V(θ_1, θ_2) → V(θ_1)
Semantics: V′(x,y) → V′(x,x)

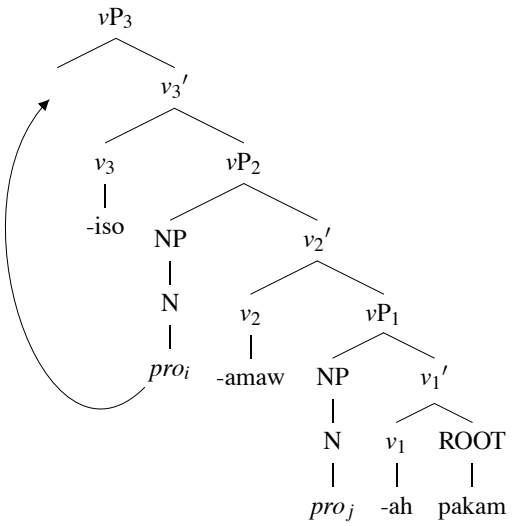
While Reinhart defines this operation in giving an account of the derivation of inherent reflexives (*e.g. The girl dresses*), Hirose argues that it can be applied to the derivation of reflexives as well. Lacking a separate level of argument structure representation in his analysis, Hirose maps this onto the syntax, claiming that the change in the argument structure (the loss of θ_2) is the result of the *pro* argument checking the θ role feature of the raising affix *iso* in the higher position [Spec, vP$_2$]. The semantic correlate of this operation is the observed reflexive semantics which retains the original external argument.

In addition to this account of a reflexive derived from a transitive predicate, Hirose also includes examples showing the interplay of the reflexive *-iso* with the applicative *-amaw*:

(329) ni-pakam-ah-amâ-**so**-n John.
1.SG-hit-by.tool-APPL-REFL-LCAL John
'I hit John for myself.' (Hirose 2003, ex 7.8)

Crucially, this is an unambiguous sentence in Plains Cree. The reflexive can only map the applied object to the subject; it is not possible to get a reading of this sentence under which the speaker hit himself for John. Hirose accounts for this easily in his syntactic analysis of the form:

(330)



The theme, pro_j indexed for *John*, is base-generated as the internal argument, checking the θ-role from the root. A contra-indexed argument, pro_i is base-generated in the position of the applied object then raises to check the θ-role feature in the highest specifier position, triggering the same argument structure reduction operation. In this case though, the two arguments which are mapped to each other are the applied object and the external argument, yielding exactly the reading reported.

Under Hirose's analysis, the blocked reading could only arise if the *pro* indexed for first person were base-generated as the internal argument and then moved up to check the highest θ-role position as the specifier of *iso*. Such a movement would be from [Spec, vP_1] to [Spec, vP_3]; to block this, Hirose invokes an uncontroversial constraint on A-movement. Specifically, having [Spec, vP_2] occupied by the contra-indexed pro_j, placing *John* as the

applied object, blocks the movement of pro_i from [Spec, vP$_1$] to [Spec, vP$_3$]. Under this account, the lack of ambiguity in (329) is predicted. Also, the third possibility which would map the two lower arguments together, is not attested. This is again predicted by the analysis, as the necessary movement would be from [Spec, vP$_1$] to [Spec, vP$_2$]. The reflexive head though, is v_3.

Hirose uses the lack of ambiguity as evidence supporting his account which uses movement to trigger an operation which reduces the valence of the predicate. Under a binding anaphoric analysis, he claims, there should be no lack of ambiguity: either a subject-direct object or a subject-applied object reflexive should be equally possible. Recalling the previous chapter, the equivalent structure in Shona *was* ambiguous; the reflexive could relate either of the two internal arguments to the subject. Hirose's claim then, along with this contrast between the Plains Cree and Shona data, provide further support for the analysis of Shona reflexives as deriving from variable binding. Exactly the ambiguity which Hirose suggests would arise from a binding account of reflexivity is found in Shona; that the variable binding account of Shona reflexives was independently motivated provides further support for Hirose's claim.

This same analysis which derives reflexive meanings from movement within the vP also makes a prediction that reflexive sentences in the language should be strictly local. This is indeed the case:

(331) John nitaweyihtam Mary ê-nipah-**iso**-t.
John know Mary CJCT-kill-REFL-3SG
'John$_i$ knows that Mary$_j$ self$_{*i/j}$-killed.' (Déchaine and Wiltschko 2002b, ex 23)

While the lack of long-distance readings for the reflexives does not further differentiate Plains Cree from Shona, it is an expected consequence of Hirose's analysis.

5.1.2 Dynamic Unaccusatives

The type of structure which Hirose describes as a dynamic unaccusative is given in (332):

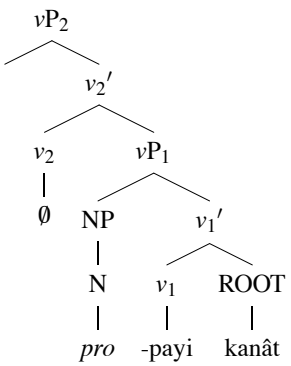
(332) a. pîko-**payi**-w wiyâkan.
break-INCH-INAN dish
'The dish broke by itself.' (Hirose 2003, ex 7.1a)

b. kanâc-**ipayi**-w.
clean-INCH-INAN
'It cleans by itself.' (Hirose 2003, ex 7.17a)

Though using inchoative morphology, these sentences convey a meaning similar to the EAA from English, indicating that there is no external force or causation taking place. While the EAA was concerned with the external argument acting without any outside assistance, the meaning here is that the action is a result of properties inherent in the internal argument. In a sense, this maps the internal argument to be both the cause and theme of a given action.

Hirose gives these structures a similar analysis to that for the reflexives, treating *payi* as another head which triggers an argument movement within the *v*P domain:

(333)



Here, there is a phonologically-null *v* head to which the argument *pro* raises, where a second theta role would be checked. However, note that unlike the reflexives, the *pro* is in a local checking relationship with the reflexive element in its base-generated position rather than its moved position. This configuration leads to a different instantiation of the reduction operation from (328):

(334) Transitive → Dynamic Unaccusative
Argument Structure: $V(\theta_1, \theta_2) \rightarrow V(\theta_2)$
Semantics: $V'(x,y) \rightarrow V'(y,y)$

Using the location (high or low) of feature checking with the reflexive element as the defining parameter, Hirose thus defines two types of reflexivisation: external (328), and internal (334).

5.1.3 Inanimate Reflexives

The *-payi* suffix also arises in reflexives involving inanimates:

(335) a. mêskoc-ihc-**ipayi**-w sîpiy.
change-TRAN-INCH-*inan* river
'The river changes its own direction.' (Hirose 2003, ex 7.4a)

b. kanâc-ihc-**ipayi**-w.
clean-TRANS-INCH-INAN
'It cleans itself.' (Hirose 2003, ex 7.17b)

The sentence in (335a) is described as being said of a river which can wend its way snakily through a valley; directional changes are natural and not the result of dams or other diversions. (335b), distinguished from (332b) by the presence of transitive morphology, is described as being the inanimate equivalent of a sentence containing the reflexive *-iso*:

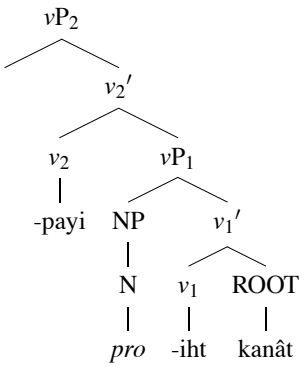
(336) a. kanâc-ihc-**ipayi**-w sîpiy.
clean-TRANS-REFL-INAN river
'The river cleaned itself.' (Hirose 2003, ex 7.3a)

b. kanâc-ihc-**iso**-w awâsis.
clean-TRANS-REFL-3.SG child
'The child cleaned him/herself.' (Hirose 2003, fn 7.3. i)

According to Hirose, the inanimate form *payi-* here can have the meaning of the river cleaning itself through its own "action", such as a swift current clearing silt and debris.

The syntactic analysis of these forms follows exactly the same structure as with external reflexivisation triggered by *-iso*:

(337)



Again, the internal argument raises into a checking relation with *-payi* in the higher position, leading to a reading which preserves the external argument.

As shown, the reflexive structures proposed for Plains Cree are quite different from what has been seen so far. In English, the reflexive pronouns did double duty in a sense, acting as functors on predicates, while adding a referent into one of the argument positions of the predicate. In their non-reflexive emphatic use, the *self* pronouns were similarly treated as functions, though instantiating a simple identity relation. In Korean and Shona, what was seen was a binding relation between an operator antecedent and a variable; in local configurations, this leads to a reflexive-like analysis. For Plains Cree, Hirose's relativised θ-criterion is the key, in that it allows one referent to carry two θ-roles, detransitivising the predicate. With his internal versus external reflexivisation, Hirose is able to capture a subtle difference between reflexive events which arise out of the argument's own internal properties (deriving a meaning similar to the EAA), versus events which arise out of the argument's actions.

5.2 Emphasis in Plains Cree

Because the reflexive meanings in Plains Cree derive from verbal suffixes which license certain manipulations in the argument structure of the predicate, the language lacks any dedicated reflexive pronominal forms. In a sense the same was true of Shona, in that while only one form, *zvi-* was linked with the reflexive, it was not the case that reflexivity was the only use for that form. In this section, I will take a brief look at the emphatics in Plains Cree, which are necessarily distinct from the expression of reflexivity, showing that these pattern somewhat like Shona.

5.2.1 *Wiya* as an Emphatic

According to Blain (1995), the non-reflexive emphatic use of the English *self* pronouns is carried by the form *wiya* in Plains Cree:

(338) a. kiya wiya ki-wî-wîkim-âw John.
2.SG EMPH 2SG.SUBJ-FUT-marry-2SG.SUBJ.3SG.OBJ John
'You yourself will marry John.' (Blain 1995, ex 17)

b. ê-pîkiskwât-ak okimâhkan wiya.
CJCT-speak-1SG.SUBJ.3SG.OBJ chief EMPH
'I talked to the chief himself.' (Blain 1995, ex 11b)

In giving her analysis of *wiya*, the first part of Blain's analysis is to conclude that *wiya* is not a pronoun in the sense of falling under Condition B. Citing Moravcsik (1972), a cross-linguistic study of such emphatics, Blain notes a general trend for such emphatics to be pronominal in nature. Further, she notes that it has been observed that in languages with rich verbal agreement, pronominal forms are used, while languages with little to no agreement use *self* pronouns. Given this connection between agreement systems, and a general lack of reflexive pronouns in the language, one should expect that *wiya* is a pronoun. Taking the examples in (338) to be examples of local binding, Blain concludes that *wiya* cannot be a pronoun as these uses should be Condition B violations. In addition to this binding evidence, Blain notes that *wiya* cannot be conjoined with a referential noun, cannot be focused, and cannot be used deictically. In the end, her conclusion is that *wiya* is a topic-sensitive anaphor.

One part of the argument for topic-sensitivity comes from the observation that *wiya* can only be used with referential antecedents:

(339) * pêyak nâpêw wiya ê-wâpam-â-t.
each man EMPH CJCT-see-DIR-3SG.SUBJ.3SG.OBJ
Each man himself saw her.' (Blain 1995, ex 11b)

Claiming that this sort of quantifier cannot meet the definition of Topic as a discourse-old referent, Blain uses the ungrammaticality of (339) to support the claim that *wiya* is topic sensitive. Further, she defines a hierarchy of topicality based on person features, noting that *wiya* cannot modify a third person referent in the same sentence as a first or second person:

(340) a. ê-pîkiskwât-ak okimâhkan wiya.
CJCT-speak-1SG.SUBJ.3SG.OBJ chief EMPH
'I talked to the chief himself.' (Blain 1995, ex 11b)

b. * Niya ê-pîkiskwât-ak okimâhkan wiya.
1.SG CJCT-speak-1SG.SUBJ.3SG.OBJ chief EMPH
'I talked to the chief himself.' (Blain 1995, ex 16b)

Where the first person subject is made overt with a pronoun, the presence of *wiya* on the third person *okimâhkan* becomes ungrammatical.

In addition to this hierarchy based on person features, Blain also shows that *wiya* interacts with the proximate/obviative system of Plains Cree, which will be discussed in the next section.

5.2.2 Proximates as Emphatic?

In his discussion of cross-linguistic patterns of emphasis, Baker (1995) draws a comparison between the emphatic *self* pronouns of English, and the system of proximate and obviative marking in Plains Cree. Baker claims that the same qualifications that make a contrasted individual eligible for emphatic marking in English are the qualifications which license proximate marking. Specifically, he makes reference to centrality in narrative, which roughly corresponds to Blain's topic. Baker further notes that Plains Cree allows only one referent per sentence to be marked as proximate, and likens this to the pragmatic oddness (though not ungrammaticality) of having two referents bearing emphatics in an English sentence.

Going back to the definition given by Hirose, the proximate/obviative distinction in Plains Cree arises whenever there are two third person arguments of a given transitive sentence. One of them, the more discourse salient, is defined as proximate, while the other is obviative. In unmarked cases, the subject is the proximate:

(341) a. John sâk-ih-ê-w Mary-wa.
John love-TRANS-A.THM-OBJ.3 Mary-OBV
'John loves Mary.'

b. ni-sâk-ih-â-w Mary.
1SG-love-TRANS-A.THM-OBJ.3 Mary
'I love Mary.' (Hirose 2003, ex 1.41a-b)

In (341a), the direct object *Mary* bears the obviative marking. In (341b), this marking has disappeared, which Hirose describes as indicating that in (341b) *Mary* is now proximate. However, the data from Blain suggest that this would only be relative to other third person referents in the sentence; there being none, it is proximate merely by default. Blain's (340b), where an overt first person pronoun made *wiya* marking on a third person entity impossible suggests that on a topicality/salience scale, first and second person still trump

all third persons, a claim Blain makes explicitly. As such, the lack of obviative marking in (341b) does not necessarily translate to proximate status.

Where there is a distinction to be made, *wiya* marking is possible only on proximates:

(342) a. John wiya ê-wâpam-â-t Mary-wa.
John EMPH CJCT-see-DIR-3SG.SUBJ.3SG.OBJ Mary-OBV
'John himself saw Mary.' (Blain 1995, ex 14b)

b. * John ê-wâpam-â-t Mary-wa wiya.
John CJCT-see-DIR-3SG.SUBJ.3SG.OBJ Mary-OBV wiya
'Johh saw Mary herself.' (Blain 1995, ex 14e)

Assuming that *wiya* is not being used redundantly in third person cases, it should be possible to tease apart the function of *wiya* from proximate marking. The function of proximate marking appears to distinctly be one of making a topic overt. Once so marked, *wiya* can be applied. Thus, Baker is right in that proximate marking identifies which elements are open for emphasis, but it is not the proximate marking itself which provides this emphasis; that role is assigned to *wiya*.

While Blain develops a clear system for defining which referents in a given sentence may take the *wiya* marking, she is less clear on exactly the function of *wiya* itself. The glosses provided give a hint as to the intended meaning, but there is no indication of whether this should be construed as the inclusive or exclusive meaning of the emphatic. As non-subjects are eligible, an exclusive reading would appear to be the most likely, but this is not explicitly stated. Furthermore, if the findings with regards to predicate type from English are generalisable here, then an exclusive reading would again be the best fit with the cited data.

5.3 Looking at Texts

While text resources for Plains Cree are even more limited than what is available for Shona, it is possible to assemble a small collection which can be used for data verification purposes. This section reports on such a procedure.

5.3.1 Corpus Assembly

The method for assembling the corpus data used here was largely the same as that used for Shona. The Algonquian Text Society has made available a number of short narratives and speeches in Plains Cree, transcribed and translated in several volumes. While these do not provide morpheme by morpheme glosses, they do reliably parallel the speaker's sentence structure, and give some sense of what the text is about.

Five Plains Cree volumes were included in this analysis: Vandall and Douquette (1987), Ahenakew (1989), Whitecalf (1993), Kâ-Nîpitêhtêw (1998), and Ahenakew (2000). All were scanned as PDF files, and these files were once again run through the PDF-OCR-X software for conversion to processable text files. Because the standardly-used roman orthography for Cree makes use of a circumflex diacritic to indicate long vowels, it was a challenge to find a method of going from PDF to text which would preserve these diacritics. Ultimately, the best solution was to set the optical character recognition to process the text as though it were French (rather than the default English). This had the best results for preserving vowel diacritics; they were not always read as the correct diacritic, but in all cases some diacritic was present in the final output. As with Shona, the text files generated by the optical character recognition software were then processed through a "cleanup" algorithm which corrected as many predictable errors as possible. Still, full manual correction of the texts against the original scans was required. In the end, a collection amounting to 25,547 words of Cree was assembled.

5.3.2 Corpus Findings

Similar to the Shona cases, an algorithm was designed which would extract sentences from the texts containing a given regular expression. In a search for the string *iso*, corresponding with the reflexive morpheme, 75 instances were extracted. Of these, after cross-checking against a Cree dictionary and the English translation, only 25 were in fact reflexives. An interesting observation in the translation from Plains Cree to English was that a form which contained the reflexive in Plains Cree would occasionally be translated as a sentence in English with a possessed direct object:

(343) ni-tipêyim-iso-n.
1SG-govern-REFL-LCAL

'I govern myself.'
Translated as 'I am my own master.' (Kâ-Nîpitêhtêw 1998, p 54)

While this not only reflects the importance of not relying strictly upon translations, it possibly sheds some light on the unexpectedly high number of reflexives found. Assuming the distribution found in this sample stayed constant, a corpus of 1,000,000 words would have almost twice as many co-argument reflexives as was found in English. If it is the case that Plains Cree speakers, for whatever reason, prefer to use verbal reflexives to express concepts which have a parallel form in a possessive, then the high number of reflexives in this corpus is accounted for. This also hearkens back to the original connection between reflexives and noun-incorporation, as it may reflect a general tendency in the language to take advantage of structures which allow for the reduction of arguments.

No instances of reflexives combining with the applicative were found. In a search for the inanimate reflexive, looking for a concatenation of the transitive and inchoative suffixes, 10 sentences were extracted, but on inspection none of these had a reflexive meaning. The string *payi* on its own was found 170 times; given that the morpheme has a non-reflexive use, and given that no inanimate reflexives were found, the instances of *payi* on its own were not analysed.

Looking finally at *wiya*, 84 instances of this pronoun were found, again somewhat more than would be expected if this were simply an emphatic pronoun. Indeed, the evidence is that this pronoun also has a purely referential use:

(344) iyâyaw wiya, wiya piko.
rather EMPH EMPH only
'(He would) rather himself. Himself only.'
Translated as 'Himself first. Only himself.' (Kâ-Nîpitêhtêw 1998, p 128)

This utterance follows a description of a selfish man, and a speculation as to who he would think about upon meeting a young woman. This example follows the analysis from Blain in that *wiya* is definitely being used to refer to a salient topic, but there is no evidence here that any nominal is being modified. In some sense this is more reminiscent of the discourse-bound cases of *caki* from Korean, where a continuing topic could serve as a covert antecedent from an operator position. Also worth noting though is that in translation, these are most naturally equated to the English reflexive pronouns, although in structures

which have more the feeling of exempt anaphors or reflexive pronouns being used in elided contexts, with an exclusive reading.

5.4 Summarising Plains Cree

With the exception of some new data extracted through a small corpus study of Plains Cree texts, this chapter has relied exclusively on previously published work, most heavily the account of Plains Cree reflexivity presented by Hirose. However, as a part of the larger thesis, this chapter plays an important role. In rounding out the Déchaine and Wiltschko typology, Plains Cree serves as the only example in this thesis of a language in which a reflexive meaning is derived from an intransitive syntax and semantics. In this respect, Plains Cree stands contrasts with Shona, which retains a transitive syntax in reflexive contexts, despite expressing reflexivity through somewhat similar morphological means. Further, through forms like the dynamic unaccusative *payi* and emphatic *wiya*, this chapter provides a glimpse at how languages which lack the full set of reflexive pronouns express the non-reflexive uses of the *self* pronouns from English.

Contrary to what was seen with the cases of coargument reflexives, with respect to the emphatic system, Plains Cree comes out looking similar to Shona. While in Shona emphasis was expressed with an appositive pronoun which agreed for ϕ features, in Plains Cree this is done using *wiya*. Forming a natural class based on the use of pronouns as emphatic forms distinguishes Shona and Plains Cree from English, as expected under the earlier-noted generalisation from Moravscik. However, with respect to the formation of reflexives, Shona and Plains Cree are distinct; by comparing the means of reflexive formation and emphasis in this way, a three-way distinction emerges. At one end is English, which uses the *self* pronouns, acting as both referential elements and functions to derive both reflexivity and emphasis. In the middle is Shona, which retains transitivity in its reflexives through the bound variables, but shifts the emphatic function to pronominal elements. Plains Cree expresses reflexivity through an intransitive predicate, but likewise uses a pronoun like element *wiya* for emphasis. Like Shona, Korean stands in the middle of this spectrum, though the picture is somewhat more complicated by the multiplicity of forms. The use of *casin* as an emphatic affix is not dissimilar to *wiya*, but the availability of the *pronoun-casin* forms, which appear to show an English-like multiplicity of functions is unexpected. However, if

it is indeed the case that the reflexive use of the *pronoun-casin* forms is a result of English-influenced usage, then the English-like multiplicity of functions is not that surprising. If the affixation of *casin* is the "core" means of expressing emphasis in Korean, then it patterns with *wiya*.

The close of this chapter represents the close of the language case studies to be presented in this thesis. Over these four chapters, a collection of different semantic forms which can lead to reflexive readings has been built up. From English, there were a number of different implementations of the *self* pronoun as a function on predicates, along with the identity function used in the appositive contexts. Korean and Shona presented two different pictures of bound variable anaphora, with variations in terms of whether or not long-distance binding is permitted. And finally Cree, which derives reflexive meanings from an underlyingly intransitive syntax. In the next chapter, I will show how all of these forms can be modelled in Synchronous Tree Adjoining Grammar.

Chapter 6

Formalising the Forest
Using Synchronous Tree Adjoining Grammar to Model Reflexivity

We never say anything unless it is worth taking a long time to say.
-Treebeard. *The Lord of the Rings*.

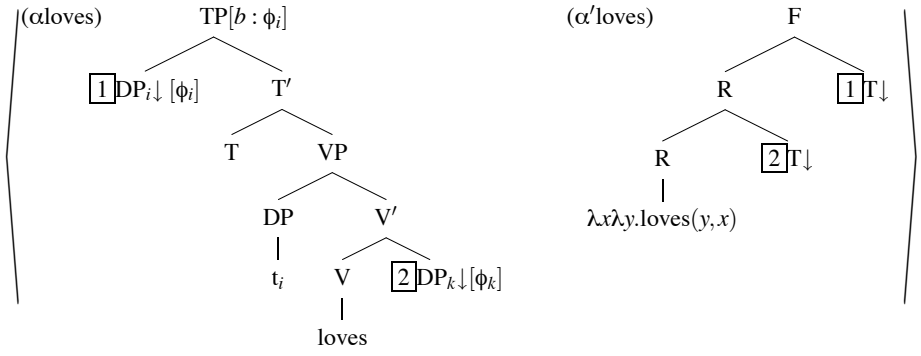
In this chapter, I present a Synchronous Tree Adjoining Grammar (STAG) account of the three major reflexive forms seen in this thesis. The English reflexives have a form in which the *self* pronoun acts as a function taking a predicate as an argument. To capture Korean and Shona, an account of bound variable anaphora within STAG will be presented. This analysis will show how a single treatment of bound variable anaphora accounts for cross-linguistic contrasts in terms of locality constraints using one constraint on the form of the STAG derivation. In a sense, as an intransitive, the Plains Cree reflexive is the simplest syntactically, but presents a representational challenge for STAG, in determining how best to illustrate the change in the argument structure. First, I begin with a presentation of the mechanics of STAG.

6.1 STAG Basics

A Tree Adjoining Grammar (TAG) is a tree-re-writing system. That is, it is a combinatorial system which uses trees, called elementary trees, as its atoms. Elementary trees are combined using two operations: substitution and adjoining. Substitution sites are defined as frontier non-terminal nodes, meaning that they are at the edges (frontiers) of elementary trees, but they are not syntactic terminals. In elementary trees, substitution sites are marked with a downward-pointing arrow, and generally correspond with argument positions. Substitution takes place when an elementary tree with a root node label matching the node label of the substitution site is placed at the empty node. Adjoining requires a specific type of elementary tree, an auxiliary tree, which has a recursive structure. That is, one of its frontier non-terminals (the foot, indicated by an asterisk) will have the same node label as the root node. A recursive auxiliary tree can then target another elementary tree, splicing in at a node whose label matches the auxiliary tree's own root and foot nodes. This adjoining operation can result in a string that has the appearance of syntactic movement, though the observed displacement of elements in the target elementary tree is in fact an effect of the splicing.

Frank (2002) presents a lexicalised TAG, in which each elementary tree is anchored by a single lexical head. Frank proposes a system in which elementary trees are built along GB/Minimalist principles, and then combined using the TAG operations. STAG extends Frank's system by generating a pair of trees for each lexical item, one syntactic and one semantic:

(345)



The tree pair in (345) shows all of the elementary tree notation which will be used through this chapter. (αloves) is anchored by the verb *love*, and is built in accordance with the principles laid out by Frank. Key among these are the TAG θ-Criterion, and the Condition on

Elementary Tree Minimality (CETM). The TAG θ-Criterion has already been introduced, and is reproduced below:

(346) TAG θ Criterion, (Frank 2002):
If H is the lexical head of elementary tree T, H assigns all of its θ-roles within T. If A is a frontier nonterminal node of elementary tree T, A must be assigned a θ-role in T.

As described in Chapter One, this provides a direct connection between the argument structure of a predicate, and the shape of the elementary tree projected by that predicate. The nodes marked with downward-pointing arrows are substitution sites; these correspond to the argument positions. The second criterion deals with the functional heads present in an elementary tree:

(347) Condition on Elementary Tree Minimality, (Frank 2002):
The syntactic heads in an elementary tree and their projections must form the extended projection of a single lexical head.

The CETM introduces the notion of an extended projection, which includes functional heads which co-occur with certain lexical classes. As shown here, the extended projection of VP goes up to TP, and could well go up into the CP domain. Similarly, the extended projection of an NP would go through the ϕP and DP projections. In the syntactic tree, local movement is shown in the movement of the subject from [Spec, VP] to [Spec, TP][1]. This movement is legal within the TAG formalism, provided it remains tree-local.

Some nodes in the syntactic tree are annotated with features in square brackets. These are added to certain nodes, in accordance to the feature-valued TAG described in Vijay-Shanker and Joshi (1988). Where feature values are labelled *t* or *b*, these refer to *top* and *bottom*. For a given node in a derivation tree, there are two sets of feature values, top and bottom. A TAG derivation using these features converges only if, at the end of all TAG operations, the top and bottom features at each node unify. An exception to the statement of nodes having two sets of features can be be found in the frontier non-terminals (argument

[1]This could easily be done with *v*P; here I am following the general tradition of the TAG literature which keeps the syntactic trees as simple as possible. This should not be taken as an explicit claim that *v*P or CP are not present; they are just not shown.

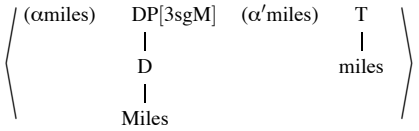
positions), which have top features only. These features constrain the argument positions in that there will need to be feature unification between the features specified in the tree of the θ-role assigning predicate, and the features specified in the tree for the argument which substitutes at that site. The top/bottom distinction comes into play with adjoining. When an elementary tree node is targeted for adjoining, it is effectively split in two; when the auxiliary tree is adjoined, the top features of the auxiliary tree root must unify with the top features at the adjoining site, and the bottom features of the auxiliary tree foot node must unify with the bottom features at the adjoining site. Once all TAG operations are complete, top and bottom features must unify at each node.

In addition to the feature annotation on the syntactic tree, some nodes are marked with boxed numerals. These link to nodes in the semantic tree, and relate to the synchronicity of derivation in the STAG extension of Frank's lexicalised TAG. As formulated in Shieber and Schabes (1990) and Shieber (1994), STAG allows for parallel derivations of two trees provided that those derivations are isomorphic. Note that this is not a constraint saying that the derived trees themselves be isomorphic, but rather that the derivations, recorded in separate structures called a derivation trees, be isomorphic. This isomorphism is further constrained by the links between members of a tree pair, indicated by the boxed numerals. In the trees (αloves) and (α′loves), boxed numerals appear on the substitution sites; what these indicate is that when a TAG operation takes place at a node, a corresponding operation takes place at the linked node in the other member of the tree pair. Thus, when a syntactic argument substitutes in at the [Spec, TP] position of (αloves), the semantic tree associated with that argument will substitute in at the linked node in (α′loves). While various methods for expressing the semantics can be used, here I follow the system presented in Han (2007) which makes use of three node labels: F(ormula), R(elation), and T(erm). Terms are elements of semantic type $<e>$, formulae are type $<t>$, and relations are functions between the two. As in the syntax, arguments are represented as substitution sites. Rather than presenting an elaborate semantic form which more directly mirrors the syntax, this system condenses all the semantic information for an elementary tree (the lexical item along with functional elements) into one unreduced lambda expression. An additional constraint is added in that a derivation can only converge if it is possible for the semantic tree to be composed using operations such as Function Application or Predicate Modification, for example.

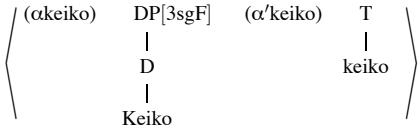
For illustration, the sentence in (348) can be derived using the tree pair from (345) and the pairs presented in (349):

(348) Keiko loves Miles.

(349) a.

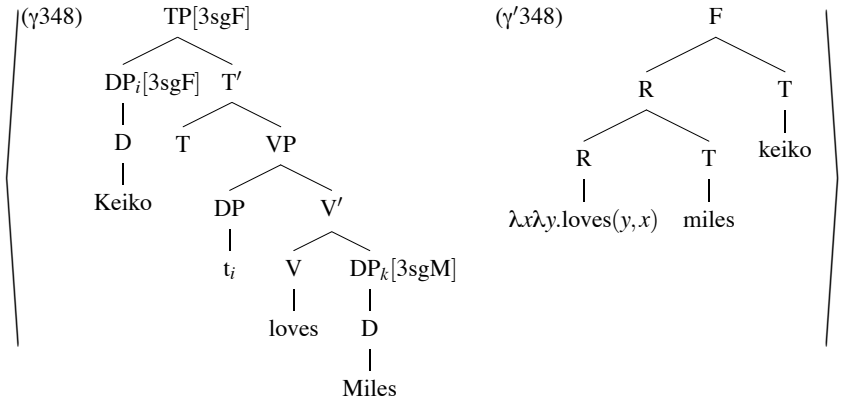


b.



These are simple trees to represent proper name arguments of semantic type $<e>$. They combine with the trees from (345) to derive the trees in (350):

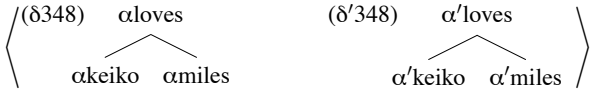
(350)



In the STAG derivation, elementary trees are combined to construct, in parallel, one tree which reflects the syntactic constituency and final string order and another tree which represents the semantic form. These trees are referred to as derived trees. A record of the TAG operations used to derive these trees is also generated, called the derivation tree. Derivation trees for (350) are presented in (351):[2]

[2]The following labelling conventions will be adopted throughout this chapter: Elementary tree names prefixed with α will be non-recursive structures. Recursive elementary trees which participate in adjoining operations will be prefixed with β. The γ prefix will indicate derived trees, while the δ prefix indicates derivation trees. For the derived and derivation trees, the numbers in the tree names will refer back to the example number of the sentence in question. Elementary trees are named according to their lexical anchors. Example numbers will continue to be used to refer to tree pairs when needed.

(351)



In deriving the trees in (350), it is the links on certain nodes in (αloves) and (α′ loves) which ensure that when the tree pairs for Miles and Keiko are substituted in, their syntactic and semantic forms substitute in at the corresponding nodes to yield the correct reading. The steps of the derivation are recorded in the derivation trees in (351). One derivation tree each is produced for the syntax and semantics; crucially, these trees are isomorphic. In other words, the same derivational steps are taken in the syntax and the semantics. The final semantic form of (348) can be derived through semantic composition on the tree (γ′348):

(352) loves (keiko, miles)

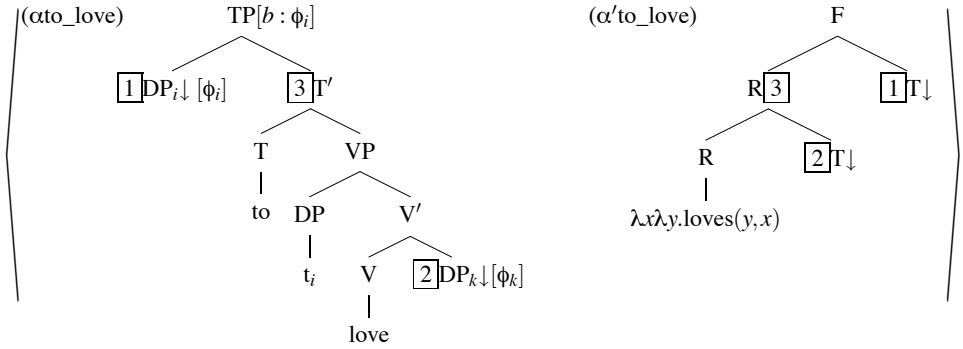
Semantic composition takes place using familiar rules of Function Application and Predicate Modification.

To illustrate adjoining, the second TAG combinatorial operation, a slightly more complex example is required:

(353) Keiko seems to love Miles.

(353) can be derived using the two tree pairs already defined for the arguments, but there is going to need to be a slightly different tree for the *love* predicate, as it is now an infinitive:

(354)



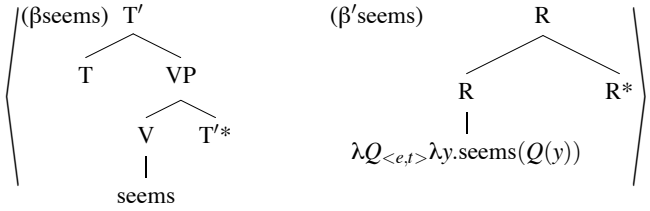
This is almost exactly the same as the tree pair for the finite version; aside from the obvious tense difference, the only new material here is a third linked node between the two trees. The T′ node of (αto_love) is linked to the R node in (α′to_love) that lies between the two arguments. Note that there is no indication of tense in the semantic form. This is primarily out of a desire to keep the semantics as simple as possible, abstracting away from

details which are irrelevant to the matter at hand. None of the examples I intend to present in this chapter depend crucially on tense, so there will be no formal presentation of the semantics of tense. Furthermore, the question of how exactly to represent tense relates to a more formal question underlying STAG which has yet to be addressed; this issue will be discussed in full later in the chapter.

Finally, a tree for *seems* is needed. In the syntax, there is the matter of the interaction between the θ-criterion and the CETM to consider. Having no θ-roles to assign, *seems* will not license any frontier substitution sites; furthermore, lacking a subject argument, the extended projection will not contain a [Spec, TP] position. Taking these facts into consideration, Frank argues for an elementary tree for *seems* which only projects to T′. Further, Frank argues that what *seems* takes as its complement is also a T′ constituent. The result is an auxiliary tree recursive on T′.

In the semantics, *seems* is a one place predicate, taking in a property and returning that same property, embedded in a *seems* predicate that has an almost evidential function in softening the speaker's assertion of the property. Taking this with the proposed syntactic form gives the tree pair in (355):

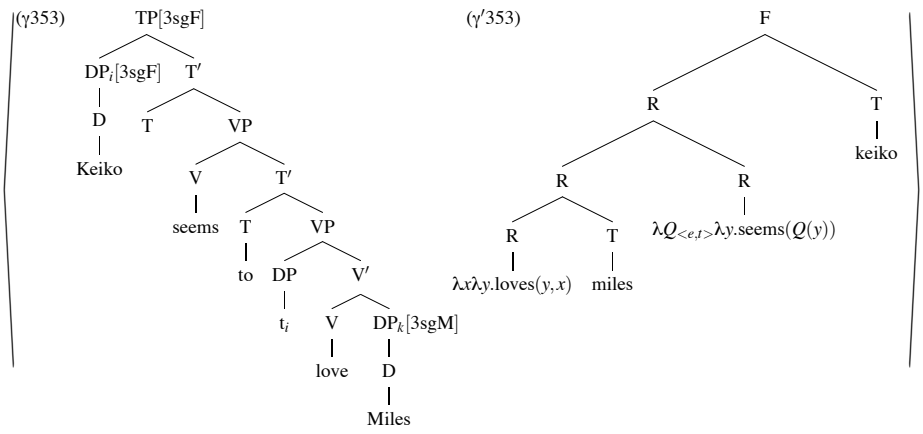
(355)



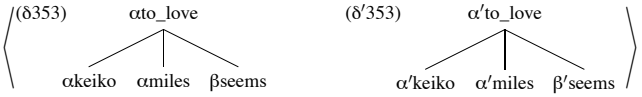
Note that both of these trees are auxiliary trees; thus (βseems) will adjoin into (αto_love) at the T′ node, and (β′seems) will adjoin at the linked node in the semantics side. In adjoining, the targeted node is split apart, the auxiliary tree is inserted in its place, and the tree material which used to be dominated by the target node is replaced, dominated by the foot node of the auxiliary tree. In effect, this is a tree-splicing operation.

The derivation of (353) from the elementary trees established so far yields the derived trees in (356) and the derivation trees in (357).

(356)



(357)



In examining the derived syntactic tree, there is one important difference between the standard raising analysis and the TAG analysis. The derived position of the matrix subject *Keiko* is a result of the adjoining of the matrix clause predicate into the embedded clause. As such, *Keiko* never undergoes any movement beyond the movement out of its base-generated VP-internal position. Frank discusses this issue at length, strongly advocating the movementless account of raising over the GB/Minimalist analysis which has movement from the embedded to the matrix clause. In the derivation trees, the derivations are once again isomorphic, meeting the criterion of synchronicity. Finally, it only remains to calculate the final semantic form of (353):

(358) seems(loves(keiko, miles))

Again, though the types are somewhat more complex, function application is all that is required.

Having worked though these examples, it should now be possible to move into the discussion of the English *self* pronouns with respect to STAG.

6.2 English *Self* Pronouns

In this section, I will cover the major uses of the *self* pronouns in English, including co-argument uses and some cross-clausal cases. In this account that the familiar requirements

from Condition A, that a reflexive pronoun have a c-commanding antecedent within a specified domain, are a consequence of the STAG analysis, rather than an external stipulation. After working through some examples covering ECM and raising, I turn my attention to the emphatic uses of the *self* pronouns. While I show how such uses described in Chapter Two can be adapted to TAG, a full analysis is not provided, as there is not as yet a fully-developed implementation of focus interpretation in STAG.

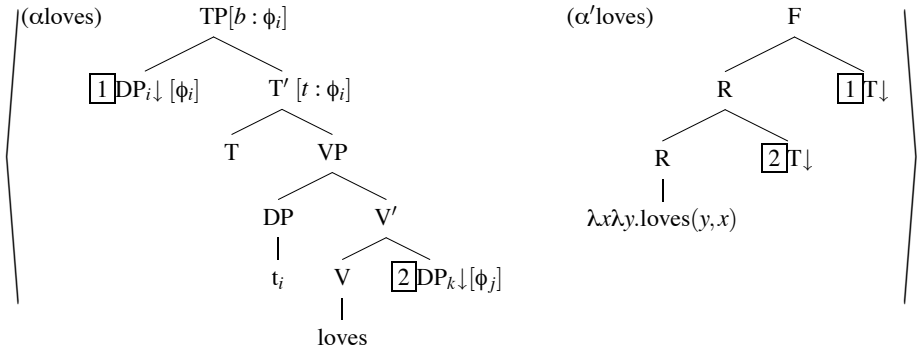
6.2.1 Coargument Reflexives

To begin the illustration of co-argument reflexives in English, I work through a simple example:

(359) John$_i$ loves himself$_i$.

The form of *John* will be a trivial rewrite of the earlier proper name trees. The necessary tree pair for *loves* is essentially identical to the one used before, though with one added syntactic detail:

(360)

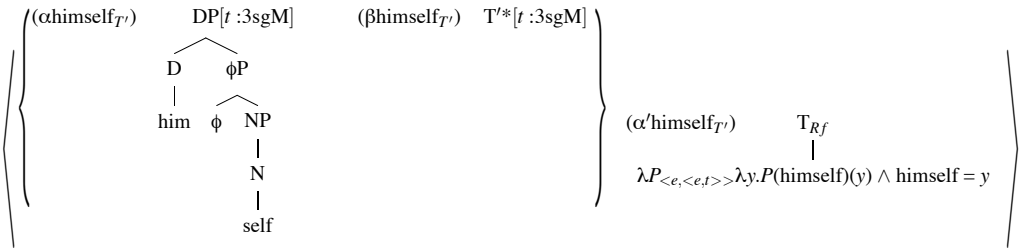


In this slightly revised version, additional feature marking is added. Specifically, the ϕ features of the subject position at [Spec, TP] are also present on the T′ sister of that specifier position. Because these features are all co-indexed, the feature value contributed by the substituted subject will be shared among these nodes. This will come into play with the introduction of the *self* pronoun.

What then are the desiderata for constructing an STAG version of the English *self* pronoun? On the syntactic side, the overt phonological form needs to appear in the correct place, and agreement between the antecedent and the reflexive needs to be observed. In the

semantics, the reflexive pronoun acts as a function, taking a predicate of type $<e,<e,t>>$ and returning a predicate of type $<e,t>$, with the assigned co-reference from Déchaine and Wiltschko encoded. The proposed form is presented in (361):

(361)



Looking first to the syntactic side of (361), the reflexive pronoun is represented as a multi-component set. Multi-component sets can be viewed as one tree in two parts. As used in Kallmeyer and Joshi (2003) to treat English quantifiers, a multi-component set is used in a derivation where a particular lexical item may need access to two parts of an elementary tree. In the Kallmeyer and Joshi implementation, quantifiers are presented as multi-components with one part substituting into the site which contains the overt form of the quantifier, and another defective auxiliary tree which would adjoin at the QR position of the quantifier, specifying its scope. In their most restricted uses, all members of a multi-component set must compose with the same elementary tree.

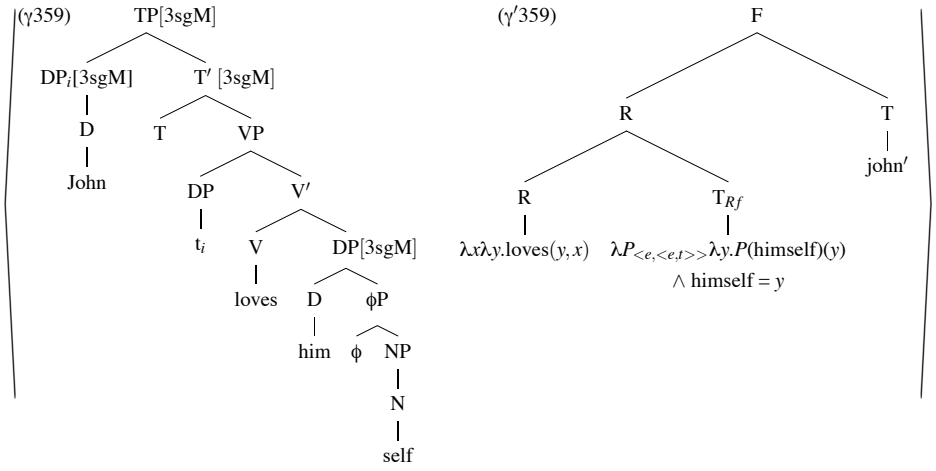
This treatment of the reflexive is similar; in (αhimself), the reflexive pronoun is a DP, carrying third person masculine ϕ features. This tree is what will be substituted into the argument position of (αloves). (βhimself) is a defective elementary tree, containing just one node, the T′. Defective trees meet the definition of auxiliary trees in that the single node is both root and foot; they must have the same label. Crucially, (βhimself) carries the same ϕ features as are present in (αhimself). This will adjoin at the T′ node of (αloves), where the ϕ features will need to unify in order for the derivation to finally converge.

On the semantics side, (α′himself) carries a form very similar to the reflexivising function defined in Chapter Two: the predicate is reorganised, and the referent for *himself* is introduced as one of the arguments of this new predicate. Because there is still a referential component to (α′himself), it is defined as a Term, able to fill a T substitution node in the semantics, rather than having to be of type R. This analysis is developed from that presented in Storoshenko et al. (2008) which had the same syntax, but used slightly different semantic forms. The present analysis is also similar to the syntactic account put forward in

Kallmeyer and Romero (2007), though crucially theirs is not an STAG analysis. Champollion (2008) also presents a version of the Kallmeyer and Romero analysis, again working in a different semantic framework. An alternative STAG implementation, couched in terms of variable binding, is described in Frank (2008). Frank's analysis has a multi-component set in the semantics, with a variable x substituting at the internal argument position, and a function $\lambda x.P(x)$ adjoining at the R node between the two arguments of a transitive predicate. This analysis is open to criticism though, as it is unclear whether the λ operator in Frank's semantic auxiliary tree can bind both x variables. Here, the present analysis is preferred as it preserves the partially-referential character of the reflexive pronouns, as well as remaining more closely unified with the emphatic uses which also treat the reflexive pronoun as a function.

Composing the trees together yields the following derived structures:

(362)

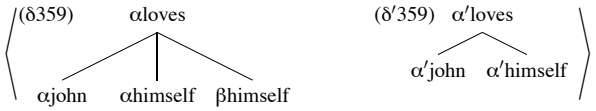


The syntactic derivation proceeds as follows: first, (αJohn) will substitute at [Spec, TP], valuing all the co-indexed ϕ features. Then, (αhimself) substitutes at the direct object position. Finally, (βhimself$_{T'}$) adjoins at the T′ node. As shown, all feature unification in (γ359) is completed without any conflicts. The crucial feature interaction is at T′ where the ϕ features from (βhimself$_{T'}$) must unify with those on (αloves), which are themselves valued by the ϕ features of (αJohn). In this way, the elementary tree from the verb mediates the necessary agreement between the reflexive and its antecedent. For the semantics, the derivation consists of two substitution operations at the linked nodes. Composition from the tree derives the following final representation:

(363) loves(john,himself) $\wedge$ himself = john

With the correct semantic form derived, the only thing left to discuss is the mechanics of the STAG derivation. The derivation trees for (359) are shown in (364):

(364)



A first striking fact here is that the derivations are not strictly isomorphic. This is due to the fact that the reflexive pronoun was a multi-component in the syntax but not the semantics. In cases such as this, an STAG derivation should be constrained to be maximally isomorphic; that is, the matching tree components (αhimself) and (α′himself) should still have matched derivations.

Along with multi-component sets, it is now important to introduce the notion of a tree-local derivation. When first proposed, a TAG derivation which made use of multi-component sets was constrained such that a derivation would only be well-formed if all components of a multi-component set composed with the same elementary tree. Such a derivation is termed 'tree-local'. The current proposed derivation conforms to this constraint in that both (αhimself) and (βhimself) are composing with (αloves). A recent variant of TAG argues that tree-locality is not necessary in all cases. Chiang and Scheffler (2008) define the notion of delayed tree-locality which allows for multi-component sets to be composed more flexibly.

Specifically, they introduce the concept of a delay on tree locality. This delay is defined in terms of derivation tree nodes. The delay for a multi-component set can be characterised as the set of derivation tree nodes separating both members of the multi-component set, tracing a path through the lowest node which dominates both members of the multi-component set, but excluding that dominating node. Thus, the delay for the multi-component set {(αhimself),(βhimself)} is as in (365):

(365) {(αhimself),(βhimself)}

The delay here contains no nodes other than the members of the multi-component set itself; this reflect the fact that the two components were composed with the same elementary tree (αloves). Under the terms defined by Chiang and Scheffler then, a tree-local derivation for a multi-component set will minimally contain two nodes. While not especially relevant to the present discussion of English, tree delays will come to play an important role in the discussion of bound variable anaphora.

As discussed in Chapters One and Two, it will be necessary to define alternative forms of the reflexive to deal with ditransitives:

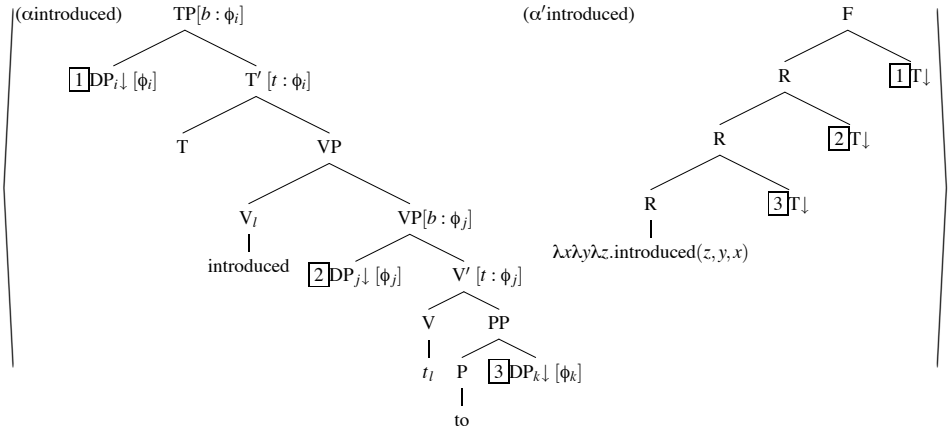
(366) a. John$_i$ introduced himself$_i$ to Bill.

b. John$_i$ introduced Bill to himself$_i$.

c. John introduced Bill$_j$ to himself$_j$.

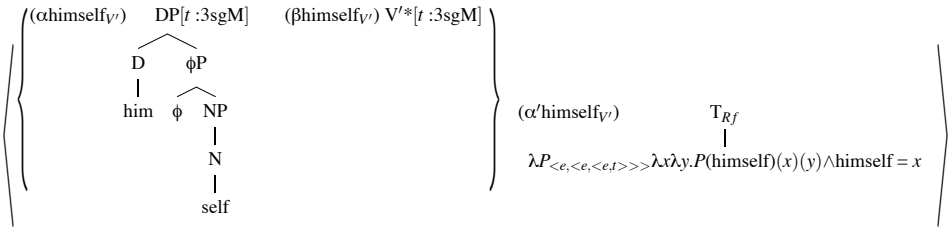
First, a tree pair for the ditransitive verb is required:

(367)



In the syntax, the ditransitive is analysed using a Larson (1988)-inspired VP shell analysis. The sentence in (366a) can be derived using the form of the *self* pronoun which has already been defined. (366b) can be derived using a subtly-altered definition which would have an identical syntax, but a semantic form from $<e,<e,<e,t>>>$ to $<e,<e,t>>$ making the identity relation hold between *himself* and the highest argument. Most interesting here is the case in (366c) which will call for a version of the *self* pronoun which is changed in both the syntax and semantics:

(368)



With this modified form of the reflexive, the two internal arguments of the ditransitive predicate are mapped into the assigned co-reference equation. Otherwise, the derivation proceeds along the same lines as for (359). The fact that (βhimself$_{V'}$) is a V′ node, rather

than a T′, controls where this form can be used, thus ensuring this formula can only access two internal arguments.

The present analysis also blocks ungrammatical cases such as that in (369):

(369) * Sue introduced himself$_i$ to John$_i$.

In principle, two different forms of the reflexive pronoun could be used to derive this example. The first, presented in (361), would take the subject in [Spec, TP] as its antecedent. Such a derivation would be blocked due to a clash of ϕ features. At first glance, the form in (368) also appears to be a viable candidate; certainly there is nothing in the syntax constraining this. (αhimself$_{V'}$) could substitute into the middle argument position, link 2, of (αintroduces), and (βhimself$_{V'}$) could adjoin at the V′ node, as the substitution at [Spec, VP] would provide a compatible valuation of ϕ features to permit the adjoining. Following the links over to the semantics though, a problem emerges. The linked T node in the semantics tree where (α'himself$_{V'}$) is not sister to a predicate of type $<e,<e,<e,t>>>$, the necessary input for the reflexive function. As the eventual derived semantics tree would not be composable, this derivation would crash. Here then, it is actually the semantics which blocks a derivation which would normally depend upon a c-command constraint to rule out (369). As argued in more detail in Storoshenko et al. (2008), no structural requirement (dominance or c-command) between the two components of the syntactic multi-component set for the reflexive is required to block ungrammatical derivations; this can all be handled either through feature agreement, or through the semantics.

6.2.2 ECM and Raising

Recalling the discussion from Chapter Two, exceptional case marking appears to allow a *self* pronoun to be bound across a clause boundary. In mapping out the proposed semantics for a *self* pronoun in such a case, the following form was defined:

(370) a. Sandy$_i$ wants herself$_i$ to win the round.

b. $[\![$to win the round$]\!] = \lambda z.\text{win}(z,\text{the round})$
$[\![$herself$]\!] = \lambda Q_{<e,t>}\lambda P_{<t,<e,t>>}\lambda x.P(Q(y))(x) \wedge x = $herself
$[\![$herself to win the round$]\!] = \lambda P_{<t,<e,t>>}\lambda x.P(\text{win}(y,\text{the round}))(x) \wedge$
$x = $herself

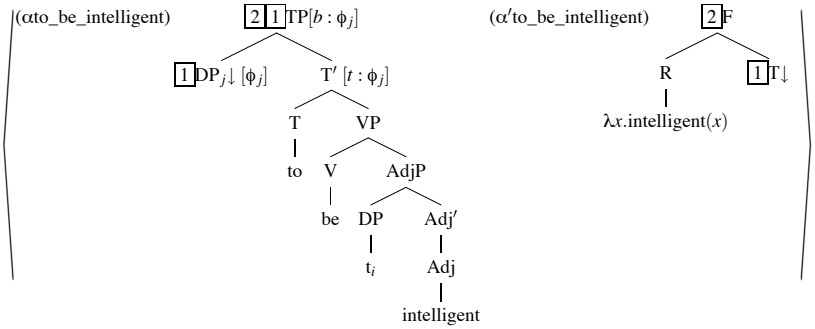
Here, the *self* pronoun again takes a predicate as an argument, mapping the assigned co-reference to the argument of what will be the matrix predicate. For the sake of simplicity, I work through an example with an intransitive embedded clause:

(371) Stephen$_i$ believes himself$_i$ to be intelligent.

To see this in the STAG context, new elementary tree pairs must be defined. Again, the tree pair for *Stephen* will be a trivial pair for a proper name mapped to a term of type $<e>$. Most important will be the trees for the two predicates, and the reflexive.

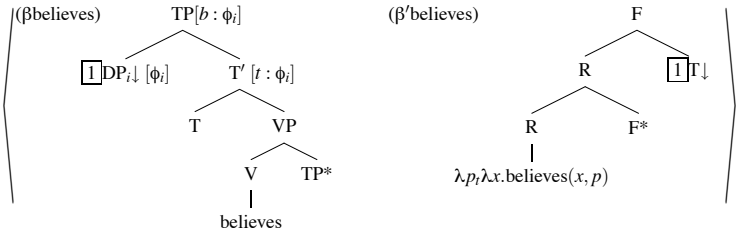
First, the tree pair for *to be intelligent* is going to be similar to what was seen in the raising example from the start of the chapter:

(372)



As was seen before, this is an infinitival tree projecting up to [Spec, TP]. The matrix predicate can be constructed as in (373):

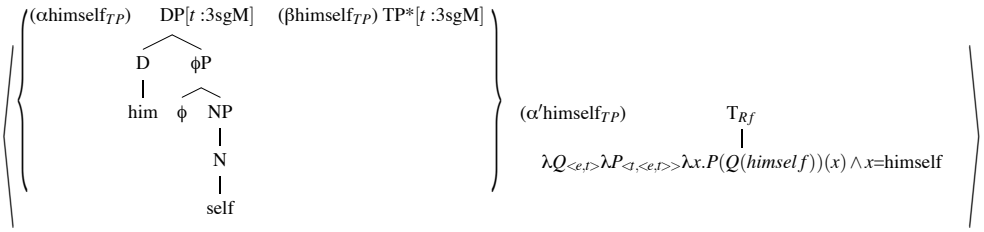
(373)



Once again, the matrix predicate is instantiated as an auxiliary tree. Different from the earlier raising tree though, this structure is recursive on TP, reflecting the fact that *believe* takes a full clause as its argument. This is mirrored in the semantics, where the tree is recursive on F, not R.

Finally, a version of *himself* for subject positions is required:

(374)

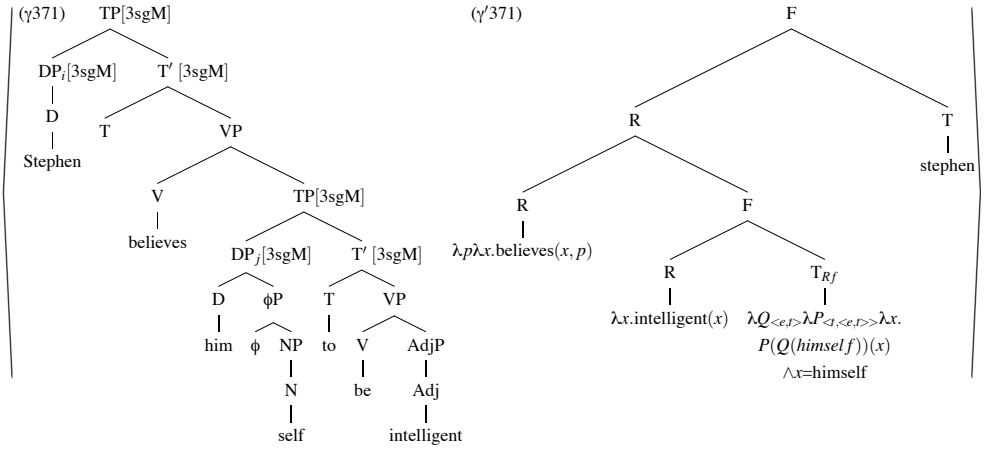


Here, the syntactic form of the reflexive contains the familiar ϕ features. (βhimself$_{TP}$) adds a top feature to the root node of (αto_be_intelligent). When this node is split, that top feature must unify with the features of the root of (βbelieves), which have been valued by (αStephen). It is crucial that the feature which controls the agreement between clauses is contributed by the *self* pronoun. If the features on the two predicates themselves had to unify, then no derivation with ϕ-feature mismatched arguments of the two clauses would be possible, ruling out grammatical cases such as (375):

(375) I believe you to be intelligent.

On the semantics side, the referential portion of *himself* becomes the argument of the embedded predicate while the output of the reflexivising function is itself another function which will also take the matrix predicate as its argument. Co-reference is mapped between the arguments of the two predicates; in essence the reflexive is acting as a bridge between *believes* and *to be intelligent*. The derived forms are presented in (376):

(376)

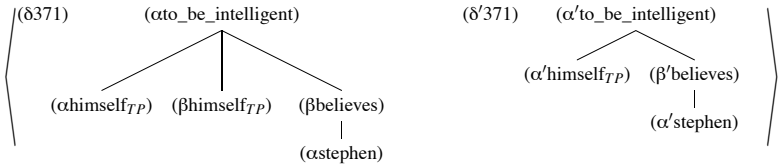


Completing the semantic composition on the derived tree in (γ′371), the final semantic form which results is shown in (6.2.2):

(377) believes(stephen, intelligent(himself)) ∧ stephen = himself

The form of the derivation is presented in (378):

(378)



Note that in the derivation tree, the only possible derivation is the one in which (β′believes) is a dependent of (α′to_be_intelligent). On the syntactic side, it would be possible for (βbelieves) to adjoin into (βhimself$_{TP}$), formalising in the derivation the bridging function being carried out by the reflexive, but such an adjoining is not possible in the semantics. As such, the maximally isomorphic derivation is presented.

The last example I present in this section is a holdover from Chapter Two:

(379) Rich$_i$ seems to himself$_i$ [t$_i$ to outperform his rivals.]

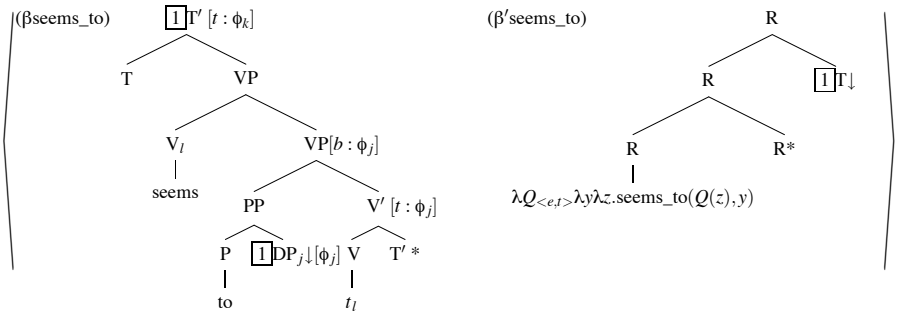
This was the case where the experiencer of *seems* is a reflexive pronoun whose antecedent is the raised subject from the embedded clause. In Chapter Two, I pointed out that because the embedded clause has no open argument positions, there is no open variable in the semantics with which some re-mapping can be carried out. Cases such as this were what led Reinhart and Reuland to ultimately use a structural constraint, the Chain Condition, to rule out some instances of reflexive pronouns.

There is a fundamental difference in the way that TAG (and STAG) handle raising examples: the appearance of movement is derived through adjoining of the raising predicate into the embedded clause tree. In the semantic derivation, the raising predicate will compose with the embedded predicate *below* the embedded clause subject. This means that examples such as this can be handled quite easily in STAG, without recourse to any constraints. Tto keep things simple, I again will illustrate with an intransitive embedded clause:

(380) Rich$_i$ seems to himself$_i$ to be intelligent.

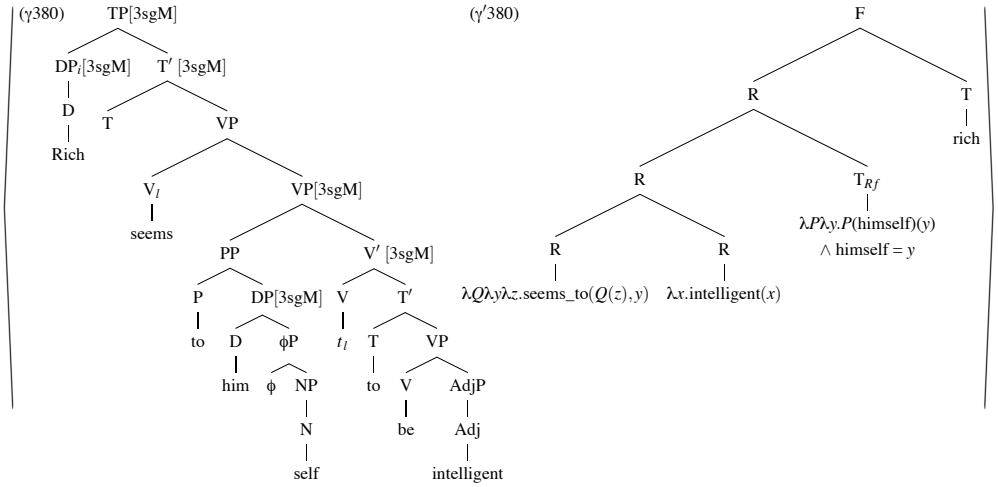
The same *to be intelligent* tree from the previous example can be recycled, and once again the tree for the proper name will be a simple variant of the ones seen before. A new tree pair for the raising predicate is required though:

(381)



As with the earlier version of *seems*, this is an auxiliary tree in both the syntax and the semantics. The major change here is the addition of the experiencer argument. For the *self* pronoun, no new form is needed; the original simple version of the reflexive pronoun used in the transitive sentence works here as the experiencer argument of *seems to*. Derived trees are presented in (382):

(382)

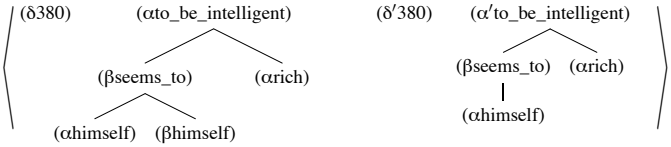


After all semantic composition, the form which emerges is:

(383) seems_to(intelligent(rich), himself) ∧ himself = rich

The derivation trees are as in (384):

(384)



Once again, the maximally isomorphic derivation is shown. It is in this case that the STAG analysis shows its superiority over other methods, not as a result of its method for the treatment of the reflexive, but as a result of TAG's unique method for analysing raising without

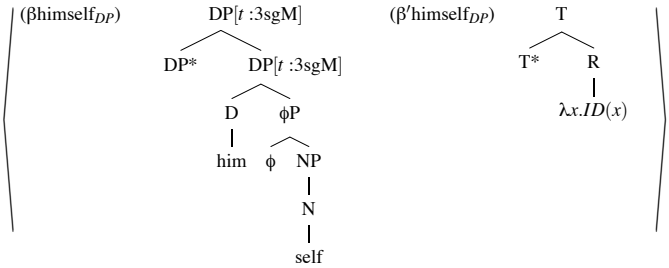
movement. This allows for the completion of the semantic analysis from Chapter Two, finally showing that co-argument reflexivity in English can be unified under this semantic account of the *self* pronouns. Taking this in combination with the earlier observation that the c-command relation between a reflexive pronoun and its antecedent falls out of the semantic analysis further highlights the suitability of STAG for this kind of analysis.

6.2.3 Emphatic Uses

As discussed in Chapter Two, three different emphatic forms of the *self* pronoun can be identified, two having an exclusive reading, and one having an inclusive reading. While a full implementation of these forms in action would require a significant extension of STAG into the domain of focus, a first step can be presented in constructing some possible forms.

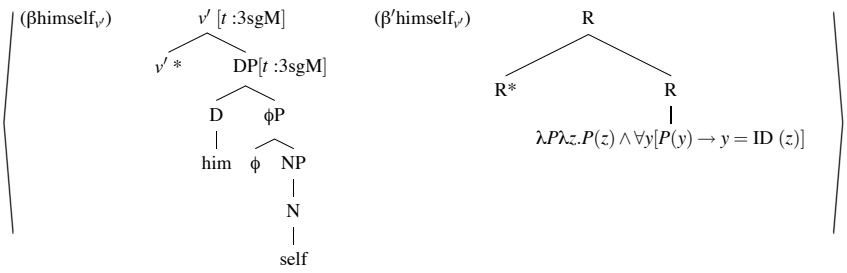
As proposed in Déchaine and Wiltschko and formalised by Gast, the emphatic use of the *self* pronoun has at its core an identity function $\lambda x.ID(x)$ which just returns its argument. The simplest form of this, the exclusive adjacent form, attaches directly to the DP it modifies, regardless of where that DP appears in the sentence:

(385)



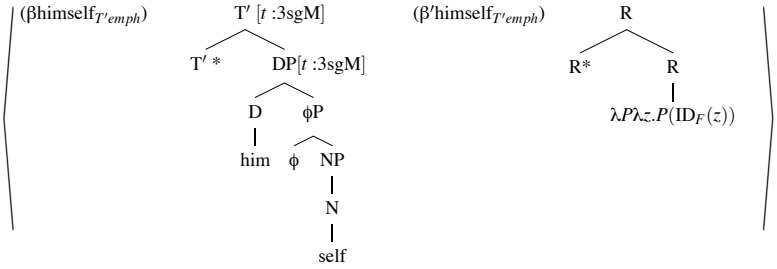
The sentence-final equivalent, the exclusive extraposed form, was shown in Chapter Two to be sensitive to theta roles, specifically connected with the external argument, suggesting a connection with the *v*P projection. This structural position was further supported by the observation that this exclusive reading appeared to be within the scope of sentential negation. Thus, the proposed semantics for the exclusive extraposed form, which borrowed from the semantics for the exclusivity particle "only", can be adapted to STAG as shown in (386):

(386)



Lastly, there was the inclusive form, which could vacillate between an adjacent or an extraposed position, but consistently appeared outside the scope of negation. Because this form is sensitive to subjects only, regardless of θ-role, the most logical placement is as an adjunct to T′:

(387)



At first glance, the forms in (385) and (387) may look similar, but recall that the difference between them will depend crucially on the contextual background. When focused, and alternative to the ID function is presupposed, yielding a presupposed set of alternative referents. If the contextual background has already established the existence of such alternatives, then an inclusive reading emerges. If no such alternatives have been established, then the exclusive meaning comes through. Also, a second version of (387) is required, reversing the two frontier nodes on (βhimself$_{T'}$) to derive the word order placing the emphatic immediately after the subject. For all three of these proposed trees, the adjoining site would be sufficiently close to the antecedent that no secondary defective tree is required to handle agreement. (385) adjoins directly into the DP with which it will agree, and (386) and (387) adjoin at the sister nodes to the agreeing DP, positions which have already been shown to carry the required ϕ feature valuation. A full formal semantic analysis must await the definition of alternative semantics for STAG.

In examining these three forms, there is one striking difference (aside from the obviously different semantics) between these forms and those used for co-argument reflexivity. In none of these cases is there a tree which can be substituted into another; all are auxiliary

trees which are generally hallmarks of modification rather than argumenthood. Given the fundamentally different function of the emphatic *self* pronouns, this difference should be expected, and provides a clear-cut distinction between the two uses.

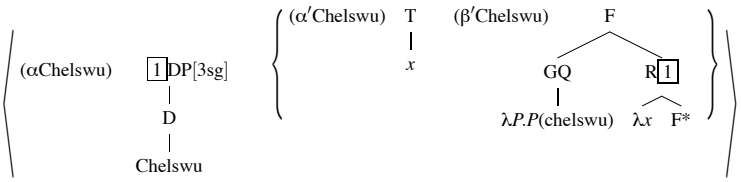
6.3 Parametrizing Bound Variables

The reflexive forms for Korean and Shona are both analysed as bound variable anaphora. Unlike the *self* pronouns, which have had some history of analysis in TAG and STAG, there is no prior work on which to base an STAG analysis of these forms. First, I begin with an analysis of a local instance of *caki* from Korean:

(388) Chelswu$_i$-ka caki$_i$-lul pinan-ha-yess-ta.
Chelswu-NOM self-ACC blame-do-PST-DECL
'Chelswu blamed self.'

For Korean, recall that the final analysis was that all subjects should be treated as generalised quantifiers. As such, the simple proper name forms from the previous section on English will not do. Rather, *Chelswu* will need to be represented as shown in (389), using forms parallel to those defined for English in Han et al. (2008):

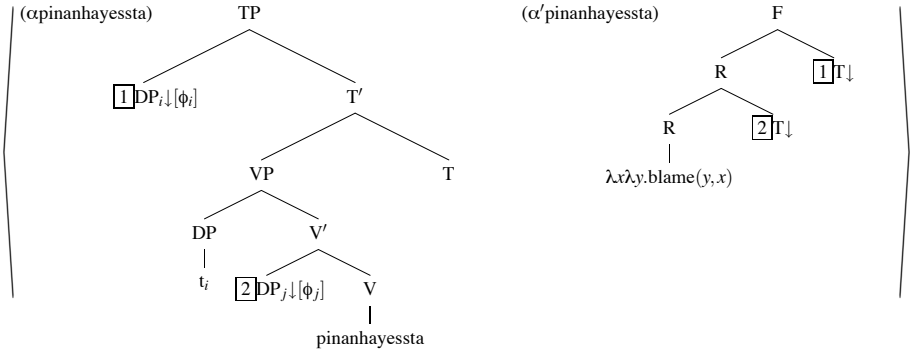
(389)



The now-familiar concept of the multi-component set re-emerges, though here only the semantic side is represented as a multicomponent. On the syntactic side, the GQ version of *Chelswu* is represented as a simple DP; this simple representation conceals the internal structure consisting of ϕP and NP, which is still assumed to be present. In the semantics, there are two trees, one a variable which will be substituted at an argument position, and the other containing the generalised quantifier portion which is an auxiliary tree recursive on F. The required structural relationship between (α′Chelswu) and (β′Chelswu), that the variable must be in the scope of the binder, can be captured under a more general constraint against unbound variables.

The elementary trees for the transitive predicate contain no surprises:

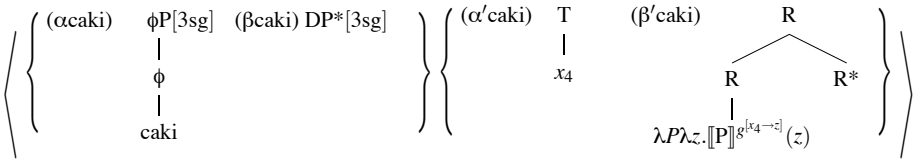
(390)



Again, the syntactic tree (αpinanhayessta) abstracts away from some finer details, most notably the derivation of the verb form with tense and mood suffixes. Another issue comes in the formation of the argument substitution sites, particularly the internal argument where *caki* will be substituted. The syntactic form for *caki* is a ϕP, but specifying the substitution site as such would block the possibility for DP to substitute at that node. For the sake of simplicity, I continue to leave open argument substitution sites labelled as DP, but it should be understood that a ϕP is equally valid for this position[3]. On the semantics side, a simple transitive predicate is presented, with the familiar links between the argument positions of (αpinanhayessta) and (α′pinanhayessta).

Finally, a tree pair for *caki* is required. This will be a formalisation of the bound variable analysis argued for in Chapter Three:

(391)



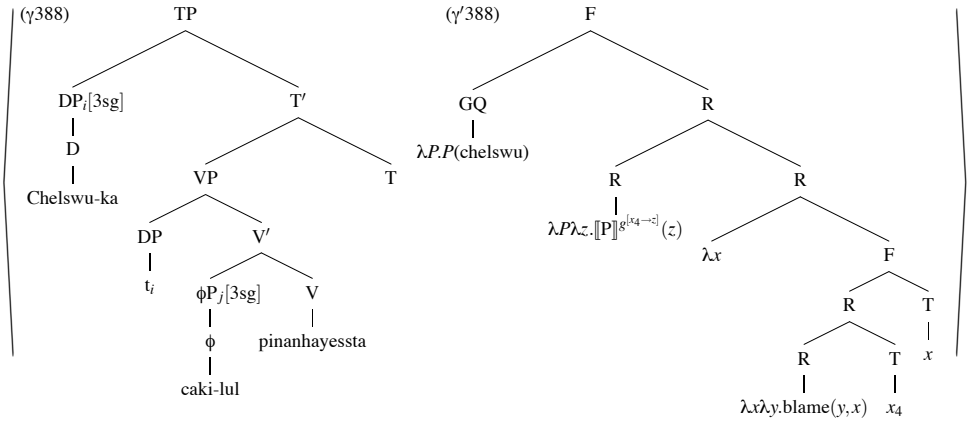
As shown in (391), *caki* is represented as a multi-component set in both the syntax and semantics. (αcaki) contains the overt phonological form, and will substitute at an argument position, while (βcaki) is a degenerate DP tree which carries the third person ϕ feature specification for *caki*. On the semantics side, (α′caki) is a variable which will substitute in at a T node. The mechanics of variable binding is housed in (β′caki), which will adjoin into the elementary tree of *caki's* antecedent. The formula in (β′caki) is a shorthand version of the Binder Index Evaluation Rule (BIER) from Büring (2005), collapsing into one step

[3]It might make more sense to use an underspecification type argument, labelling the substitution site as ϕP, with the further statement that anything containing a ϕP node can substitute there (covering ϕP and DP), but I am erring on the side of making the elementary trees look as intuitive as possible.

what Büring does in two. Büring's formulation of variable binding has a binder index as a sister to a predicate, and his BIER is defined as an interpretation of just that semantic structure. The form presented here does away with the separate variable binder and treats the BIER as a function of its own without needing a new interpretational rule.

Composition of these trees yields the derived trees in (392):[4]

(392)

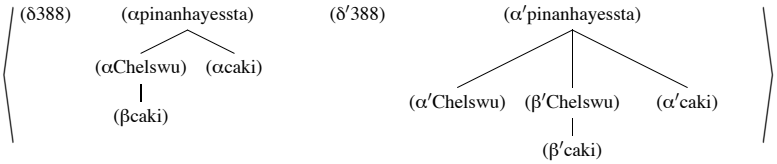


With the full derived trees, the semantic function of the (β′caki) tree is clearer. Walking up the semantic derivation, the lowest F node is a type $<t>$ formula with both of the argument variables having undergone λ-conversion. Then, the *lambda*-operator contributed by (β′Chelswu) binds the variable from (α′Chelswu). Next, the contribution of (β′caki) manipulates the assignment function on its sister, converting all instances of the *caki* variable to a new *z* variable. This same *z* variable is also introduced into the one place predicate created by the binder portion of the (β′Chelswu) tree. The result is a reflexive semantics:

(393) blame(chelswu, chelswu)

The derivation trees for (394) show that this is the most complicated derivation yet:

(394)



[4]Another thing not shown is case assignment. Frank (2002) provides a description of case assignment which also makes use of features, and having elements enter into familiar checking relationships. This is fully implementable here, but is not shown to avoid cluttering up the trees. For this reason, the case markers appear on arguments in the derived trees as if by magic.

This is the first derivation so far which has involved the non-local composition of a multi-component set: (αcaki) composes with (αpinanhayessta), while (βcaki) composes with (αChelswu). A parallel derivation takes place in the semantics with the added wrinkle of the introduction of a multi-component set for *Chelswu* as well. Isomorphism is maintained in that on both sides of the derivation, (βcaki) composes with the *Chelswu* tree set; that set just happens to be a singleton set in the syntax. The delays in syntax and semantics for the *caki* multi-component sets are shown in (395):

(395) a. Delay in Syntactic Derivation:
{(αcaki),(βcaki),(αChelswu)}

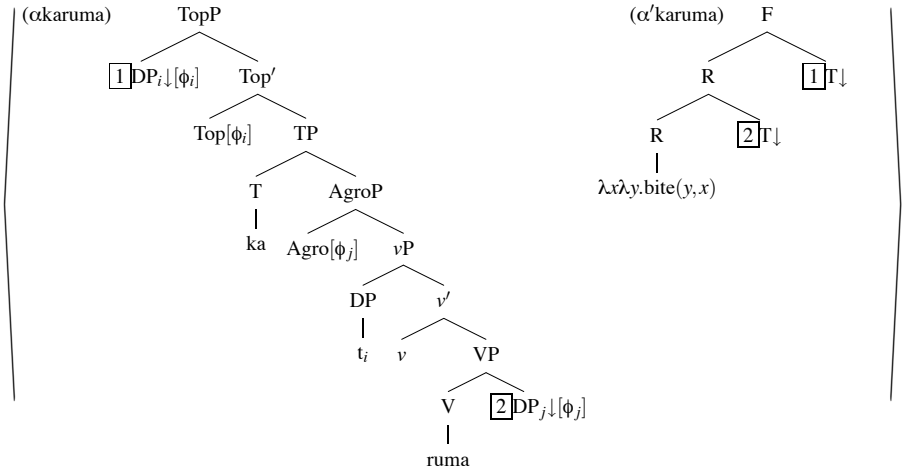
b. Delay in Semantic Derivation:
{(α′caki),(β′caki),(β′Chelswu)}

Recalling that in the semantics *Chelswu* was also a multi-component set, there is a second delay in the semantics as well, consisting just of the two members of that multi-component set. Because (β′Chelswu) participates in the delays for *Chelswu* and *caki*, this derivation is described as having two simultaneous delays, according to the terms defined by Chiang and Scheffler. Looking at the delays in (395), these both have a cardinality of three, which can be seen as the minimal non-trivial delay. However, while this has the appearance of having crossed a threshold in terms of derivational locality, a look back at the syntax shows that this is still local binding; assuming a generalised quantifier analysis which can interact with the form of the bound variable exemplified by *caki*, all instances of reflexivity which derive from bound variable anaphora will have a delay with a cardinality of three. This can be illustrated for Shona as well as Korean:

(396) Imbwa y-oga-yoga ya-ka-zvi-rum-a.
dog.9 CL9-every-REDUP SUBJ.9-PST-REFL-bite-FV
'Every dog bit itself.'

First of all, a tree pair for the predicate needs to be defined:

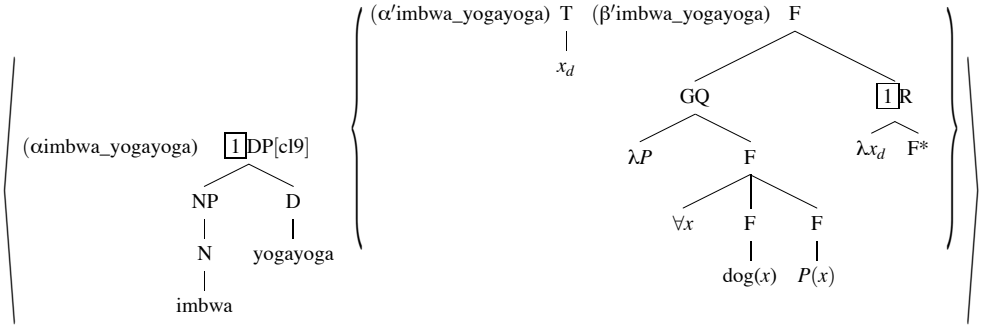
(397)



As with Korean, this syntactic tree abstracts away from some detail. *karuma* 'bite' is chosen as the name of the tree, indicating that tense has been marked, but all agreement is still left open. Here though, *v*P is represented, whereas it was left out of the trees for English and Korean. The choice of whether or not to illustrate the *v*P node reduces to a matter of rhetorical necessary; because these trees already carry a great deal of notation, a simplified syntax is generally used to keep the analysis as clear as possible, but this should in no way be construed as a strong claim that any given head is not present.

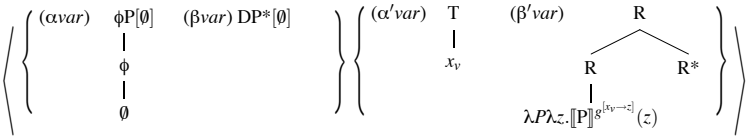
The tree pair for *imbwa yogayoga* 'every dog' will be constructed as a generalised quantifier:

(398)



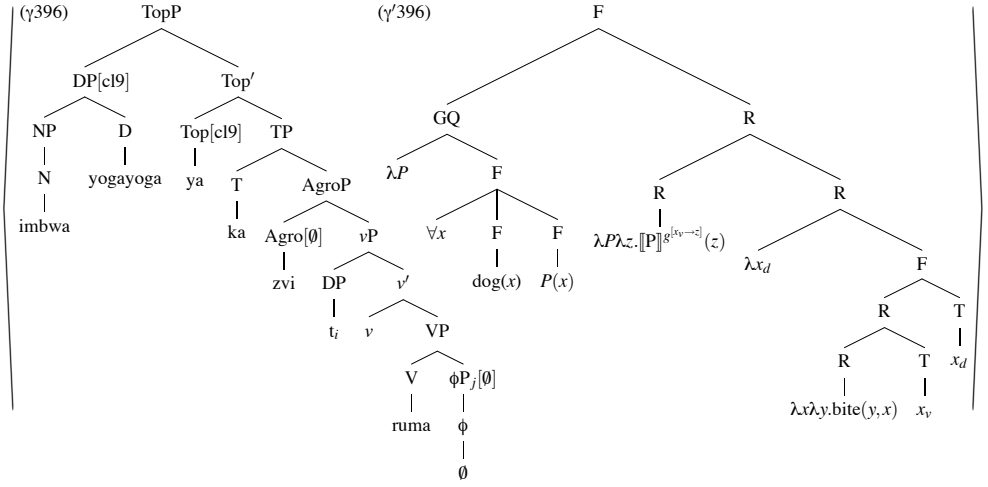
In the syntax, noun class agreement is marked as a ϕ feature specification which values the ϕ features at the node where (αimbwa_yogayoga) substitutes. In the semantics side, (α′imbwa_yogayoga) provides an indexed variable, while (β′imbwa_yogayoga) contains the generalised quantifier. Lastly, there is the covert bound variable:

(399)



In structure, (399) is identical to *caki*, and with a minor difference in variable indexing, the semantics are identical. The syntax however is quite different. As described in Chapter Four, the Shona reflexive is derived from a covert variable which lacks ϕ features. Thus, on the syntax side, this *var* is for all intents and purposes invisible, while there is a full semantic specification. Once composition of all trees has taken place, the result is as shown in (400):

(400)

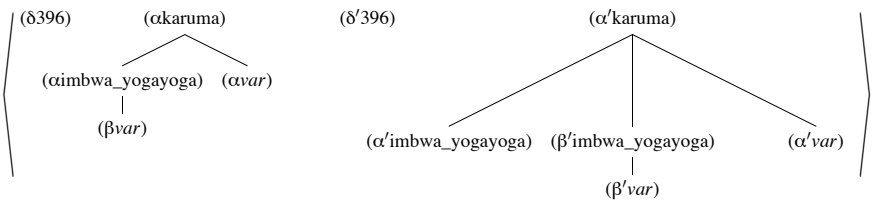


The syntactic tree does not reflect the final head movement to construct the verb stem, though because such movement would be through nodes of the original (αkaruma) tree, there is nothing preventing this movement from occurring. An agreement relation between the AgrO head and the ϕP variable is responsible for the presence of class 8 *zvi* object agreement. In terms of the semantics, this is just a slight extension of the Korean example, with the addition of a universal quantifier:

(401) $\forall x[\ \text{dog}(x)][\text{bite}(x,x)]$

Here, the final semantic form can be expressed in a quantifier-restrictor-nuclear scope structure. Again, as with the case from Korean, the predicate in the quantifier's scope has a reflexive semantics. In terms of the derivation, this looks exactly like the one seen for (388):

(402)



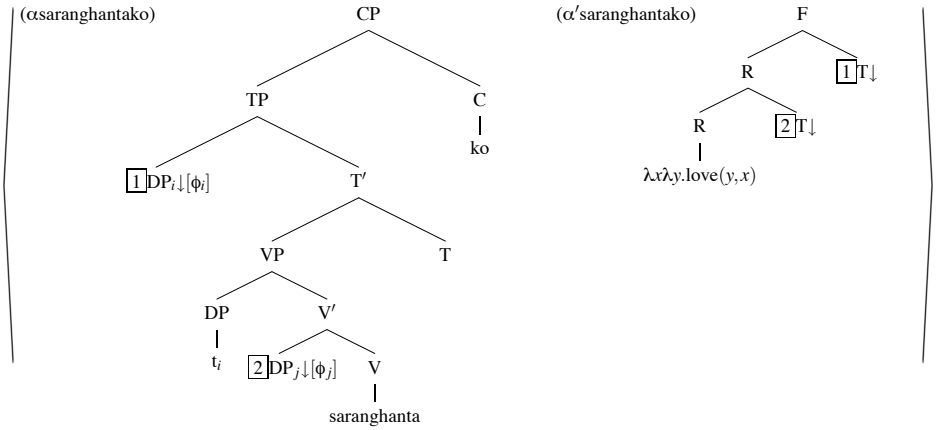
Like the derivation for (388), this too involves two simultaneous delays in the semantics, one a trivial delay for a locally-composed multi-component set, and another delay for the reflexive variable which will have a cardinality of three once again.

Here is where Korean and Shona diverge though: *caki* allows for long-distance binding, while the reflexive variable of Shona does not. To illustrate how long-distance binding for Korean works under this analysis, I work through a simple example:

(403) Swuni$_i$-nun [Minji$_j$-ka caki$_{i/j}$-lul sarang-han-ta-ko] sayngkak-ha-n-ta.
Swuni-TOP Minji-NOM self-ACC love-do-DECL-COMP think-do-PRES-DECL
'Swuni thinks Minji loves self.'

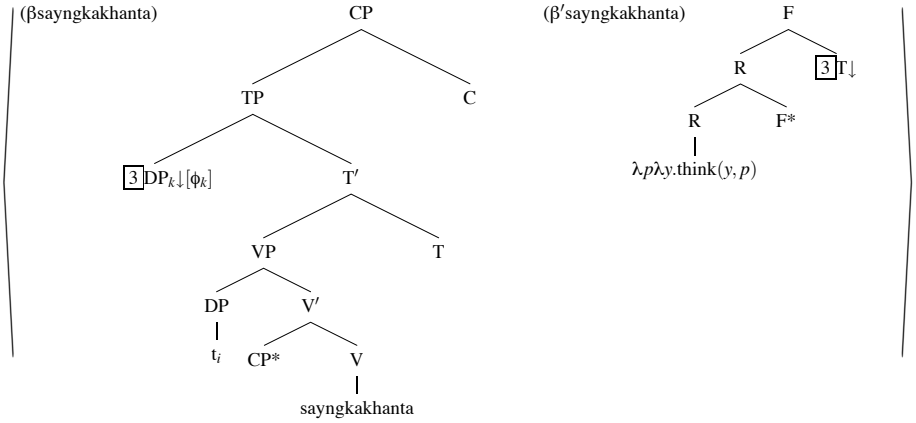
While the sentence in (403) is ambiguous between a local and a long-distance reading of *caki*, of most interest here is the long-distance one. To derive this sentence, the familiar *Chelswu* tree pair from (389) can be adapted. Similar pairs can be defined for the proper names *Swuni* and *Minji*, again using a generalised quantifier analysis, with uniquely indexed variables x_s and x_m for disambiguation purposes. The embedded predicate looks very similar to the earlier tree for *pinanhayessta* 'blamed', but with an extension to accommodate the complementiser:

(404)



The tree pair for the matrix predicate can also be extended to the CP domain; taking a CP complement, this would make *sayngkakhanta* 'thinks' an auxiliary tree, recursive on CP:

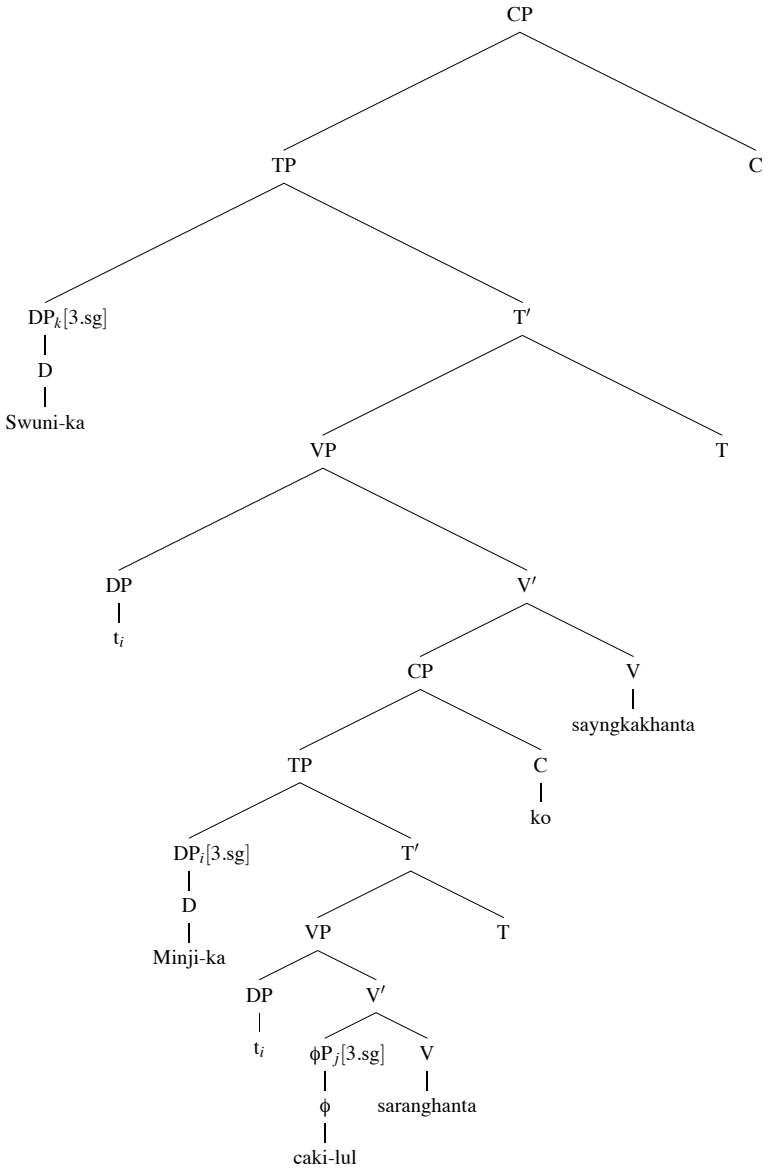
(405)



Finally, no new form of *caki* is required. Both readings can be derived from the trees as they are currently defined. The derived trees for the long-distance reading (γ403) and (γ′403) are given in (406) and (407), respectively.[5]

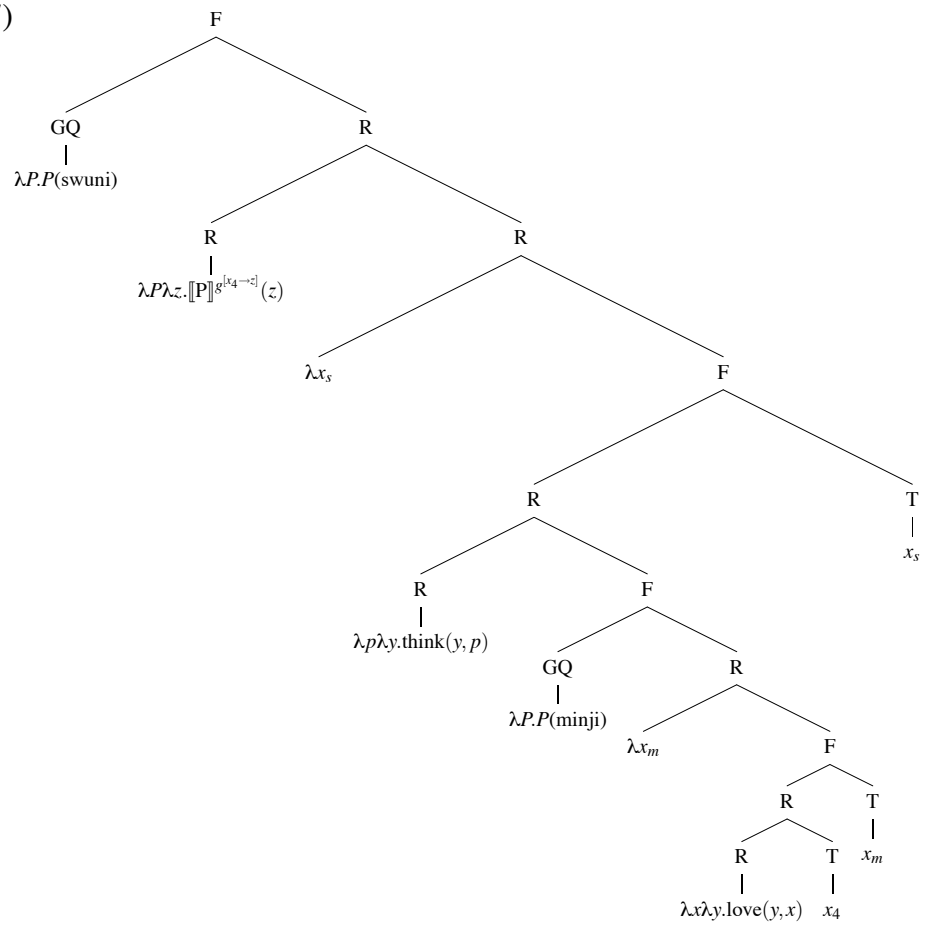
[5]Owing to the size of the trees, the traditional side-by-side presentation will be temporarily abandoned.

(406)



Looking first at (406), there is nothing in the syntax to differentiate which reading, the local or the long-distance, is going to emerge. This work is done in the semantics, particularly in the choice for where to adjoin (β′caki):

(407)

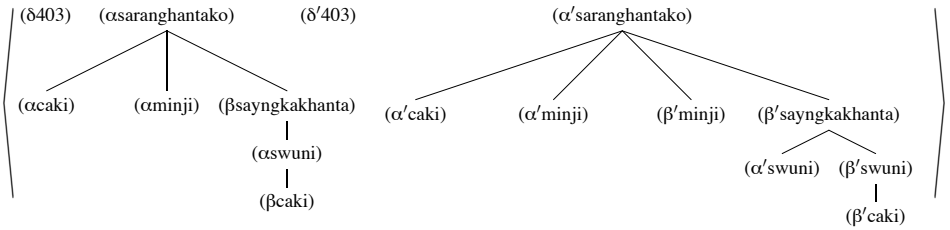


As shown here, (β′caki) has been adjoined into (β′swuni), indicating that *Swuni* should be the antecedent for *caki*. While this is not visible in the syntax, a parallel step takes place with the degenerate (βcaki) adjoining into (αswuni), assuring agreement. The final semantic form is given in (408):

(408) think(swuni, (love(minji, swuni)))

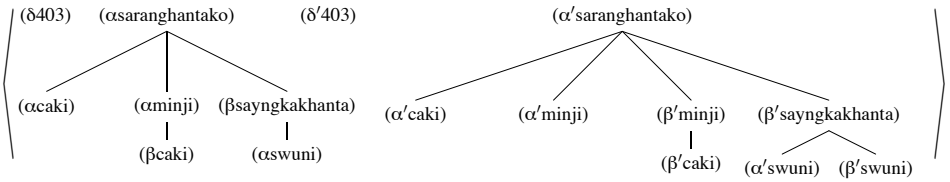
The full derivation trees are shown below:

(409)



Again, because the generalised quantifier is a multi-component set in the semantics, the derivations are not isomorphic, but they are inasmuch as (βcaki) and (β′caki) both compose directly with members of the *Swuni* tree sets. Deriving the local reading, where (βcaki) and (β′caki) composed with *Minji* would have the following derivational trees:

(410)



In examining these derivation trees, it emerges that the difference between local and long-distance binding can be described in terms of the delays for *caki*:

(411) a. Delay in Syntactic Derivation of (403) with local binding:
{(αcaki),(βcaki),(αMinji)}

b. Delay in Semantic Derivation of (403) with local binding:
{(α′caki),(β′caki),(β′minji)}

c. Delay in Syntactic Derivation of (403) with long-distance binding:
{(αcaki),(βcaki), (βsayngkakhanta), (αSwuni)}

d. Delay in Semantic Derivation of (403) with long-distance binding:
{(α′caki),(β′caki), (β′sayngkakhanta), (β′swuni)}

The difference is clear: the cardinality of the delay for the long distance cases is four, while the cardinality for the local cases remains three. Further clausal embedding, or embedding of *caki* within a possessive structure would add to this delay. According to Nesson and Shieber (2009), it is possible within a delayed tree-locality derivation to define constraints based on the length of a delay, here represented as the cardinality of the set of nodes participating in the delay. Here then, strictly local uses of bound variables can be defined as being restricted to a delay of three. Unconstrained uses will have no such constraint.

To round out the discussion on bound variables, it is worth noting that in English, bound variable pronouns may not be used in reflexive contexts:

(412) a. Every girl$_i$ saw herself$_i$.

b. * Every girl$_i$ saw her$_i$

c. Every girl$_i$ knows [that she$_i$ is intelligent].

d. Every girl$_i$ saw her$_i$ father.

English bound variable pronouns can be given exactly the same analysis as *caki* and the reflexive variable *var* proposed for Shona. The difference is that reflexive contexts, where the derivation of *her* would have a delay cardinality of three (412b) is ungrammatical. In such cases, a *self* pronoun must be used, as shown in (412a). Long-distance uses of English bound variable pronouns are fine, as shown in (412c). The possessive example in (412d) will also have a delay cardinality of four, as the *father* elementary tree and the *saw* elementary tree would both intervene between the components of the bound variable, as shown by Storoshenko and Han (2010). So, while *caki* was unconstrained, and the reflexive variable for Shona is strictly local, the English bound variable pronouns are strictly anti-local. All three of these can be captured in terms of the cardinality of a delay, shown in Table 6.1.

Table 6.1: Parameter Settings for Bound Variables

	Delay Cardinality	Example
Local	$= 3$	Shona Reflexive
Anti-Local	> 3	English Bound Variable Pronouns
Unconstrained	≥ 3	Korean *caki*

The choice to formalise this in terms of a cardinality of three is not arbitrary. Recalling the earlier discussion of delays, a delay with a cardinality of three represents the first step in a derivation which becomes non-tree-local; every multi-component set will, by definition, have a delay with a cardinality of at least two, so three is the lowest non-trivial value. For this reason, there is no line in the table for "≤ 3" or "< 3"; any interesting delays will have a cardinality of at least three. Also, three is a key value in that it represents a binding relation which is syntactically local. Thus, this parameter on the types of delay permitted for a bound variable multi-component set can be used as a way to constrain a given variable to be strictly local, strictly anti-local, or unconstrained.

6.4 Representational Challenges

In this section, I will discuss some of the issues encountered in considering an STAG analysis for some of the more complex derivations in Shona, leading into an account for the Plains Cree reflexive.

6.4.1 Bantu Verbal Morphology meets STAG

As discussed in Chapter Four, Shona has a number of verbal suffixes which alter the argument structure of a predicate:

(413) Shingi a-ka-bik-ir-a Mufaro ma-nhanga.
Shingi SUBJ.1-REM.PST-cook-APPL-FV Mufaro CL6.PUMPKIN
'Shingi cooked pumpkins for Mufaro.'

The applicative morpheme *-ir* adds a third argument to the predicate, essentially making this a ditransitive. In the semantics, this could easily be represented as a function on a predicate. First, the semantic form of the verb *bika* 'cook' can be re-cast into a form consistent with Davidsonian event semantics:

(414) $\lambda x \lambda y \lambda e.\text{cook}(x,y)(e)$

Once arguments have been converted in for x and y, a simple existential closure operation can be carried out, deriving a statement to the effect that there is an event of y cooking x. A semantic form of the applicative can be constructed to re-write (414) as in (415):

(415) $\lambda x \lambda y \lambda z \lambda e.\text{cook}(x,y)(e) \wedge \text{benefit}(e,z)$

In (415), a third argument is added, though it is not added directly to the *cook* predicate. Rather, the new argument participates in a benefit relation with the event itself. Expressed in this fashion, the applicative morpheme has some non-trivial semantic content, and in an STAG derivation, one may treat the applicative as a separate elementary tree, composing with a simple transitive predicate. Similarly, it might not be unreasonable to think that tense should be presented as an individuated atom in the analysis.

Returning to the case of the Shona applicative, TAG does not readily allow for this analysis giving the applicative its own elementary tree. In adding an argument to a verb,

the applicative head can be treated as a functional element. As such, according to the CETM, it should be present as a part of the extended projection of the verb, rather than in its own elementary tree. Furthermore, given that in a fully lexicalised TAG each elementary tree has a unique lexical anchor, the only way to treat the applicative as having its own elementary tree would be to view the applicative as lexical, rather than functional.

Recall from Chapter Four that the final word order of the Shona sentence is derived by head movement of the verb through all functional projections up to the Top^0 head. If the applicative head were a part of a different elementary tree that had combined with the predicate, the necessary head movement would cross elementary trees, not permitted in Frank's formalisation of TAG. This exact issue is brought up in Frank's discussion of the following example from Chichewa:

(416) Mtsikana a-nan-gw-ets-a mtsuko.
girl.1 SUBJ.1-PST-fall-CAUS-FV waterpot
'The girl made the waterpot fall. (Frank 2002, ex 2.30b)

Here, Frank considers an analysis in which the causative morpheme is a separate predicate from the *fall* predicate, and (416) a result of a verb-incorporation operation. He ultimately rejects this analysis (or at least deems it to be incompatible with TAG) because it would require a movement between elementary trees. Rather, he concludes that the causative morpheme be treated as a part of the lexical form of the verb.

An open question at this point would be to ask whether there is any synchronisation between the construction of the syntactic and semantic elementary trees for a given lexical item. As currently formulated, STAG is only concerned with synchronous TAG derivations: the elementary tree combination operations. This kind of semantic detail might show up in a finer examination of the elementary tree-construction process which creates the structures which are ultimately manipulated by the substitution and adjoining operations, but at present there is no formal account of how this is carried out in an STAG context. Clearly, this an area for future research, but at present, the formalism does not allow for this kind of fine analysis of a language such as Shona.

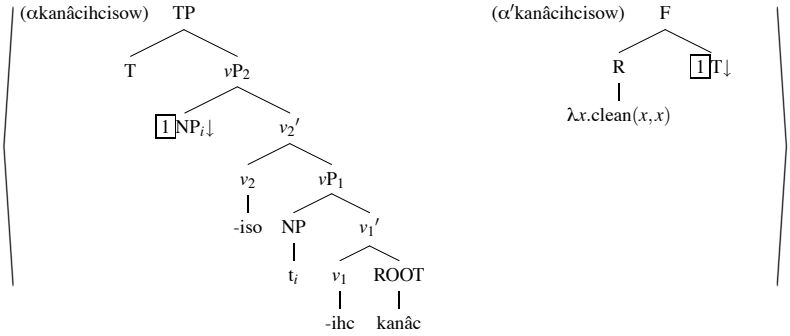
6.4.2 Plains Cree Reflexives

The prior discussion also has a bearing on the STAG analysis for the Plains Cree reflexive. Following Hirose's analysis, the reflexivising morpheme *-iso* is a functional element, and so should be present as a part of the verb root's extended projection in STAG. Thus, an example such as (417) will have a very simple derivation:

(417) kanâc-ihc-iso-w awâsis.
clean-TRANS-REFL-3.SG child
'The child cleaned him/herself.'

Extending Hirose's finer *v*P structure would yield the following tree pair:

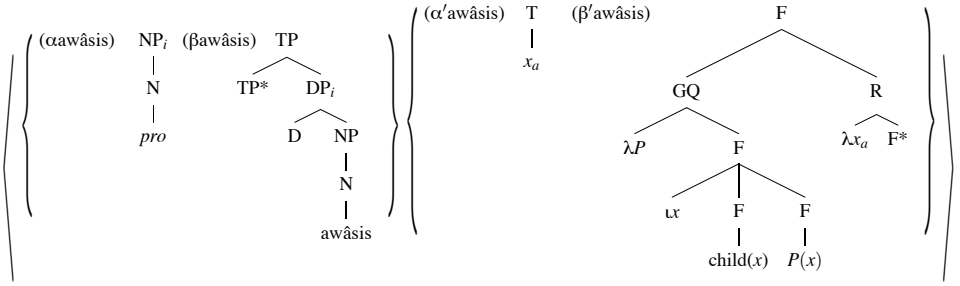
(418)



The *v*P-internal movement of the argument *pro* which Hirose argues to be the source of the reflexive reading is not problematic, as the TAG θ-Criterion does not appear to directly speak to the issue of using movement to have more than one θ-role assigned to one entity. To avoid a proliferation of traces, the final head movement which derives the morpheme order is not shown; the verb-final agreement morpheme is not given its own syntactic position as Hirose argues that this arises out of a checking relation between the verb stem and its argument, not from any specific head. As with the prior discussion of Shona, the semantic tree (α′kanâcihcisow) does not show any derivation of a reflexive reading; because the reflexive morpheme is considered to be a part of the functional projection of the verb, the reflexive form of the verb is presented directly in the semantic tree. An obvious question at this point is to wonder where exactly the overt subject is going to go.

According to Blain (1995), the overt arguments in Plains Cree can be treated as adjuncts to TP. As such, the best treatment would be to see the argument as a multi-component set consisting of a *pro* which can substitute into the argument position, and an auxiliary tree which carries the overt DP:

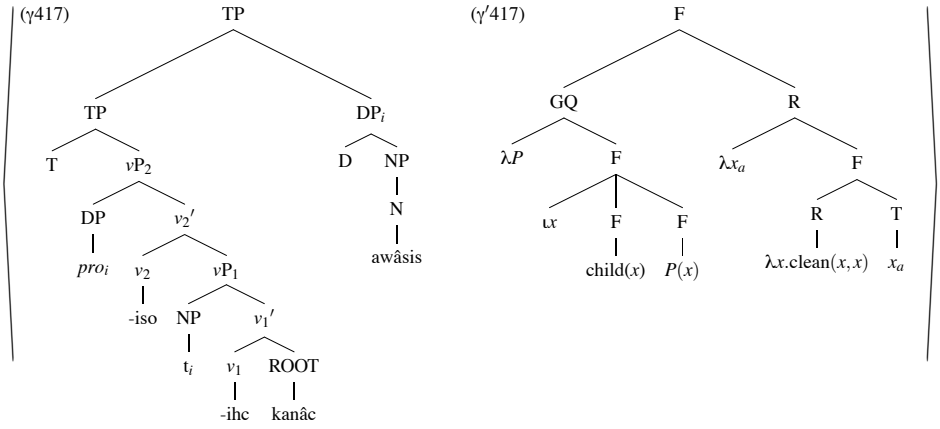
(419)



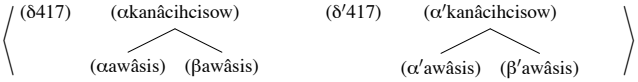
In the semantics, the relationship between the *pro* and the overt DP can be represented as a generalised quantifier, exactly as in the Korean and Shona cases. For cases where the overt DP is not used, the semantic form (α′awâsis) would be free, though open to contextual interpretation. Not only does this derive exactly the syntactic tree proposed by Blain, but it stays true to the generalisation in TAG that adjuncts are introduced by adjoining. Also, this approach sidesteps any potential conflicts with the TAG θ-Criterion which may arise should the sentence-final position be considered a substitution site. An alternative approach would be to consider these two be two distinct elementary trees, a *pro* tree and an adjunct nominal tree. While technically implementable, such an approach would require a more complicated semantics, equating the *pro* and the overt argument via the predicate. If these adjuncts can only occur in the presence of a co-indexed *pro*, their connection as a multi-component set can be justified. However the argument is implemented though, there is no change in the formation of the reflexive, as the *kanâcihcisow* 'clean' tree will still support only one argument.

The trees compose as in (420), according to the derivation structures in (421):

(420)



(421)



Semantic composition on ($\gamma'417$) yields the expected reflexive form:

(422) $\iota x[\text{child}(x)][\text{clean}(x,x)]$

In terms of the STAG derivation, Plains Cree is by far and away the simplest, owing largely to the fact that the semantic consequences of the movements within the elementary tree itself are hidden at the level of the STAG derivation. To get at the finer details, more work would be needed to explore the stages of elementary tree derivation in STAG, a project well beyond the scope of the present thesis.

6.5 Final Summary

In this chapter, it has been shown that the STAG formalism is not only able to adequately model the three different types of reflexivity, but that the mechanisms of STAG can be applied to four typologically distinct languages. In English, by taking advantage of TAG's unique account of raising, a structure which proved to be a challenge for a Reinhart and Reuland style semantics-based approach to reflexives was easily brought in line with the established analysis of co-argument reflexives. Furthermore, both parts of Chomsky's Condition A can be seen as consequences of the proposed STAG analysis. In order to ensure proper semantic composition, the reflexive pronoun must appear in a syntactic position c-commanded by its antecedent. This is not formalised as a constraint on the syntax, merely a consequence of the links between the semantic and syntactic derivations. Further, the locality restriction encoded in Condition A is reflected in the fact that the reflexive multi-component set composes tree locally, with a maximum delay cardinality of two.

The introduction of the concept of delay cardinality allows for the definition of a simple three-way parametric variation to define three types of bound variable anaphora: local, anti-local, and unconstrained. With this, the differences between Korean *caki*, the Shona reflexive variable, and English bound variable pronouns can all be described as different settings of a very simple parameter constraining the derivation of sentences containing those bound variables.

While STAG was able to capture the reflexive forms of Shona and Plains Cree well enough, the examination of those languages shed light on an area of STAG which needs further development: an exploration of the construction of the elementary trees themselves.

Because certain of the functional elements in the syntax have very clear semantic effects, it would be desirable to see STAG taken to a deeper level of representation which showed in more detail how these functional elements operate. So while it is not impossible to give an STAG treatment of languages such as Shona or Plains Cree, the analysis of those languages would benefit from further work on STAG itself.

To sum up, the three different reflexive forms, NP, ϕP, and DP come out of the STAG derivations with the following final semantic forms:

(423) a. $\iota x[\text{child}(x)][\text{clean}(x,x)]$

b. $\text{blame}(\text{chelswu}, \text{chelswu})$

c. $\text{loves}(\text{john}, \text{himself}) \wedge \text{himself} = \text{john}$

Expressed in a generalised quantifier format, the example from the Plains Cree NP reflexive still contains at its core the reflexive meaning illustrated in (1), repeated below as (424):

(424) $[\![\text{hit}]\!] = \lambda x \lambda y.\ \text{hit}(y,x) \rightarrow \text{hit}(a,a)$

While there is some variable binding at play in the semantic form of (423a), the reflexive part of the nuclear scope was derived independently. Using a GQ for a proper name, the Korean example is even more clearly reflexive, though crucially this form is derived through variable binding. As a last observation, the form of the reflexive used for English is somewhat jarring, in that it does not match so cleanly with the final form of the other two. Given that the *self* pronoun in English leads a mixed existence between a function on predicates, a function on nominals, and, in its exempt uses, a purely referential element, the present analysis is motivated, but it would be no bad thing to bring English DP reflexives in line with the other two.

This can be accomplished simply enough by breaking (423c) into its component conjuncts and constructing a small proof:

(425)

$\text{loves}(\text{john}, \text{himself})$

$\text{himself} = \text{john}$

$\therefore\ \text{loves}(\text{john}, \text{john})$

With this last logical step, all three forms of the reflexive now reduce to the same semantic form defined at the outset of Chapter One.

Chapter 7

Out of the Woods
Conclusions and Future Directions

There is still so much to do...and still so much to learn.
-Jean-Luc Picard. *Star Trek: The Next Generation.*

In this closing chapter, I provide a brief re-cap of the main points of the thesis, before moving on to a discussion of potential future research avenues.

7.1 Looking Back

This thesis began with a very simple question, posed in relation to the example repeated below:

(426) $[\![\text{hit}]\!] = \lambda x \lambda y.\ \text{hit}\ (y, x) \rightarrow \text{hit}\ (a, a)$

The question was one of determining how it is that a transitive predicate can lead to a reflexive interpretation — what's going on at that arrow? Using data from four different languages, it has been shown that there are three broad strategies which can be employed, all of which lead to the same reflexive semantic form.

For English, it was argued that the *self* pronouns can be represented in the syntax as DPs, having the ability to stand on their own in referential contexts. This structure not only

reflects the historical development of these forms, from roots as a referential expression, but the fact that referential uses for *self* pronouns persist, both in the classic cases of exempt anaphora and in a sort of politeness form uncovered in the corpus study of English *self* pronoun use. As reflexives, this referential meaning is combined with a predicate by way of adding a statement of assigned co-reference. The emphatic heritage of these forms is also reflected in their use as focus-sensitive forms. While I have argued against an analysis which derives all emphatic uses of the *self* pronouns from a single underlying form, I have borrowed the fundamentals of that analysis, supporting the idea that the *self* pronouns have a significantly different semantics in their reflexive and emphatic uses.

Korean has a multitude of reflexive forms, with varying claims in the literature as to their status as restricted to local binding, anti-local binding, and whether or not they can be unified with a treatment of long-distance anaphors which predicts subject-orientation through covert head movement to the matrix clause. Of the Korean reflexive forms, the one with probably the most inconsistent data is *caki*. While there is a back-and-forth in the literature on whether *caki* should be considered a pronoun or an anaphor, here I have argued for a third line, also hinted at in the literature, which treats *caki* as a bound variable. This is verified through a corpus analysis which shows that the overwhelming majority of the uses of *caki* can be made compatible with a bound variable analysis. Further psycholinguistic experimentation using eye-tracking not only verifies that *caki* is far less contextually-sensitive than referential pronouns, but provides evidence that the subject orientation observed for *caki* appears to be a default for Korean anaphora in general, as the same initial bias for subject antecedents was observed for the referential pronouns as well.

In Shona, reflexivity is overtly expressed as a form of object agreement which is more broadly used for indefinites, and for objects which lack ϕ-features. This, I argue, is not a co-incidence, as the reflexive can be bound by any subject antecedent, regardless of ϕ-features or noun class. Under my proposed analysis, Shona reflexive forms arise out of the binding of a ϕ-featureless bound variable which is restricted to local binding only. This variable binding serves a disambiguating function, as elicitation data suggests that Shona allows for unmarked, but ambiguous cases of co-argument reflexivity as well.

Plains Cree manifests reflexivity through an overt verbal suffix which licenses a reorganisation of the *v*P domain to derive a form which is semantically reflexive, but syntactically intransitive. An examination of a small selection of Plains Cree texts shows an

unexpectedly high number of reflexive sentences, proportionally almost twice as many as English. Cross-checking these reflexive cases against their English translations reveals that Plains Cree makes use of reflexives in many cases where an English paraphrase would make use of a bound possessive pronoun. This hints at a connection back with the Plains Cree use of noun incorporation as another detransitivising strategy, suggesting that the language may make use of an overall strategy of incorporating internal arguments directly into the predicate where possible.

All four languages received an analysis in Synchronous Tree Adjoining Grammar. At the outset, STAG appeared to be an ideal choice for modelling reflexivity, as it presents parallel derivations for syntax and semantics. In applying STAG to the analysis of English *self* pronouns, it was found that STAG actually allowed for a more elegant account of cases where there was interaction between the reflexive and a raising predicate, and the two components of Condition A, a structural condition in the form of c-command and a domain restriction in the form of the binding domain, are both direct consequences of the STAG analysis, rather than external stipulations. For Korean and Shona, an STAG account of bound variable anaphora is introduced, making use of delayed tree-locality in the STAG derivation to define a parameter which dictates whether a given variable will demonstrate local, anti-local, or unconstrained binding. In looking at Shona and Plains Cree though, new problems arise. The STAG analysis is somewhat limited by the formalism itself, as there is currently no established means within STAG of examining the construction of elementary trees in the syntax and semantics. As a result, there is no easy way to tease apart the semantic contributions of various functional heads, which renders the STAG analysis of Plains Cree reflexivity in particular very simple indeed.

7.2 Looking Ahead

This apparent weakness of STAG is one of the clearest avenues for future work arising from this thesis. While this thesis has met its goal of minimally demonstrating that STAG can be used for a diverse collection of languages, the application of STAG to Shona and Plains Cree has demonstrated that the STAG formalism needs to be extended inward. While the synchronisation of the TAG operations is well-implemented, and extensions such as delayed tree-locality have proved to be fruitful, more attention needs to be paid to the con-

struction of the elementary tree pairs themselves. An interesting avenue to explore would be to determine whether or not the construction of the paired elementary trees themselves is also synchronous.

Recalling the discussion of the English emphatics, a full STAG account of those forms was not given due to the fact there has been no formalisation of an application of alternative semantics and focus interpretation within STAG. Information structure sensitivity in general should be encoded, as Korean, Shona, and Plains Cree have all demonstrated that topical or discourse-old status is a key component of the grammar in those languages. There is as yet no formal means of encoding this in STAG. If STAG is to be seriously pursued as a model for the syntax-semantics interface, then this inclusion of information structure also appears to be necessary. In a sense then, STAG as it stands now is at the middle of where it needs to be; inward growth in terms of the examination of the elementary trees themselves and outward growth to discourse sensitivity are both called for.

For each language, there is also more work that can be done. Looking first at English, the proposed forms of the emphatic uses of the *self* pronouns are tentative, and could do with more study. Whether this means a more detailed examination of the existing corpus, or possibly a totally new corpus study, more detailed accounts of the inclusive and exclusive forms would be desirable, as well as a search for evidence that the two forms can indeed co-occur naturally, outside of constructed examples. This can be further tested in more comprehensive experimentation where sentences with the emphatics are presented in contexts that are explicitly compatible with either an inclusive or an exclusive use. Rather than an acceptability rating task, this can be accomplished using a truth value judgement task, to more precisely determine which readings are possible for a given sentence. Furthermore, the so-called "creeping" reflexives, I believe are worthy of more detailed study. A diachronic approach might be revealing here, using corpora from different periods to determine whether there has been any extension of the use of *self* pronouns in non-reflexive contexts. The data suggest that there may be a movement toward using the *self* pronouns as a polite or higher register form of the referential pronouns; ongoing research may track such a change in progress if it exists.

In Korean, the eye-tracking experiment revealed that there is a default subject bias, but yielded virtually no instances of non-subject binding of *caki*. In future research, the existing eye-tracking study can be remounted with a slight alteration of the background

context for the target sentences, actually playing the dialogue between the two characters making it explicit which character *caki* should refer to. Instead of a forced-choice question asking participants to determine the antecedent of *caki*, this can also be implemented as a truth value judgement task, in which participants judge a controlled reading of the target sentence. Running this test will give a more conclusive answer to the question of whether or not the generally-observed subject orientation for *caki* is hard-coded into the lexical item itself, or is just a by-product of the proposed default subject orientation for the language, able to be overridden when context demands.

Turning to Shona, one of the first tasks would seem to be a more thorough inventory of what is covered under class 8 agreement, as the data suggest that class 8 covers more than what is hinted at in the published grammars. It would also be worthwhile to examine more Bantu languages to see what generalisations can be made in terms of connecting class 8 agreement and reflexivity. Also, of the four languages covered in this thesis, Shona is the one whose reciprocal form is most markedly different from its reflexive. While this has not been discussed in any great detail, a comparison of the reflexive and reciprocal in Shona and Bantu in general would be interesting, as my preliminary data points to a conclusion that the Shona reciprocal is similar to the Plains Cree reflexive, in that it is a detransitiviser. Also, while I have made every effort to remain agnostic on this point within the confines of this thesis, there is still a question of the syntactic status of object nominals which occur in conjunction with object markers. More tests need to be conducted to determine whether or not object markers in particular should be treated as agreeing directly with the overt object, or agreeing with a *pro* in the argument position, while the overt object is a displaced modifier, along the lines of the pronominal argument hypothesis.

Finally, the corpus work on Plains Cree uncovered an unexpected density of reflexives. Through the collection of a larger corpus, combined with new fieldwork, it would be interesting to explore the idea that these reflexives are being used at the exclusion of possession. Also, along with Shona, Plains Cree appears to be an ideal candidate to serve as a test-case for the extension of STAG, as this thesis has barely scratched the surface of the complexities of the Algonquian languages. Getting back to the subject of topicality, the proximate/obviate distinction is a good example of something that needs a more formal account. All of this, however, lies in the realm of future work.

Bibliography

Ahenakew, Alice. 2000. *Âh-âyîtaw isi ê-kî-kiskêyihtahkik maskihkiy = They knew both sides of medicine: Cree tales of curing and cursing told by Alice Ahenakew*. Winnipeg MB: University of Manitoba Press.

Ahenakew, Freda, ed. 1989. *Kiskinahamawâkan-âcimowinisa : Student stories*. Winnipeg: Algonquian and Iroquoian Linguistics.

Baker, C. L. 1995. Contrast, discourse prominence, and intensification, with special reference to locally free reflexives in British English. *Language* 71:63–101.

Baker, Mark. 1988. *Incorporation: A theory of grammatical function changing*. Chicago: University of Chicago Press.

Barss, Andrew, and Howard Lasnik. 1986. A note on anaphora and double objects. *Linguistic Inquiry* 17:347–354.

Barwise, John, and Robin Cooper. 1981. Generalized quantifiers and natural language. *Linguistics and Philosophy* 4:159–219.

Bellusci, David Christian. 1991. Serialization patterns in Shona verbal morphology. Doctoral Dissertation, University of Calgary.

Bickerton, Derek. 1987. *He Himself*: Anaphor, pronoun, or...? *Linguistic Inquiry* 18:345–348.

Blain, Eleanor M. 1995. Emphatic *wiya* in Plains Cree. In *Papers of the Algonquian Conference*, volume 26, 22–34.

Bliss, Heather. 2009. Comparing APPL's and oranges: The syntax of Shona applicatives. In *Proceedings of the 39th Annual Conference on African Linguistics*, ed. Akinloye Ojo and Lioba Moshi, 100–109.

Bliss, Heather, and Dennis Ryan Storoshenko. 2008. Passivization and A' movement in Shona. In *Proceedings of the 2008 Annual Conference of the Canadian Linguistics Association*.

Bliss, Heather, and Dennis Ryan Storoshenko. in prep. All the world's a stage: Locative stage topics in Shona. In *Proceedings of the 2010 Annual Conference of the Canadian Linguistics Association*.

Brauner, Siegmund. 1995. *A grammatical sketch of Shona*. Köln, Köppe.

Bresnan, Joan, and Jonni M. Kanerva. 1989. Locative inversion in Chicheŵa: A case study of factorization in grammar. *Linguistic Inquiry* 20:1–50.

Bresnan, Joan, and Sam A. Mchombo. 1987. Topic, pronoun, and agreement in Chicheŵa. *Language* 63:741–782.

Bresnan, Joan, and Lioba Moshi. 1990. Object asymmetries in comparative Bantu syntax. *Linguistic Inquiry* 21:147–185.

Büring, Daniel. 2005. *Binding theory*. Cambridge, UK: Cambridge University Press.

Canac-Marquis, Rejean. 2005. Phases and binding of reflexives and pronouns in english. In *Proceedings of the HPSG05 Conference*, ed. Stefan Muller, 482–502.

Champollion, Lucas. 2008. Binding theory in LTAG. In *Proceedings of the 9^{th} International Workshop on Tree Adjoining Grammars and Related Formalisms*, ed. Claire Gardent and Anoop Sarkar, 1–8.

Chang, Sun. 1977. Korean reflexive pronoun *caki* and its referent NP's point of view. *Language Research* 13:35–48.

Cheshire, Jenny, Viv Edwards, and Pamela Whittle. 1993. Non-standard English and dialect levelling. In *Real English: The grammar of English dialects in the British Isles*, 53–96. London: Longman.

Chiang, David, and Tatjana Scheffler. 2008. Flexible composition and delayed tree-locality. In *Proceedings of the 9th International Workshop on Tree Adjoining Grammars and Related Formalisms*, ed. Claire Gardent and Anoop Sarkar, 17–24.

Cho, Dong-In. 1996. Anaphor or pronominal. *Language Research* 32:621–636.

Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht: Foris.

Chomsky, Noam. 1986. *Barriers*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1995. *The minimalist program*. Dordrecht: Foris.

Chomsky, Noam. 2005. On phases. Manuscript.

Cole, Peter, Gabriella Hermon, and Li-May Sung. 1990. Principles and parameters of long-distance reflexives. *Linguistic Inquiry* 21:1–22.

Creissels, Denis. 2001. A typology of subject marker and object marker systems in African languages. In *Proceedings of the International Symposium: Typology of African Languages*.

Creissels, Denis. 2002. Valence verbale et voix en Tswana. *Bulletin de la Société de Linguistique de Paris* 97:371–426.

D'Alessandro, Roberta. 2007. *Impersonal si constructions*. New York: Mouton de Gruyter.

Dalyedwa, Ntombizodwa Cynthia. 2002. Valency reducing processes in Xhosa. Doctoral Dissertation, University of Essex.

Déchaine, Rose-Marie, and Martina Wiltschko. 2002a. Decomposing pronouns. *Linguistic Inquiry* 33:409–442.

Déchaine, Rose-Marie, and Martina Wiltschko. 2002b. Deriving reflexives. In *Proceedings of the 21st West Coast Conference on Formal Linguistics*, 71–84.

Diercks, Michael. in press a. The morphosyntax of Lubukusu locative inversion and the parametrization of agree. *Lingua* .

Diercks, Michael. in press b. Parametrizing case: Evidence from Bantu. *Syntax* .

Fortune, George. 1955. *An analytical grammar of Shona*. London: Longmans, Green and Co.

Fortune, George, ed. 1974. *Ngano*. Department of African Languages, University of Rhodesia.

Frank, Robert. 2002. *Phrase structure composition and syntactic dependencies*. Cambridge, MA: MIT Press.

Frank, Robert. 2008. Reflexives and TAG semantics. In *Proceedings of the 9th International Workshop on Tree Adjoining Grammars and Related Formalisms*, ed. Claire Gardent and Anoop Sarkar, 97–104.

Gast, Volker. 2006. *The grammar of identity: Intensifiers and reflexives in Germanic languages*. New York: Routledge.

Gast, Volker, and Peter Siemund. 2006. Rethinking the relationship between SELF-intnesifiers and reflexives. *Linguistics* 44:348–381.

Gill, Kook-Hee. 1999. The long-distance anaphora conspiracy: The case of Korean. *University of Pennsylvania Working Papers in Linguistics* 6:171–183.

Han, Chung-hye. 2007. Pied-piping in relative clauses: Syntax and compositional semantics using Synchronous Tree Adjoining Grammar. *Research on Language and Computation* 5:457–479.

Han, Chung-hye, David Potter, and Dennis Ryan Storoshenko. 2008. Compositional semantics of coordination using Synchronous Tree Adjoining Grammar. In *Proceedings of the 9th International Workshop on Tree Adjoining Grammars and Related Formalisms*, ed. Claire Gardent and Anoop Sarkar, 33–41.

Han, Chung-hye, and Dennis Ryan Storoshenko. 2009. A bound-variable analysis of the Korean anaphor *caki*, evidence from corpus. In *Current issues in unity and diversity of languages: Collection of the papers selected from the 18th International Congress of Linguistics*. The Linguistic Society of Korea.

Han, Chung-hye, Dennis Ryan Storoshenko, and Calen Walshe. in press. An experimental study of the grammatical status of *caki* in Korean. In *Proceedings of the 19th Japanese/Korean Linguistics Conference*. University of Hawaii at Manoa.

Harford-Perez, Carolyn. 1985. Aspects of complementation in three Bantu languages. Doctoral Dissertation, University of Wisconsin-Madison.

Henderson, Brent. 2006. Multiple agreement and inversion in Bantu. *Syntax* 9:275–289.

Hirose, Tomio. 2003. *Origins of predicates*. Outstanding Dissertations in Linguistics. New York: Routledge.

Hirst, Graeme. 1981. *Anaphora in natural language understanding: A survey*. Berlin: Springer-Verlag.

Huang, Yan. 2000. *Anaphora*. Oxford, UK: Oxford University Press.

Jackendoff, Ray. 1972. *Semantic interpretation in generative grammar*. Cambridge, MA: MIT Press.

Jaeggli, Osvaldo. 1986. Passive. *Linguistic Inquiry* 17:587–622.

Jayaseelan, Karattuparambil A. 1997. Anaphors as pronouns. *Studia Linguistica* 51:186–234.

Jelinek, Eloise. 1984. Empty categories, case, and configurationality. *Natural Language and Linguistic Theory* 2:39–76.

Joshi, Aravind K., L. Levy, and M. Takahashi. 1975. Tree adjunct grammars. *Journal of Computing Systems Science* 10:132–163.

Kâ-Nîpitêhtêw, Jim. 1998. *Ana kâ-pimwêwêhahk okakêskihkêmowina: The counselling speeches of Jim Kâ-nîpitêhtêw*. Winnipeg MB: University of Manitoba Press.

Kahari, G. P. 1981. The history of the Shona protest song: A preliminary study. *Zambezia* 9:79–101.

Kaiser, Elsi, Jeffrey T. Runner, Rachel S. Sussman, and Michael K. Tanenhaus. 2009. Structural and semantic constraints on the resolution of pronouns and reflexives. *Cognition* 112.

Kallmeyer, Laura, and Aravind K. Joshi. 2003. Factoring predicate argument and scope semantics: Underspecified semantics with LTAG. *Research on Language and Computation* 1:3–58.

Kallmeyer, Laura, and Maribel Romero. 2007. Reflexives and reciprocals in LTAG. In *Proceedings of the Seventh International Workshop on Computational Semantics*, ed. Jeroen Geertzen, Elias Thijsse, Harry Bunt, and Amanda Schiffrin, 271–282.

Kang, Beom-Mo. 1988. Unbounded reflexives. *Linguistics and Philosophy* 11:415–456.

Kang, Beom-Mo. 2001. The grammar and use of Korean reflexives. *International Journal of Corpus Linguistics* 6:134–150.

Kang, Young-Se. 1986. Korean anaphora: *Caki* as a resumptive pronoun. *Language Research* 22:215–228.

Katada, Fusa. 1991. The LF representation of anaphors. *Linguistic Inquiry* 22:287–313.

Keller, Frank, and Ash Asudeh. 2001. Constraints on linguistic coreference: Structural vs. pragmatic factors. In *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*.

Kim, Ji-Hye, and James H. Yoon. 2009. Long-distance bound local anaphors in Korean - An empirical study of the Korean anaphor *caki-casin*. *Lingua* 119:733–755.

Kim, Soo-Yeon. 2000. Acceptability and preference in the interpretation of anaphors. *Linguistics* 38:315–353.

Kim, Wha-Chun Mary. 1976. The theory of anaphora in Korean syntax. Doctoral Dissertation, Massachusetts Institute of Technology.

Kioko, Angelina Nudku. 2005. *Theoretical issues in the grammar of Kikamba*. Munich: Lincom Europa.

König, Ekkehard, and Volker Gast. 2006. Focused assertion of identity: A typology of intensifiers. *Linguistic Typology* 10:223–276.

König, Ekkehard, and Peter Siemund. 2000a. The development of complex reflexives and intensifiers in English. *Diachronica* 17:39–84.

König, Ekkehard, and Peter Siemund. 2000b. Locally free *self*-forms, logophoricity, and intensification in English. *English Language and Linguistics* 4:183–204.

Kroch, Anthony, and Aravind Joshi. 1985. The linguistic relevance of Tree Adjoining Grammar. Technical Report MS-CS-85-16, Department of Computer and Information Sciences, University of Pennsylvania.

Kunene, Euphrasia. 1975. Zulu pronouns and the structure of discourse. *Studies in African Linguistics* 6:171–183.

Kuno, Susumu. 1987. *Functional syntax: Anaphora, discourse and empathy*. Chicago: University of Chicago Press.

Larson, Richard. 1988. On the double object construction. *Linguistic Inquiry* 19:335–391.

Lee, Chungmin. 1973. Abstract syntax and Korean with reference to English. Doctoral Dissertation, University of Indiana.

Lee, Chungmin. 1988. Issues in Korean anaphora. In *Papers from the 6^{th} International Conference on Korean Linguistics*, 339–358.

Lee, Gunsoo. 2000. *Kucasin* is a long-distance anaphor. In *Proceedings of the 12^{th} Western Conference on Linguistics*, 305–315.

Lee, Gunsoo. 2004. On the dual nature of Korean bimorphemic anaphors. *Harvard Studies in Korean Linguistics* 10:558–567.

Lee, Hong-Bae. 1976. Notes on pronouns, reflexives, and pronominalization. *Language Research* 12:253–263.

Lee, Ik-Hwan. 1978. Pronominal anaphora in Korean. *Language Research* 14:63–99.

Lees, R. B., and E. S. Klima. 1963. Rules for English pronominalization. *Language* 39:17–28.

Li, Yafei, and Chioko Takahashi. 1995. On the logophoric AND syntactic nature of reflexivization. *Cornell Working Papers in Linguistics* 13:49–84.

Lidz, Jeffrey. 1996. Dimensions of reflexivity. Doctoral Dissertation, University of Delaware.

Madigan, Sean, and Masahiro Yamada. 2006. Asymmetry in anaphoric dependencies: A cross-linguistic study of inclusive reference. In *University of Pennsylvania Working Papers in Linguistics 13-1: Proceedings of Penn Linguistics Colloquium 30*, 183–196.

Manzini, Maria Rita. 1986. On Italian *si*. In *Syntax and semantics v. 19: The syntax of pronominal clitics*, 241–262. New York: Academic Press, Inc.

Marcus, Mithcell P, Beatrice Santorini, Mary Ann Marcinkiewicz, and Ann Taylor. 1999. *Treebank-3*. Philadelphia: Linguistics Data Consortium.

Moon, Seung-Chul. 1995. An optimality theory approach to long distance anaphors. Doctoral Dissertation, University of Washington.

Moravcsik, Edith A. 1972. Some crosslinguistic generalizations about intensifier constructions. In *Papers from the 8^{th} Regional Meeting of the Chicago Linguistic Society*, 271–277.

Nesson, Rebecca, and Stuart M. Shieber. 2009. Efficiently parsable extensions to Tree-Local Multicomponent TAG. In *Proceedings of NAACL 2009*, 92–100.

O'Grady, William. 1984. The syntax of Korean anaphora. *Language Research* 20:121–138.

O'Grady, William. 1987. The interpretation of Korean anaphora: The role and representation of grammatical relations. *Language* 63:251–277.

Park, Sung-Hyuk. 1986. Parametrizing the theory of binding: The implication of *caki* in Korean. *Language Research* 22:229–253.

Reinhart, Tanya. 1997. Syntactic effects of lexical operations: Reflexives and unaccusatives. *UiL-OTS Working Papers of Theoretical Linguistics* 97.

Reinhart, Tanya, and Eric Reuland. 1993. Reflexivity. *Linguistic Inquiry* 24:657–720.

Rooth, Mats. 1992. A theory of focus interpretation. *Natural Language Semantics* 1:75–116.

Rooth, Mats. 1996. Focus. In *The handbook of contemporary semantic theory*, ed. Shalom Lappin, 271–297. Oxford UK: Blackwell Publishers.

Runner, Jeffrey T. 2007. Freeing possessed NPs from binding theory. *University of Rochester Working Papers in the Language Sciences* 3:57–90.

Russi, Cinzia. 2006. Morphosyntactic functions of Italian reflexive *si*: A grammaticalization analysis. In *Historical romance linguistics: Retrospective and perspectives*, ed. Randall S. Gess and Deborah Arteaga, 357–374. Amsterdam, NE: John Benjamins Publishing Company.

Sæbø, Kjell Johan. 2009. Self intensification and focus interpretation. *Oslo Studies in Language* 1:1–21.

Sells, Peter. 1987. Aspects of logophoricity. *Linguistic Inquiry* 18:445–480.

Shieber, Stuart M. 1994. Restricting the weak generative capacity of Synchronous Tree Adjoining Grammars. *Computational Intelligence* 10:371–385.

Shieber, Stuart M., and Yves Schabes. 1990. Synchronous tree adjoining grammars. In *Papers Presented to the 13th International Conference on Computational Linguistics*, volume 3, 253–258.

Simango, Silvester Ron. 2006. Verb agreement and the syntax of ciNsenga relative clauses. *Southern African Linguistics and Applied Language Studies* 24:277–290.

Sohng, Hong Ki. 2003. Topics in the syntax of East Asian languages: Long-distance anaphora and adverbial case. Doctoral Dissertation, University of Washington.

Sohng, Hong-Ki. 2004. A minimalist analysis of X^0 reflexivization in Chinese and Korean. *Studies in Generative Grammar* 14:375–396.

Son, Gwangrak. 2003. The contrastive study of monomorphemic reflexives: Japanese *zibun* and Korean *casin*. *MIT Working Papers in Linguistics* 45:233–252.

Stevens, S.S. 1975. *Psychophysics: Introduction to its perceptual, neural, and social prospects*. New York: John Wiley.

Storoshenko, Dennis Ryan. 2007. A bound variable account of the Korean reflexive *caki*. In *Proceedings of the 26th West Coast Conference on Formal Linguistics (WCCFL 26)*, ed. Charles B. Chang and Hannah J. Haynie, 438–444.

Storoshenko, Dennis Ryan. 2009. External argument adjunct phrases in English. In *Proceedings of the 2009 Annual Conference of the Canadian Linguistics Association*.

Storoshenko, Dennis Ryan, and Chung-hye Han. 2010. Binding variables in English: An analysis using delayed tree locality. In *Proceedings of the 10th International Workshop on Tree Adjoining Grammars and Related Formalisms*, ed. Srinivas Bangalore, Robert Frank, and Maribel Romero, 143–150.

Storoshenko, Dennis Ryan, Chung-hye Han, and David Potter. 2008. Reflexivity in English: An STAG analysis. In *Proceedings of the 9th International Workshop on Tree Adjoining Grammars and Related Formalisms*, ed. Claire Gardent and Anoop Sarkar, 149–157.

Tajima, Kazuhiko. 1987. On the functional determination of variables. *University of Washington Working Papers in Linguistics* 9:1–10.

Tenny, Carol. 2003. Short distance pronouns in representational noun phrases and a grammar of sentience. URL `www.linguist.org`, manuscript.

Vandall, Peter, and Joe Douquette. 1987. *Wâskahikaniwiyiniw-âcimowina = Stories of the house people: told by Peter Vandall and Joe Douquette*. Winnipeg MB: University of Manitoba Press.

Vijay-Shanker, K., and Aravind Joshi. 1988. Feature structure based Tree Adjoining Grammars. In *Proceedings of COLING'88*, 714–719.

Visser, Marianna. 2008. Definiteness and specificity in the isiXhosa determiner phrase. *South African Journal of African Languages* 28:11–29.

Whitecalf, Sarah. 1993. *Kinêhiyâwiwininaw nêhiyawêwin: The Cree language is our identity: the La Ronge lectures of Sarah Whitecalf*. Winnipeg MB: University of Manitoba Press.

Williams, Edwin. 1997. Blocking and anaphora. *Linguistic Inquiry* 28:577–628.

Woolford, Ellen. 1995. Why passive can block object marking. In *Theoretical approaches to african linguistics*, ed. Akinbiyi Akinlabi, 199–215. Africa World Press.

Xue, Ping, and Fred Popowich. 2002. Middle-distance reflexives. *Journal of Linguistics* 38:71–86.

Yang, Dong-Whee. 1982. Control and binding in Korean. *Linguistic Journal of Korea* 7:257–283.

Yoon, Jeong-Me. 1989. Long-distance anaphors in Korean and their cross-linguistic implications. In *Papers from the 25th Annual Meeting of the Chicago Linguistic Society*, ed. Caroline Wiltshire, Randolph Graczyk, and Music Bradley, 479–495. Chicago Linguistic Society.

Zribi-Hertz, Anne. 1989. Anaphor binding and narrative point of view: English reflexive pronouns in sentence and discourse. *Language* 65:695–727.