Grammar of Binding in the Languages of the World: Innate or Learned?

Authors:

Peter Cole University of Delaware and MPI EVA Leipzig

University of Delaware

125 East Main Street

Newark, DE – 19716, USA

email: pcole@udel.edu

Gabriella Hermon University of Delaware and MPI EVA Leipzig

University of Delaware

125 East Main Street

Newark, DE – 19716, USA

email: gaby@udel.edu

Yanti Universitas Katolik Indonesia Atma Jaya

Universitas Katolik Indonesia Atma Jaya

Pusat Kajian Bahasa dan Budaya

Jl. Jendral Sudirman, 51 - Gedung K2 Lantai 2

Jakarta – 12930, Indonesia

Telephone: (+62)-21-5703306 ext. 213; Fax: (+62)-21-5719560

email: reginayanti@yahoo.com

Corresponding author: Yanti

Abstract

Languages around the world often appear to manifest nearly identical grammatical properties, but, at the same time, the grammatical differences can also be great, sometimes even seeming to support Martin Joos's (1958) claim that "languages can differ from each other without limit and in unpredictable way" (p. 96). This state of affairs provides a puzzle for both nativist approaches to language like Generative Grammar that posit a fixed "Universal Grammar", and for approaches that minimize the contribution of innate grammatical structure. We approach this puzzling state of affairs by looking at one area of grammar, "Binding", the system of local and long distance anaphoric elements in a language. This is an area of grammar that has long been central to the Generative approach to language structure. We compare the anaphoric systems found in "familiar" (European-like) languages that contain dedicated classes of bound and free anaphors (pronouns and reflexives) with the anaphoric systems in endangered Austronesian languages of Indonesia, languages in which there is overlap or no distinction between pronouns and reflexives (Peranakan Javanese and Jambi Malay). What is of special interest about Jambi anaphora is not only that conservative dialects of Jambi Malay do not distinguish between pronouns and reflexives, but that Jambi anaphora appear to constitute a live snapshot of a unitary class of anaphora in the process of grammaticalization as a distinct system of pronouns and reflexives.

We argue that the facts of Jambi anaphora cannot be explained by theories positing a Universal Grammar of Binding. Thus, these facts provide evidence that complex grammatical systems like Binding cannot be innate. Our results from Austronesian languages are confirmed by data from signed and creole languages. Our conclusion is that the human language learning capacity must include the ability to model the full complexity found in the syntax of the world's

GRAMMAR OF BINDING

languages. From the perspective of child language acquisition, these conclusions suggest that Universal Grammar does not provide a general solution to the problem of poverty of the stimulus, and the solution to that problem must reside at least in part in special properties of the grammar construction tools available to the language learner rather than simply in a fixed set of grammatical rules hard wired into the brains of speakers.

Keywords: anaphora; binding; innate; learned; syntax; Indonesian

1. Introduction

Unrelated languages in diverse parts of the world are often very similar in many aspects of their grammatical structure, but, despite this, the range of structural differences can be startlingly great.¹ Linguists have responded to this conundrum in different ways. At one end of the spectrum, Martin Joos (1958) is famous for having made the often derided claim that “languages can differ from each other without limit and in unpredictable ways” (p. 96). This position has been reaffirmed by Evans and Levinson (2009), who state that “languages differ so fundamentally from one another at every level of description (sound, grammar, lexicon, meaning) that it is very hard to find any single structural property they share. The claims of Universal Grammar... are either empirically false, unfalsifiable, or misleading in that they refer to tendencies rather than strict universals” (p. 429).

At the other end of the spectrum, within Generative Grammar, the emphasis has been on explaining the recurring similarities among languages and on attempting to demonstrate that the dissimilarities are more apparent than real. Crain, Khlentzos, and Thornton (2010) sees the core problem for linguistic theory to be that of providing an explanation for “the fact that the same cluster of linguistic properties crops up in language after language, including typologically different languages” (p. 2671).

Prior to *The Minimalist Program* (Chomsky, 1995) it was widely assumed among Generative linguists that it is the initial state for infants acquiring their native language that is invariant, leading to the claim that many properties of language are hard wired in the “language

¹ This contrast was expressed clearly by Emmon Bach, in his Presidential Address at the 1996 Annual Meeting of the Linguistic Society of America in San Diego.

GRAMMAR OF BINDING

acquisition device” and constitute Universal Grammar (UG). While the initial state of the language learner was assumed to be invariant, this state was assumed to be subject to alternation on the basis of experience (exposure of the child language learner to actual languages), resulting in adult languages that could be very different from the initial state dictated by UG. In *The Minimalist Program* this stance was strengthened, and the claim was put forward that “the primary [task of linguistics] is to show that the apparent richness and diversity of linguistic phenomena is illusory and epiphenomenal, the result of interaction of fixed principles under slightly varying conditions” (Chomsky, 1995, p. 8). The existence of “fixed principles” is taken to mean that all languages share the same grammatical rules and constraints, and that the seeming differences among rules in different languages are in fact only attributable to properties of the vocabulary items found in each individual language rather than to variations in the rule systems themselves: Languages are hypothesized not to exhibit differences in rule systems despite appearances to the contrary (pace e.g. Crain, 2010 and many others):²

The standard idealized model of language acquisition takes the initial state S_0 to be a function mapping experience (primary linguistic data, PLD) to a language. UG is concerned with the invariant principles of S_0 and the range of permissible variation. Variation must be determined by what is “visible” to the child acquiring language, that is, by the PLD. It is not surprising, then, to find a degree of variation in the PF [phonetic form] component, and in aspects of the lexicon: Saussurean arbitrariness (association of

² This is a much stronger position than the claim that “universal grammar” provides the initial state for the acquisition of grammar, but that children modify this initial state, and, as adults, end up with rule systems that can be quite different from each other.

concepts with phonological matrices), properties of grammatical formatives (inflection, etc.), and readily detectable properties that hold of lexical items generally (e.g., the head parameter). Variation in the overt syntax or LF component would be more problematic, since evidence could only be quite indirect. **A narrow conjecture is that there is no such variation: beyond PF options and lexical arbitrariness (which I henceforth ignore), variation is limited to nonsubstantive parts of the lexicon and general properties of lexical items.** If so, there is only one computational system and one lexicon apart from this limited kind of variety. Let us tentatively adopt that assumption—extreme perhaps but it seems not implausible—as another element of the Minimalist Program (Chomsky, 1995, p. 169-170, emphasis added).

Thus, with respect to the rule system, UG is taken to fully determine the syntactic system of natural languages: at least with regard to syntactic principles, all languages are fundamentally the same.

The Minimalist view that all languages are essentially the same is motivated by theoretical considerations, but not by theoretical considerations alone. This stance, in fact, reflects an experience that is shared by linguists of all stripes, the experience that when a new language is examined, the vast majority of the structural properties of the language will be familiar, once idiosyncratic aspects of the language like morphology and the lexicon are factored out and attention is turned to the syntax and semantics. For instance, to anticipate the topic of this paper, it is unsurprising to linguists when there exist in a language two types of anaphoric forms, those requiring c-commanding antecedents and subject to stringent locality restrictions (reflexives like *himself*), and those that cannot be used when there is a c-commanding, local antecedent

(pronouns like *him*).³ From the familiar division between reflexives and pronouns it is a short step to the assumption typically made by Generative linguists (an assumption often is not shared by non-linguists or by non-Generative linguists) that the similarities found among languages with respect to the properties of anaphoric elements are due to the existence of a module of an innate language faculty (UG), specifically to some variant of the principles of the “Binding Theory”. The Generative approach is to be contrasted with an alternative approach to explaining these facts, one that does not claim that these properties are determined by UG, which would be that factors such as the nature of the communicative act, perceptual principles determining the overall characteristics of the human mind etc., principles which are not strictly linguistic in nature, lead to the fact that similar divisions of anaphoric elements are found in languages that are otherwise quite dissimilar. We will return to alternatives to UG with respect to Binding Theory later in this article.

Our purpose in this article is to approach these issues by examining further the implications of the system of local and long distance anaphoric elements generally referred to as Binding Theory with respect to the general claim that the complex properties of the syntactic component of language must be largely innate and that the seeming diversity of language structure can be reduced to the interaction of fixed principles in diverse contexts. The study of

³ C-command is a structural relationship between elements in a syntactic tree: A c-commands B iff the lowest branching node that dominates A also dominates B. C-command has been shown to be fundamental to specifying the syntactic relations that are required between grammatical elements (e.g. Frank, 2002). Locality restrictions specify how far in the tree elements of a particular type may be from each other.

GRAMMAR OF BINDING

Binding is appropriate for this purpose because Binding was one of the original component modules of UG, and, more important, it remains an area that Generative linguists view as central to the Generative enterprise. See, for example, the recent work by Safir, Reuland and others on anaphora and language design (e.g. Reuland, 2011 and many earlier works; Safir, 2004 and many additional works).⁴

What is most important for our purposes is that Binding displays the sort of complexity that Poverty of Stimulus arguments for the centrality of UG would claim to be unlearnable from experience: The syntactic facts of the anaphoric system (which sentences are well formed and which are not) are under determined by the data available to the language learner. Since the language learner is exposed primarily to grammatical sentences and does not receive any information regarding which sequences are ill-formed, he lacks critical information that is used e.g. by the linguist to determine the syntactic rules of the language. Therefore, if the speaker can be shown empirically to have knowledge of e.g. the anaphoric system of a sort unavailable from input, there is an apparent contradiction: Speakers have knowledge that they cannot have learned.

UG constitutes a general solution to the problem of the poverty of the stimulus, and would be expected to provide the solution to this dilemma in the realm of Binding as well as in other areas of syntax. If, however, it should turn out that Binding is not determined by UG and, hence, must be learned by the language learner (as we claim below in fact is the case), then Binding

⁴ Reuland's theory of Binding is considerably more complex than the classic Binding Theory, but the core claim that anaphoric relations are determined by principles of universal grammar is maintained.

GRAMMAR OF BINDING

must be learnable *a priori* claims to the contrary. For this reason, Binding constitutes a good test case for the claim that the facts of syntax generally cannot be accounted for without making the assumption that much of our knowledge of syntax is innate (i.e. determined by UG).

With this general question in the background, we turn to the organization of the paper. We shall start our discussion, as a point of reference, with the familiar system, discussed in much of the Generative literature on English, consisting of a two-way contrast between reflexives and pronouns, which has been taken as the model for the Binding Theory. From there, we shall turn to more exotic systems found in the Austronesian languages of Indonesia, first to languages in which part of the anaphoric system consists of “Binding Theory exempt anaphors”, languages that include anaphors that appear not to respect the distinctions between reflexives and pronouns found in systems displaying a two-way contrast. The languages we shall discuss also contain Binding Theory compatible anaphors (anaphors that do accord with that distinction). We then turn to a language in which two closely related dialects occur. In one dialect, the Binding Theory as a set of grammatical constraints does not apply at all. Rather, in this language there are only Binding preferences, which appear to be determined purely by functional considerations rather than by strict grammatical rules. We compare this dialect with a closely related dialect in which hard and fast grammatical Binding constraints do occur. The preferences exhibited by the first dialect are clearly not predicted by Binding Theory since, by hypothesis, UG determines what is possible in language, not what is preferred for functional reasons. In contrast, the second dialect exhibits a more rigid version of the same preferences: what are preferences in the first dialect are hard and fast rules in the second. The pattern observed in the second dialect is in fact essentially identical with that predicted by Binding Theory. We argue that a UG-based theory of anaphoric relations would have to claim that the systems exhibited in the two dialects are

unrelated, one determined by Binding Theory and the other by pragmatic principles, but this claim would miss the fact that two systems are nearly identical, the only difference being the fact that what is preferred in one dialect is required in the other. We therefore suggest that these dialects constitute, in effect, a live snapshot of Binding Theoretic reflexives and pronouns in the process of grammaticalization as a distinct system of pronouns and reflexives. These facts appear incompatible with a theory in which Binding, and complex syntax in general, are specified by UG rather than learned.

In the final sections of the paper we consider the implications of our claims for anaphora with regard to pidgins and signed languages, both (potentially, at least), instances of recently created languages. In both cases we predict that the use of a single indeterminate form in both reflexive and non-reflexive contexts will occur with much greater frequency than in languages displaying a long history of slow grammatical changes. The facts appear to support our claim that the distinction between reflexive and non-reflexive pronouns is based on the grammaticalization of functional aspects of language rather than on hard wired principles of universal grammar which would be expected to be especially visible in languages created *de novo*.

Our main claim in this article is that when the domain of Binding is considered across languages, including languages displaying “exotic” systems of Binding, it can be seen that the rule systems involved in Binding are, in fact, learned rather than based entirely on “the

interaction of fixed principles under slightly varying conditions.”⁵ In evaluating the results of our study, we return to the core contrast between pronouns and reflexives. Our overall claim in the paper is that the distinction between these core constructs of Binding is not attributable to UG, but rather it appears to derive from the grammaticalization of functionally-based properties of anaphoric noun phrases. Despite this, languages contain complex and formally characterizable rule **systems** governing when each of these grammaticalized forms is sanctioned in the language. This suggests that language specific rule systems cannot be epiphenomenal, as the Minimalist Program proposes. If our claims are correct, it cannot be Universal Grammar plus properties of the vocabulary of the language alone that constitute the totality of our grammatical competence. Language learners must be capable of acquiring complex syntactic **systems** on the basis of imperfect data. Thus, UG does not provide a general solution to the problem of poverty of the stimulus, and the solution to that problem must reside at least in part in special properties of the grammar construction tools available to the language learner rather than simply in a fixed set of grammatical rules hard wired into the brains of speakers.

2. Reflexives and Pronouns

We will start our discussion by presenting as a reference point an instance of a language that makes a clear distinction between pronouns and reflexives. We could, of course, use English examples for this purpose, but, in order to emphasize the fact that the contrasting

⁵ We would like to emphasize that we argue that UG as it has been proposed does not account for the range of facts to be discussed here. We cannot, of course, demonstrate that no system of UG could account for these facts.

categories of reflexive and pronoun recur in genetically unrelated and typologically dissimilar languages around the world, we shall use Mandarin Chinese.⁶ The core characteristics of reflexives and pronouns, according to the classical Binding Theory, are roughly as in (1) and (2):

(1) Reflexives (bound anaphora)

- a) an antecedent is required;
- b) the antecedent must c-command the reflexive;
- c) the antecedent must stand in a local relationship to the reflexive;⁷

⁶ Please note, however, in addition to garden variety reflexives and pronouns, Mandarin and other Chinese languages also have “long distance reflexives,” which, for reasons of space we will not discuss in this paper. Much of the literature on long distance reflexives is aimed at showing that these forms do not constitute evidence against the universality of the classical Binding Theory. (See Cole, Hermon, & Huang, 2006; Huang, Li, & Li, 2009, in which it is argued that this effort has been successful, but that long distance reflexives are subject to significant pragmatic constraints that make it difficult to observe clearly which aspects of the distribution of long distance reflexives are due to Binding, which are due to constraints on movement and which are due to pragmatic considerations).

⁷ The definition of “local” has always been problematic. As used in the literature, a local antecedent includes at the very least c-commanding clause mates of the anaphoric form, more typically the domain of the closest “subject” that c-commands the anaphor. The exact definition of locality is not important for this paper.

GRAMMAR OF BINDING

(2) Pronouns (free anaphora)

- a) the presence of an antecedent is optional;
- b) the antecedent may either c-command the pronoun or not;
- c) the antecedent must not stand in a local relationship to the pronoun.

These characteristics of reflexives and pronouns are illustrated with examples from English and Mandarin.

- (3) a. Mary_i hurt herself_{i,*j} =someone other than Mary
b. Mary_i's sister_j /Mary_j hurt herself_{*i,j,*k} =someone other than Mary or her sister

- (4) a. Mali_i haile [ta ziji]_{i,*j} =someone other than Mary
Mary hurt 3 self
'Mary hurt herself.'
b. Mali_i de jiejie_j haile [ta ziji]_{*i,j,*k} =someone other than Mary
Mary_i's sister_j hurt 3 self
'Mary's sister hurt herself.'

- (5) John_k thinks Sam_j knows Ted_i likes himself_{i,*j,*k}

- (6) Zhangsan_k renwei Lisi_j zhidao Wangwu_i xihuan ta ziji_{i,*j,*k}
Zhangsan think Lisi know Wangwu like 3 self
'Zhangsan thinks Lisi knows Wangwu likes himself.'

We can see from these examples that the reflexive forms *himself* and *ta ziji* follow the same pattern. The antecedent for the reflexive must c-command the reflexive, as seen in (3)-(4):

Mary/Mali c-commands *herself/ta ziji* when the subject is a simple noun phrase, *Mary/Mali*, but

GRAMMAR OF BINDING

not when the subject is a possessive noun phrase, *Mary's sister/Mali de jiejie*. Similarly, the antecedent for the reflexive must be local with respect to the reflexive. In (5) and (6) the only possible antecedent for the reflexive is the subject of the most deeply embedded clause.

Furthermore, reflexives, in contrast to pronouns, receive a sloppy rather than a strict interpretation in ellipsis contexts:

(7) a. In VP ellipsis and similar constructions, reflexives show a strong preference for a sloppy interpretation.

b. In VP ellipsis and similar constructions, pronouns exhibit both strict and sloppy interpretations.

(8) Mary hurt herself; her sister did too.
'Mary and her sister each hurt herself,' and not 'Mary's sister hurt Mary.'

(9) Mali_i haile [ta ziji]_i; ta de jiejie yiyang.
Mary hurt herself she 's sister the.same
'Mary and her sister each hurt herself,' and not 'Mary's sister hurt Mary.'

Pronouns in English and Mandarin also appear to follow the same pattern:

(10) a. Mary_i's sister_j hurt her_{i, *j}, k=someone other than Mary or her sister

b. Mary_i hurt her_{*i, j=someone other than Mary}

(11) a. Mali_i de jiejie_j/Mali_j haile ta_{i, *j}, k=someone other than Mary
Mary's sister Mary hurt 3
'Mary hurt her.'

GRAMMAR OF BINDING

- b. Mali_i haile ta_{*i, j}=someone other than Mary
Mary hurt 3
'Mary hurt her.'

The same pattern is also found in Chinese and English with respect to VP ellipsis, but we will omit examples for reasons of space.

The pattern just described receives a plausible explanation in terms of the classical Binding Theory:

Principle A: An anaphor is bound in a local domain.

Principle B: A pronominal is free in a local domain.

Principle C: An R-expression is free (in the domain of the head of its maximal chain).

(Chomsky, 1986, p. 166)

(For reasons of space, we will not discuss how the less obvious properties of anaphors, e.g. the differing properties of pronouns and reflexives in elliptical contexts, might follow from the classical Binding Theory.)

We know that Mandarin and English are about as distantly related as are any two languages on the planet, and the languages are typologically quite dissimilar. Additionally, the pattern just described is shared with a very large number of unrelated and otherwise dissimilar languages. Thus, it is reasonable (although, we will ultimately conclude, wrong) to suppose that the similarities in Binding between Mandarin and English reflect the characteristics of a Binding component that is an integral sub-system in the rule system of UG.

3. Binding Theory Exempt Anaphors

We will now turn to Binding phenomena that do not appear to be reducible to the classical Binding Theory. In many varieties of Malay/Indonesian (Cole & Hermon, 1998a, 1998b, 2005), and in some other Austronesian languages of Indonesia, like Peranakan Javanese (Cole, Hermon, Tjung, Sim, & Kim, 2007, 2008), the anaphoric system consists of three forms, a garden variety local reflexive, a well behaved personal pronoun and a form that is problematic with respect to the classical Binding Theory because it is indeterminate or unspecified with regard to whether it is a pronoun or a reflexive.⁸ This form is one that, we shall argue, is exempt from the principles that restrict the distribution of pronouns and reflexives in the language. We shall first discuss anaphora in Peranakan Javanese (essentially the same facts hold in Standard Malay/Indonesian). In a later section we will then turn to the anaphoric forms in Traditional Jambi Malay (TJM), and compare the facts of TJM with those of the dialect of Jambi Malay spoken in Jambi City (Jambi City Malay, or JCM). Our purpose is to show that both the Peranakan Javanese and the Jambi facts present a serious challenge to UG-based approaches to binding. We will argue that TJM is an instance of a language that lacks Binding Theory compliant anaphora altogether.

As illustrated in (12), the anaphoric expression *awake dheen* in Peranakan Javanese appears on initial examination to be a local reflexive like English *himself*.⁹

⁸ Malay (Bahasa Malaysia) and Indonesian (Bahasa Indonesia) are two slightly different standardizations of the same language.

⁹ Anaphoric expressions using the morpheme *awak* are only etymologically related to ‘body’. Body parts are frequent historical sources for reflexives. In the case of *awak*, when used as part

(12) a. *John* saw **himself** in the mirror.

b. *Tono_i ketok awake dheen_i nggon kaca.*
 Tono see body.3 3SG in mirror
 ‘Tono saw himself in the mirror.’

However, *awake dheen* exhibits a distribution different from that of a local reflexive like English *himself*. First, as seen in (13), *awake dheen* can refer to either a c-commanding antecedent like *gurue Tono*, a non-c-commanding antecedent like *Tono*, or a discourse antecedent (*k* indicates a discourse antecedent).

(13) [*Gurue Tono_j*]_i *ketok awake dheen_{i/j/k} nggon kaca.*
 teacher.3 Tono see body.3 3SG in mirror
 ‘Tono’s teacher saw himself/him/her in the mirror.’

Second, unlike local reflexives, *awake dheen* can refer either to a local antecedent, a non-local antecedent or to a discourse antecedent, as illustrated in (14).

of an anaphoric expression it is clearly bleached of its original meaning and can be used in contexts where ‘body’ would be inappropriate:

Sak-durunge Mary_i nulis surat nggo awake dheen_{i/j/k}, John_j wis balek omah
 before-3 Mary N.write letter for body.3 3SG John already return house
 ‘Before Mary had written a letter to herself, John already returned home.’

GRAMMAR OF BINDING

- (14) *Ali_j ngomong nek aku pikir [Tono_i ketok awake dheen_{i/j/k} nggon kaca].*
 Ali N.say COMP 1SG think Tono see body.3 3SG
 in mirror
 ‘Ali said that I thought that Tono saw himself/him in the mirror.’

The data seen in (12)–(14) indicate that, from the perspective of the classical Binding Theory, *awake dheen* has an unexpected distribution. Either it is ambiguous between a pronoun and a reflexive or it is a general anaphoric form that is not specified as either pronominal or reflexive.

We shall examine the predictions of these hypotheses below, but before we do so, we shall survey the other anaphoric forms available in Peranakan Javanese. In addition to *awake dheen*, Peranakan Javanese has two other forms, *awake dheen dhewe* and *dheen*, that appear well behaved from the perspective of the classical Binding Theory, and which therefore can be categorized as a reflexive and a pronoun, respectively. First, as seen in (15), *awake dheen dhewe* behaves exactly as expected for a local reflexive, i.e., it can only be bound by a c-commanding antecedent like *gurue Tono* ‘Tono’s teacher’; it cannot refer to a non-c-commanding potential antecedent like the possessor DP *Tono*, or a discourse antecedent.

- (15) [*Gurue Tono_j ketok awake dheen dhewe_{i/*j/*k} nggon kaca.*]
 teacher.3 Tono see body.3 3SG self in mirror
 ‘Tono’s teacher saw himself in the mirror.’

Second, as shown in (16), *awake dheen dhewe* cannot refer to a non-local antecedent.

- (16) *Ali_j ngomong nek aku pikir [Tono_i ketok **awake dheen**¹⁰*
 Ali N.say COMP 1SG think Tono see body.3 3SG
***dhewe**_{i/*j/*k} nggon kaca].*
 self in mirror
 ‘Ali said that I thought that Tono saw himself in the mirror.’

The form *dheen*, in contrast, behaves exactly as expected for a pronoun. As the examples below show, *dheen* exhibits the typical distribution of a pronoun and is in complementary distribution with *awake dheen dhewe*, i.e., it cannot be bound by a local c-commanding antecedent like *gurue Tono* ‘Tono’s teacher’, but it can refer to a non-c-commanding antecedent like the possessor DP *Tono*, a discourse antecedent, or a non-local c-commanding antecedent.

- (17) [*Gurue Tono_j]_i ketok **dheen**_{*i/j/k} nggon kaca.*
 teacher.3 Tono see 3SG in mirror
 ‘Tono’s teacher saw him/her in the mirror.’

- (18) *Ali_j ngomong nek aku pikir [Tono_i ketok*
 Ali N.say COMP 1SG think Tono see
***dheen**_{*i/j/k} nggon kaca].*
 3SG in mirror
 ‘Ali said that I thought that Tono saw him/her in the mirror.’

In order to capture the properties of *awake dheen*, we propose that it should be analyzed as an anaphoric form that is indeterminate, or unspecified in the lexicon, for the features [α pronominal/ β anaphor]. That is, *awake dheen* is analyzed as a form that is anaphoric in the non-

¹⁰ There are no blocking effects like those found in Chinese. See Cole et al. (2007) for details.

technical sense of “anaphor”, but that is not subject to the Binding Theory.¹¹ (We will later compare this analysis with one according to which *awake dheen* is ambiguous between a pronoun and a reflexive.)

Let us see how the unspecified anaphor analysis accounts for the data in (12b), (13), and (14). According to our analysis, *awake dheen* is neither a reflexive nor a pronoun. Thus, it is able to take a c-commanding antecedent like *Tono* in (12b), or a non-c-commanding antecedent like *Tono* in (13). A long distance or a discourse antecedent would also be possible since the Binding Theory imposes no locality restrictions on *awake dheen*, and it therefore can refer to a non-local antecedent like *Ali* in (14).

If this analysis of *awake dheen* is correct, it constitutes a serious challenge for UG-based approaches to Binding. The presence in a language of a form that is used anaphorically but which is exempt from the Binding requirements of UG would impose a considerable burden on the child acquiring the language. The problem is that a child learning to speak a language would need to learn which forms that are functionally anaphoric in their use are subject to UG principles of Binding and which are not. The existence of UG sanctioned categories for anaphora simplifies learning only if all anaphoric elements (in the non-technical sense of “anaphoric” that includes both pronouns and reflexives) are subject to UG principles. Thus, in the context of a system containing a mixture of UG compliant and UG exempt elements, UG principles do not provide a solution to the poverty of the stimulus problem. At the least, the distinction between exempt and non-exempt forms must be learned from experience, i.e. from observed characteristics of the language, and it is doubtful that the data is sufficiently structured

¹¹ See discussion in Cole et al. 2007, 2008.

to make this distinction learnable on the basis of the distributional properties of the data that are available to the child. Thus, the existence of UG exempt anaphors has the effect of greatly vitiating the value for language acquisition of a UG theory of Binding.

In light of the serious consequences of positing Binding Theory Exempt anaphors, it is necessary to examine alternative analyses for the facts. In Cole et al. (2007, 2008), we consider a variety of alternative analyses and show why they do not account for the distribution of *awake dheen*. In the present article, we will restrict ourselves to showing why we reject the strongest potential counter analysis, that *awake dheen* is ambiguous between a local reflexive and a pronoun rather than indeterminate with regard to the distinction. We shall argue that ambiguity and indeterminacy make different empirical predictions about what interpretations are possible when *awake dheen* is used in sentences involving VP ellipsis, and that in these cases the ambiguity hypothesis makes the wrong predictions.

Let us look at the consequences of treating *awake dheen* as ambiguous between a local reflexive and a pronoun. If the Ambiguity Hypothesis is correct each of the forms might be Binding Theory compliant, and, hence, not a problem for linguistic theory. We shall show, however, that the Ambiguity Hypothesis makes incorrect predictions, and, as a result, that *awake dheen* cannot be analyzed as specified for the features [α pronominal/ β anaphor].

Our argument is based on the interpretations available for *awake dheen* in ellipsis constructions. The critical examples to consider are those involving the local use of *awake dheen* in VP-ellipsis.

- (19) Tono ketok **awake dheen** nggon kaca, Siti yaya.
 Tono see body.3 3SG in mirror Siti also
 ✓ Strict: ‘Siti saw Tono in the mirror.’
 ✓ Sloppy: ‘Siti saw Siti in the mirror.’

In (19), both the strict and sloppy interpretations are possible, as depicted in (20). It can mean either ‘Tono saw Tono in the mirror and Siti saw Siti in the mirror’ (the sloppy reading) or ‘Tono saw Tono in the mirror and Siti saw Tono in the mirror’ (the strict reading).

(20) Strict Reading Well-formed; Sloppy reading Well-formed

a. Sloppy Reading

$\lambda x.[x \text{ saw } x \text{ in the mirror}] (\text{Tono}) \wedge \lambda y.[y \text{ saw } y \text{ in the mirror}] (\text{Siti})$

b. Strict Reading

$[\text{Tono saw } x_1 \text{ in the mirror} \wedge \text{Siti saw } x_1 \text{ in the mirror}] \text{ }^g [x_1 \rightarrow \text{Tono}]$

Bearing the above facts in mind, let us now see how the Ambiguity Analysis would account for (19). We shall then compare the predictions of the Ambiguity Analysis versus those of the Indeterminacy Analysis.

Under the Ambiguity Analysis, there are two possibilities if a sentence contains *awake dheen*. The overt form *awake dheen* must be either a pronoun or a reflexive. If the overt VP contains a reflexive, the unpronounced clause in (19) needs to contain a reflexive as well, since by hypothesis, the unpronounced element is an identical copy of that overt element. Let us assume that (19') underlies (19) under the Ambiguity Analysis.

(19') Tono ketok **awake dheen** [+reflexive] nggon kaca,
 Tono see body.3 3SG in mirror
 Siti yaya ketok **awake dheen** [+reflexive] nggon kaca.
 Siti also see body.3 3SG in mirror
 ‘Tono saw himself in the mirror, and Siti also saw herself in the mirror.’

GRAMMAR OF BINDING

As shown in (19'), when *awake dheen* is used as a reflexive in the first clause, it refers to *Tono*. Note that if it is used as a pronoun, it cannot refer to *Tono*, so it must be interpreted as a reflexive when it refers to *Tono*. In the second clause it must also be a reflexive and, therefore, can refer only to the local c-commanding antecedent, *Siti* (the sloppy reading). The strict reading is predicted not to be available in the second (ellipsis) clause. This, however, is incorrect. Thus, the Ambiguity Analysis fails to account for the fact that when used locally in VP-ellipsis, the unpronounced instance of *awake dheen* in the ellipsis clause can receive either the sloppy reading or the strict reading. We conclude that the Ambiguity Analysis fails to account for the distribution of the pseudo-reflexive *awake dheen*. We therefore claim that *awake dheen* cannot be ambiguous between a reflexive and a pronoun.

Let us see now how the Indeterminacy Analysis accounts for the peculiar distribution of *awake dheen* in (19). It will be remembered that, according to the Indeterminacy Analysis, *awake dheen* is neither a reflexive nor a pronoun. Thus, it is able to take a c-commanding antecedent like *Tono* in (21a) or a non-c-commanding antecedent like *Tono* in (21b).

- (21) a. Tono_i ketok **awake dheen**_i nggon kaca.
 Tono see body.3 3SG in mirror
 ‘Tono saw himself in the mirror.’
- b. [Gurue Tono_j]_i ketok **awake dheen**_{i/j/k} nggon kaca.
 teacher.3 Tono see body.3 3SG in mirror
 ‘Tono's teacher saw himself/him/her in the mirror.’

A long distance or a discourse antecedent is also possible since the Binding Theory imposes no locality restrictions on *awake dheen* and it can therefore refer to a non-local antecedent like *Bowo* in (22).

GRAMMAR OF BINDING

- (22) Bowo_j ngomong nek aku pikir [Tono_i ketok
 Bowo N.say COMP 1SG think Tono see
awake dheen_{i/j/k} nggon kaca].
 body.3 3SG in mirror
 ‘Bowo said that I thought that Tono saw himself/him in the mirror.’

Let us now turn to the critical question at issue, the use of *awake dheen* in VP-ellipsis. Because *awake dheen* is analyzed as lacking grammatical features specific to anaphoric forms like [α pronominal]/[β anaphor], it would be expected that such a form would receive the union of the interpretations available for the pronoun *dheen* (strict and sloppy interpretations) and the reflexive *awake dheen dhewe* (sloppy interpretation). Thus, *awake dheen* should receive both the strict and the sloppy interpretations. Consider again example (19), repeated below.

- (19) Tono ketok **awake dheen** nggon kaca, Siti yaya.
 Tono see body.3 3SG in mirror Siti also
 ✓ Strict: ‘Siti saw Tono in the mirror.’
 ✓ Sloppy: ‘Siti saw Siti in the mirror.’

As we saw earlier, the Ambiguity Analysis would predict that the only possible interpretation is the sloppy interpretation. According to the Indeterminacy Analysis, however, *awake dheen* in the antecedent (overt) clause is neither a reflexive nor a pronoun. Thus, it is able to take *Tono* as an antecedent since no anti-locality requirements (like those on pronouns) apply to this form. In the ellipsis clause, the same requirements hold as in the antecedent clause. *Awake dheen* can refer to either a local antecedent, a long distance antecedent, or a discourse antecedent. There is, however, an additional requirement in the ellipsis clause: *Awake dheen* in the ellipsis clause must be identical to *awake dheen* in the antecedent clause. This can be

GRAMMAR OF BINDING

satisfied in either of two ways, either by referring to the same individual as *awake dheen* in the antecedent clause (leading to the strict reading) or by constituting the same variable as *awake dheen* in the antecedent clause (leading to the sloppy reading). These possibilities exist in Peranakan Javanese because (by hypothesis) features specific to anaphoric forms like [α pronominal]/[β anaphor] are not present, and, as a result, identity with respect to these features is not part of the identity requirements for ellipsis. Thus, under the strict reading, the elliptical instance of *awake dheen* satisfies the identity requirements on VP ellipsis even though the referent of *awake dheen* is local in the antecedent clause and a discourse referent in the ellipsis clause.

Similarly, when *awake dheen* in the ellipsis clause receives a sloppy interpretation, it also satisfies the identity requirements on VP ellipsis. In this case, the relevant property of overt *awake dheen* is taken to be the fact that it is a bound variable. Thus, the relevant identity requirement satisfied by covert *awake dheen* in the ellipsis clause is that it is a bound variable as well. The fact that both the covert and the overt instances of *awake dheen* have local antecedents is of no relevance to the grammaticality of the sentence since neither the overt instance of *awake dheen* nor the covert instance in the ellipsis clause is subject to the locality requirements of Condition A and Condition B of the Binding Theory since neither is marked as [α pronominal]/[β anaphor]. We therefore conclude that *awake dheen* in Peranakan Javanese should be analyzed as unspecified in the lexicon for the syntactic features [α pronominal]/[β anaphor] and not as an ambiguous form that can be either a pronoun or a reflexive.

To summarize our argument so far, we have shown that Peranakan Javanese, like some other Austronesian languages of Indonesia, contains a form used as an all purpose anaphor that is unspecified for the defining features of pronouns and reflexives. This form exists in addition to

reflexives and pronouns that are compliant to the Binding Theory. The existence of Binding Theory exempt forms is a serious challenge for a UG account of Binding phenomena because it is unclear how, on the basis of the data available to the child, the language learner would be able to determine when UG principles should or should not apply to a given anaphoric form.

4. A Language without Binding Theory Compliant Anaphors

In the sections that follow we shall discuss the anaphoric forms in Traditional Jambi Malay (TJM), and compare the facts of TJM with those of the dialect of Jambi Malay spoken in Jambi City (Jambi City Malay, or JCM).¹² Our purpose is to show that the Jambi facts present an even more serious challenge to UG-based approaches to binding than do the Peranakan Javanese facts. We will argue that TJM is an instance of a language that lacks Binding Theory compliant anaphora altogether. If we are correct, this fact is of interest in and of itself, but it is not entirely novel. Similar claims were made for earlier forms of English, for instance, by Keenan (2002), Kiparsky (2002) and others. In Cole, Hermon, and Yanti (2010), we discuss the absence of dedicated reflexives in Jambi Malay and in a number of Austronesian languages, especially Oceanic languages. For instance, as is shown in Besnier (2000), there is no dedicated reflexive in Tuvaluan. We will not repeat that discussion here. What we claim is of special interest about Jambi anaphora is that Jambi anaphora may constitute a live snapshot of Binding Theoretic reflexives and pronouns in the process of grammaticalization as a distinct system of pronouns and reflexives. We will claim that the anaphoric forms in TJM are instances of a functionally-based system in the stage prior (perhaps immediately prior) to grammaticalization. In contrast,

¹² See Cole, Hermon, and Yanti (2010) and Yanti (2010), for a more detailed discussion.

GRAMMAR OF BINDING

the dialect of Jambi Malay found in Jambi City (JCM), just across the Batang Hari River from the locus of the traditional dialects, shows the stage in development in which the anaphoric forms have already undergone grammaticalization. Thus, the examination of these dialects together facilitates an understanding of the nature of the change from a functionally defined system to one that is grammatically determined. The grammaticalized system observed in JCM shows the emergence of a dichotomy between reflexives and pronouns, precisely the system which the classic Binding Theory takes to be part of the genetically predetermined human language faculty. If the sequence of stages that we observe in Jambi is responsible for the classic pattern of reflexives and pronouns, then the division of anaphora into reflexives and pronouns cannot be simply a matter of compliance with UG principles, but should be attributed to the results of language change, determined by the principles governing grammaticalization rather than UG principles that determine the structure of all human languages. The facts we show could, perhaps, be interpreted differently, at least in some details, but they suggest strongly that the pattern modeled by Binding Theory is not primarily due to the interaction among principles of UG but is rather an artifact of our ability to construct grammatical systems by means of grammaticalization. The frequent occurrence among languages of a pattern in which anaphora are divided into two types, pronouns and reflexives is best analyzed as due to the independent recurrence of similar paths of grammaticalization in unrelated languages (analogous to the independent development of similar forms of e.g. shelter in different cultures due to the need to adapt to similar environmental conditions) rather than as the emergence of Binding Theory as a part of our innate knowledge of language structure.

It may be useful to start by presenting some background information about TJM, and about the relationship between Jambi Malay and the official languages of Malaysia and Indonesia,

GRAMMAR OF BINDING

Standard Malay/Indonesian (SMI). Although most theoretically-oriented studies of Malay/Indonesian have been based on SMI, in reality SMI is problematic as an object of study. Although SMI is generally viewed as a major world language with about 250 million speakers in Indonesia, Malaysia, Brunei Darussalam, and Singapore, and, is taught in schools and employed for official purposes, in fact, almost no children learn SMI as their first language. Instead, they acquire as their first language one of many colloquial varieties of Malay/Indonesian, or one of the many local languages spoken in the region (for instance, Peranakan Javanese). Unlike the colloquial varieties of Malay, the grammar of SMI, to at least some extent, has been determined by the fiat of the language planners in Malaysia and Indonesia, with, as we mentioned earlier, somewhat different standardizations in each country. Indeed, it could be argued that SMI as described by the language planners is an artificial creation rather than a natural language. The artificiality of SMI is moderated by the existence of “near standard” Malay/Indonesian, which a number of authors (including the authors of this paper) have argued is an appropriate subject for linguistic investigation, since speakers of near standard have strong and consistent intuitions about grammaticality, and there are many newspapers, books, and other publications that reflect these intuitions in naturally occurring corpora. The intuitions of speakers of near standard Malay/Indonesian often contradict both the fiat of the language planners and the grammars of the colloquial varieties of Malay/Indonesian spoken by these same individuals in other contexts. Thus, near standard can be studied as a natural language. Still, despite the legitimacy of examining the characteristics of the near standard, it is of even greater importance to study local varieties of Malay/Indonesian so as to learn the properties of the language as acquired in an unselfconscious fashion by native speakers as young children.

Our work on this topic is part of an NSF-supported investigation we conducted on the Malay spoken in the villages around the city of Jambi in Sumatra, Indonesia, and on the language spoken in the city itself. Village Jambi Malay is a “traditional” Malay dialect, spoken in the Malay heartland of Central Sumatra, the region where Malay is widely believed to have originated. Although there has been extensive contact with other languages (especially Javanese) and with other dialects of Malay/Indonesia, the village dialects neighboring Jambi City have been transmitted in an uninterrupted fashion from generation to generation, as far as we know, back to the earliest speakers of Malay. These village varieties are to be contrasted with contact varieties of Malay/Indonesian like Riau Indonesian, Jakarta Indonesian, and also with the local Jambi koiné spoken in Jambi City, which will play an important role in our discussion later. We will not discuss Jakarta Indonesian here. We limit ourselves to noting that, according to Wouk (1999), Tjung (2006), Sneddon (2006), and many others, contact varieties like Jakarta Indonesian developed through a process of accommodation among a number of sources, among them Betawi Malay, Chinese Malay, Standard Malay/Indonesian, Javanese, and Sundanese, which were the languages/dialects spoken by the major population groups living in the city of Batavia (later Jakarta) at the time Jakarta Indonesian came into existence. Such a mixed origin cannot be posited for the village dialects of Jambi. Although there was unquestionably influence from outside sources like Javanese, the differences among the village dialects clearly instantiate slow divergence over time among neighboring speech communities, similar to what has been observed for dialects of Romance, Germanic, etc. Thus, it is unsurprising if TJM displays grammatical characteristics that are distinct from those found in varieties of colloquial Malay in which language contact played a central role in their creation.

GRAMMAR OF BINDING

The data we will discuss from TJM are drawn from two dialects of Jambi Malay, Tanjung Raden and Mudung Darat (named after the villages in which the dialects are spoken). All varieties of Jambi Malay are mutually comprehensible (including the city dialect that we will discuss later), but there are differences in the grammar of the dialects that are noted and, sometimes, made the topic of jokes by native speakers. (As a whole Jambi Malay is not mutually comprehensible with Standard Indonesian, Jakarta Indonesian, and other dialects not native to the area around Jambi City, but speakers of Jambi Malay typically have at least a passive knowledge of SMI and Jakarta Indonesian from school and the media.) We draw our data from a combination of naturalistic and elicited data, with an emphasis on naturally occurring narratives that were recorded as part of our documentation project.

With this background in mind, we now turn to TJM anaphora. Unlike SMI, in TJM “pronouns” (i.e. the base form of the anaphor, the forms cognate with pronouns in Standard Malay/Indonesian) can corefer with local and LD antecedents, but speakers have strong preferences about what interpretation is most likely. (Preferred interpretations are underlined. MD = Mudung Darat; TR = Tanjung Raden. *Dio* and *jo* are alternative forms of the third person pronoun, related to *dia* and *-nya* in SMI. Unlike SMI *-nya*, Jambi *jo* is an independent pronoun rather than a clitic – at least in MD.)

- (23) a. *dio cinto dio* (MD)
3 love 3
'He loves himself/him.'
- b. *dio cinto samo jo-la* (MD)
3 love with 3-EMPH
'He loves himself/ him.'

GRAMMAR OF BINDING

- (24) a. yanti neŋo? dio di tipi (MD/TR)
 Yanti ACT.see 3 on TV
 ‘Yanti saw herself on TV.’/ ‘Yanti saw him on TV.’
- b. yanti neŋo? dio-la di tipi (MD/TR)
 Yanti ACT.see 3-EMPH on TV
 ‘Yanti saw herself on TV.’/ ‘Yanti saw him on TV.’

(The choice of *dio* versus *no* in (23)-(24) is accidental and does not affect the interpretation.)

As is seen in (23)-(24), the addition of the emphatic focus marker *-la* changes the preferred interpretation but not the range of possible interpretations for the pronouns. When *-la* is present ((23b) and (24b)), the preferred interpretation is that the antecedent for *dio/no* is local. In contrast, when *-la* is absent (as in (23a) and (24a)), the preferred antecedent is non-local, i.e. a discourse antecedent. A similar effect is observed when the intensifier *dewe?* (etymologically related to Javanese *dhewe*) follows the pronoun:

- (25) arna cinto djan dio? dewe? (TR)
 Arna love with 3 INT
 ‘Arna loves herself/(only)him.’
- (26) mama? nik cinto dio dewe? (MD)
 mother Nick love 3 INT
 ‘Nick’s mother loves herself/ (only) him.’

In fact, the primary use of *dewe?* is to signal contrast. Like SMI *sendiri* and Javanese *dhewe*, *dewe?* is often translated as ‘only’ or ‘alone’. In examples like (25)-(26) the effect of *dewe?* is similar to that of *-la* in that it changes the preference for the interpretation of *dio* from

GRAMMAR OF BINDING

‘him’ to ‘himself. *-La*, unlike *dewe?*, is purely an emphatic or focal marker, and its usual translation is not ‘only’ or ‘alone’ but rather as a marker of emphasis.

Dewe? and *-la* can also be used in combination:

- (27) *kami cinto kami dewe?-la* (MD)
1 love 1 INT-EMPH
‘I love only myself.’

The presence of *dewe?* and *-la* both emphasizes the pronoun *kami* and their combination further increases the likelihood that the pronoun will be interpreted as coreferential with its antecedent. (This is the effect regardless of the person features of the pronoun. In the case of first person pronouns, as in (27), no other interpretation is possible.)

Despite the fact that *dewe?* and *-la* make a local coreferential interpretation likely, long distance and discourse interpretations are not ruled out:

- (28) *yanti pikir [arna cinto dio dewe?/ dio-la]* (MD/TR)
Yanti think Arna love 3 INT/ 3-EMPH
‘Yanti thinks Arna loves Arna/Yanti/him.’

Furthermore, the absence of *dewe?* and *-la* does not block the possibility of a “reflexive” interpretation:¹³

¹³ These judgments hold for both TR and MD speakers, but responses from our primary TR speaker are somewhat inconsistent, and seem to vary depending on the choice of predicate and

- (29) yanti pikir [arna cinto dio] (MD)
 Yanti think Arna love 3
 ‘Yanti thinks Arna loves Arna/Yanti/him/her.’

While in isolation the most likely interpretation for *dio* in (29) is ‘Yanti’ or a discourse antecedent, the sentence is well formed to mean ‘Yanti thinks that Arna loves herself.’ Furthermore, *dio dewe?* does not display subject orientation or require a c-commanding antecedent, so it does not display the constellation of properties associated in the literature with “long distance reflexives” (e.g. in Cole, Hermon, & Huang, 2006):

- (30) yanti ba-cakap samo arna tntan dio dewe? (MD)
 Yanti INTR-talk with Arna about 3 INT
 ‘Yanti talked to Arna about Arna/Yanti/him/her.’

Dio dewe? in (30) can refer to the object of the preposition *samo* ‘with’ as well as to the c-commanding subject (*Yanti*). (This contrasts with e.g. Chinese LD reflexives, in which the antecedent nearly always must c-command the reflexive.)

We have shown that both the bare pronouns (*dio/dio?*) and pronoun + *-la* and pronoun + *dewe?* can be used for local, long distance and discourse reference. In addition, these forms can be used regardless of whether or not the antecedent c-commands the anaphoric form, as was shown in (26) (repeated) for *dio dewe?* and by (31) for *diola/nola*.

whether the pronominal form is *dio* or *no*. In contrast, in MD the judgments are completely consistent.

GRAMMAR OF BINDING

- (26) *mama? nik cinto dio dewe?* (MD)
mother Nick love 3 INT
'Nick's mother loves herself/ (only) him.'

- (31) *mama? nik cinto ana?-e, laki-e jugo djan dio-la/ po-la* (MD)
mother Nick love child-3 husband-3 also with 3-EMPH/3-EMPH
'Nick's mother loves her child, her husband and also herself/him.'

In both (26) and (31) the antecedent for the anaphoric form can be the c-commanding nominal (*mama? nik*) or the non-c-commanding possessor (*nik*).

To summarize, both the bare pronouns (*dio/dio?*) and pronoun + *-la* and pronoun + *dewe?* are underspecified anaphoric forms like *awake dheen* in Peranakan Javanese. They are vague in their referential possibilities and can be used for local, long distance coreference and Binding, as well as for discourse coreference, whenever the context is right. That is, TJM does not divide anaphora into two types (pronouns and reflexives) as in English, into three types (pronouns, reflexives and "long distance reflexives") as in Chinese, or into three (different) types (pronouns, reflexives and underspecified anaphors), as in Peranakan Javanese and Standard Malay/Indonesian. Rather, **all** anaphors are unspecified with respect to local/long distance, c-command, etc. There is, however, a clear distinction among forms with regard to which interpretation is preferred. The preferred interpretation for the bare forms corresponds to that of Binding Theoretic pronouns in a language like English, while the preferred interpretation for the complex forms (pronoun + *-la* and pronoun + *dewe?*) corresponds to that for Binding Theoretic reflexives. We emphasize, however, that these are interpretative preferences and not grammatical requirements. All interpretations are possible for all forms.

5. More on Intensifiers and Emphatics in Traditional Jambi Malay

We have seen that intensifiers and emphatics play a critical role in obtaining a “reflexive” interpretation in TJM, but that they do not establish grammatical requirements. In order to clarify the role of intensifiers and emphatics in anaphora, we will next examine the broader role of intensifiers and emphatics in the grammar of the language.

As was mentioned previously, *dewe?* is cognate with and was probably borrowed from Javanese *dhewe*, and is used as an intensifier akin to Javanese *dhewe* and Madurese *dibi*. It expresses many of the same meanings described for *sendiri* by Sneddon (1996) (Standard Indonesian) and Gil (2001) (Riau Indonesian).

(32) Riau Indonesian	‘deagentive’
<p>Bisa di-buka <i>sendiri</i> pintu-nya.
can AT-open sendiri door-ASSOC
‘Its doors can open by themselves.’</p>	
(Gil, 2001, p. 15)	
(33) Riau Indonesian	‘dealiative’
<p>Aku <i>sendiri</i> tak ada rumah.
1:SG sendiri NEG exist house
‘Only I don’t have a house.’</p>	
(Gil, 2001, p. 16)	
(34) Riau Indonesian	‘intensive’
<p>Mister <i>sendiri</i> yang kasi
white.person sendiri REL give
‘You yourself gave them to me’</p>	
(Gil, 2001, p. 22)	

GRAMMAR OF BINDING

In the examples that follow, *dewe?* modifies a referring expression (rather than an anaphor) or has an adverbial function (modifying a predicate) [examples are taken from Cole et al., 2010:337]:

- (35) *dolɔ* *aku* *da?do* *me?* *jeliya* *dewe?-la*
sibling 1SG NEG.exist aunt Jeliya INT-EMPH
'I don't have sisters or brothers, it's only aunt Jeliya.'
['I don't have sisters or brothers except for aunt Jeliya.']

- (36) *sila* *makɪn* *dewe?* *makɪn* *mapa?* *umo-e*
Sila more INT more a.lot paddy.field-3
'The more Sila works alone, the more paddy field she has.'

- (37) *motonɔ* *sayo* *dewe?-la*
ACT.cut 1SG INT-EMPH
'I myself cut it.'

In (35) *dewe?* modifies *me? jeliya* and indicates exclusivity ('only X', 'Aunt Jeliya and no other'). In contrast, in (36) *dewe?* modifies *makɪn*, i.e. the predicate, and indicates 'alone'. In (37) *dewe?* can be understood as predicate modification, 'I cut it alone.' or as modifying *sayo* 'I myself', 'I and no other'. In all the examples *dewe?* indicates exclusivity or contrast, either with respect to the arguments or with respect to the predicate. Thus, the general effect of *dewe?* is contrastive focus, 'X as distinct from/in contrast with Y', where X can be an argument or a predicate.

Some further examples, also taken from our naturalistic corpus, illustrating the emphatic use of *-la*, *dewe?* and *dewe?la* are the following. In these examples *-la*, *dewe?* and *dewe?la* occur in conjunction with *dio?*, but the interpretation is not that of a reflexive anaphor.

GRAMMAR OF BINDING

- (38) Suharto-tu militer dioʔ-la
 Soeharto-that military 3-EMPH
 ‘Soeharto was his own soldier.’ [Speakers is explaining how Soeharto succeeded Soekarno as President of Indonesia. *Dioʔla* refers to the discourse antecedent, Soekarno, not to Soeharto.]
- (39) mntla-e dioʔ-la kan
 before-3 3-EMPH PART
 ‘He used to do it by himself. [‘Dio’ refers to the father of the individual under discussion.]
- (40) biaʔ dioʔ makan papon dioʔ makan dioʔ deweʔ-la k’dalam
 although 3 eat thing 3 eat 3 INT-EMPH to.inside
 ‘Even though he ate many things he ate by himself inside.’
- (41) teŋoʔ dioʔ deweʔ ŋan makan papoan daʔ oaŋ bakuŋ ŋbut
 look 3 INT very eat thing NEG person Bakung N.mention
 ‘See, he himself doesn’t eat many things, said Bakung people.
 [Bakung people said that he himself just eats whatever he finds.]

None of the uses of pronoun + *deweʔ* or *deweʔla* in (38) – (41) have a reflexive interpretation.

The distribution of *deweʔ* and *-la*, whether alone or in combination, appears to be determined by their core meanings as intensifiers and emphatics. There is no sign that either of these morphemes has been reanalyzed as part of a dedicated anaphor. What we would like to emphasize is that while *deweʔ* and *-la* do not convert bare anaphoric forms into dedicated reflexives, they do affect the ease with which a “reflexive” interpretation can be assigned to an anaphoric form. Bare anaphoric forms are likely to be interpreted similarly to the pronominals that occur in languages that have a distinct pronominal category, while anaphor plus *deweʔ* or *-la* is likely to be interpreted similarly to a dedicated reflexive. These, however, are interpretive

tendencies, varying from individual to individual and from situation to situation, and not grammatical rules. All forms can be used for the full range of interpretations since intensification and emphasis only increase the likelihood of a “reflexive” interpretation. Such an interpretation is in no sense forced by the use of these forms.

6. Jambi City Malay

We have looked at the anaphoric system found in TJM. We will turn now to Jambi City Malay (JCM), and will show that in JCM the system is nearly identical to that in TJM, with one important exception. What constitutes an interpretative tendency in TJM is an obligatory grammatical rule in JCM.

We turn first to bare pronominal forms. In the examples that follow, we see that bare forms must be interpreted as Binding Theoretic pronominals:

(42) a. dio cinto dio
 3 love 3
 ‘He loves him.’

b. yanti neŋo? dio di tipi
 Yanti ACT.see 3 LOC TV
 ‘Yanti saw her on TV.’

(43) mama? nik cinto dio?
 mother Nick love 3
 ‘Nick’s mother loves him.’

The examples in (42)-(43) are nearly identical to those we have seen from the village dialects, but, unlike TJM, these and similar examples **cannot** be interpreted as involving coreference

GRAMMAR OF BINDING

between the subject and the object. Rather, disjoint reference is obligatory. For the same reason, in (43), *dio?* can only refer to the non-c-commanding possessor, *Nik*, and not to the full c-commanding subject, *mama? nik* ‘Nick’s mother’. Thus, the bare pronominal forms like *dio?* display all the usual characteristics of Binding Theoretic pronouns.

We turn next to pronoun + intensifier (*dio? dewe?*). (Pronoun + emphatic *dio?la* does not occur in JCM in reflexive contexts although it is understood by JCM speakers.) The “reflexive” interpretation for *dio? dewe?*, which is the preferred but not the sole interpretation in TJM, is the only interpretation available in JCM. A reflexive interpretation is also the only interpretation available when *diri-ɲo dewe?* is used (modeled on the SI reflexive *dirinya sendiri*):

- (44) a. *mama? nik cinto dio? dewe?*
 mother Nick love 3 alone
 ‘Nick’s mother loves herself.’
- b. *mama? nik cinto diri-ɲo dewe?*
 mother Nick love self-3 alone
 ‘Nick’s mother loves herself.’

In (44a-b) the only possible referent available for *dio? dewe?* and *diri-ɲo dewe?* is *mama? nik*.

Thus, with these forms the antecedent must c-command the anaphor (*Nik* is only part of the subject noun phrase, so it cannot c-command the object) and discourse referents are also not possible, so we see that *dio? dewe?* and *diri-ɲo dewe?* cannot be interpreted as “free anaphors”. That is, as in the case of the bare pronominal forms, the readings available in JCM are exactly those readings demanded by the classical Binding Theory. In this case, the readings are those expected for Binding Theoretic reflexives.

GRAMMAR OF BINDING

To summarize the situation, from the perspective of classic Binding Theory, the binding systems of TJM and JCM could not be more different. In JCM the classic Binding Theory holds, at least with respect to *dio?*, *dio? dewe?* and *diri-no dewe?*. Thus, binding in JCM is determined solely by presumptive principles of UG controlling the possibility of forms that are [+anaphoric] and [+pronominal]. In contrast, in TJM, all forms are Binding Theory exempt anaphors. Exempt anaphors are by definition outside the scope of the Binding Theory because, as we saw earlier, they lack the features [+anaphoric] and [+pronominal] altogether. Whatever the principles are that control preferences among exempt anaphors, they would have to be entirely separate from the subsystem of UG that licenses Binding Theoretic pronouns and bound anaphors. Thus, any connection between the binding systems of TJM and JCM would have to be viewed as accidental, i.e. due to the happenstance of historical change.

From the perspective of the speakers, however, the same anaphoric system appears to hold in the village and city dialects. In both, it is the forms with intensifiers and/or emphatics that tend to get reflexive interpretations, and those employing bare pronominal forms that are interpreted as disjoint in reference. From this point of view, there is only a single, “minute” difference separating the village and city dialects. What constitutes a preference in the village dialects is obligatory in the city. Thus, from the viewpoint of the speakers themselves, there is no perceptible difference between the city and village dialects in their use of forms like *dio?*, *dio? dewe?* and *diri-no dewe?*. The fact that city speakers display greater consistency in their use of these forms would be noticed only through text counts or by asking for grammaticality judgments, neither of which occur in normal interaction among speakers.

7. Explaining the Preferences in TJM

Any analysis that treats the distribution of anaphors in TJM and JCM as controlled by unrelated modules of grammar is less than optimal because it makes the implicit claim that the great similarities among these closely related dialects are accidental, i.e. the result of historical change rather than of universal grammar.¹⁴ Thus, it is necessary for us to explore how a unified

¹⁴ A defender of Classical Binding Theory might argue that it is in fact desirable that TJM and TCM would, at the “underlying” level, receive two entirely unrelated analyses while appearing on the surface to be nearly identical. We would argue, however, that such a move would be self-defeating. The strongest argument for a universal Binding Theory was poverty of the stimulus, the claim that the complexities of Binding are unlearnable on the basis of the evidence available to the child, so they must be “hard-wired” rather than learned. Note, however, that the facts of TJM and TCM are identical except with respect to the obligatoriness of the distribution of anaphoric forms. This means that the speakers of TJM must learn the full complexities of these facts, by hypothesis, without the assistance of (this aspect of) UG. Such a scenario, however, severely undercuts the argument for specifying Binding Theory in UG since the distribution specified found in TJM is the same as that in TCM (obligatoriness aside). Since by hypothesis, the facts of TJM are learned without the benefit of Binding Theory, and since these facts are the same as those in TCM, it follows that the facts of TCM can be learned without the benefit of the Binding Theory. Thus, an analysis that treats the facts of TJM as learnable from experience, would also predict that those of TCM are learnable. However, if that is the case, these facts provide no basis for a poverty of the stimulus argument in favor of a UG-based Binding component.

analysis of the cross-dialectal data might be possible. As a first step toward this goal, we need to have an analysis of how the preferences found in TJM should be accounted for. We will then ask in what way the analysis can be extended to JCM.

As we discussed earlier, we make the assumption that the principles of UG set the parameters distinguishing between possible and impossible grammars, but UG cannot explain why one of two possible grammatical forms is preferred. Rather, the explanation for preferences is not found in UG, but must be sought elsewhere, e.g. in the cognitive modules that interface with language, and which govern the various aspects of language use. That is to say, preferences are not determined by UG, but instead, to use Minimalist terms, by the interface between the computational component that is determined by UG, and the conceptual-intentional system. As a result, our first step will be to look for an explanation in the conceptual-intentional interface for why in TJM pronoun + intensifier and pronoun + emphatic are preferentially interpreted as coreferential with the subject of the clause, while bare object pronouns are preferentially interpreted as non-coreferential. Accounts for how intensifiers could lead to reflexive interpretations are provided in the work of a number of scholars, among them König and Siemund (2000), Paul Kiparsky (2002), Edward Keenan (2002), and König and Gast (2006). We shall base our discussion primarily on ideas proposed in a number of papers by König and colleagues, in particular on König and Siemund (2000), where the development of modern English reflexives from Old English pronoun + intensifier is examined in detail.¹⁵ We take their

¹⁵ A variation on this analysis is given in König and Gast (2006), who analyze adnominal intensifiers as denoting an identity function ID, which takes a referent and maps it onto an output which is identical to the input. This semantically trivial operation produces a pragmatic effect

analysis of Old English as a model for our treatment of TJM. König and Siemund consider the interpretation of examples like:

through focusing. As a result of focusing, a set of alternatives is evoked. The alternatives are different from but must be pragmatically related to the input in order to be evoked by the input. This analysis makes essentially the same predictions as that of König and Siemund (2000).

The discussion in Keenan (2002) is also especially enlightening. Keenan shows that a number of steps were required in the transition from the OE anaphoric system to that in ME, in particular the fact that dedicated reflexives came into the language before dedicated pronouns came to exist. While König and Siemund's discussion spans a broad set of topics, they give special attention to the point of interest for us here, how an intensifier might affect the likelihood that an object pronoun will be treated as coreferential with the subject of the clause. Another notable work dealing with this topic is Levinson (2000), who presents an analysis that is similar in many ways to that of König and Siemund (2000), but which builds on the idea that there is a general presumption in language that the arguments of a predicate are not coreferential. This is combined with a principle that marked forms like pronoun + intensifier implicate the complement of unmarked forms like bare pronouns. Since unmarked forms (bare pronouns) implicate disjoint reference of coarguments, marked forms (pronoun + intensifier) will implicate coreference.

GRAMMAR OF BINDING

(45) The director himself will talk to us.

In ... the examples given above, alternative persons are brought into the discussion by the intensifier; the director as opposed to other possible speakers, the driver as opposed to the passengers, etc. Relating a given value to a set of salient alternative values is a characteristic property of focusing devices, to which intensifiers therefore clearly belong. There is, however, yet another property that intensifiers manifest in the examples under discussion: not only do they evoke alternatives, they also characterize the alternatives in some way. Possible alternatives in an example like [(44)] are people inferior to the director in the relevant professional context, i.e. his secretary, his substitute, his collaborators, etc. Generalizing from this example, we can say that intensifiers relate a periphery of alternative values (Y) to a center (X), identified by the expression they combine with...

Figure 1. Center and Periphery

(König and Siemund, 2000, p. 42)

Thus, in (45) *the director* is characterized as the center. The intensifier calls up a set of alternatives to the center, which are characterized as being in some way related to the center. The effect of the intensifier, then, is contrastive focus, one of establishing the intensified nominal as “central” to the event described and of calling to mind alternatives to the central participant.

Let us apply König and Siemund’s analysis to TJM. In König and Siemund’s terms, in (25) (repeated below) the center is the subject, *Arna*.

- (25) arna cinto djan dio? dewe? (TR)
 Arna love with 3 INT
 ‘Arna loves herself/(only) him.’

This is because the verb *cinto* ‘love’ is a prototypically outward acting, non-reflexive predicate.

In terms of theta roles, the expected theme is not the same individual as the actor. (Contrast ‘love’ with a prototypically reflexive predicate like ‘dress’, for which the theme is expected to be identical to the actor—though this expectation can be discharged.) For *cinto* ‘love’ the actor is understood to be the center and the theme the periphery.

The presence of the intensifier *dewe?* has two effects. The first is contrastive focus, to call up a set of alternatives to *dio?*. This alone, however, would not explain the propensity to interpret *dio? dewe?* as referring to *Arna*. The second, and for us, more important effect is to establish *dio?* as the center (the individual with whom the periphery is contrasted). The indication that *dio?* is the center that is provided by the intensifier creates a potential conflict in the interpretation of the sentence. Because *cinto* ‘love’ is a prototypically outward acting predicate, the subject is taken to be the center, and the theme the periphery, but the occurrence of *dewe?* indicates that it is the theme that is the center. This conflict in presumptions can be resolved without contradiction if the theme and the actor are understood to be the same individual. Thus, the interpretation that is most easily compatible with understanding both the actor and the theme as the center is to interpret the theme as the same individual as the actor. If it is assumed that the emphatic particle *-la* in sentences like (23b) is also a focus particle, then *-la* also establishes the nominal to which it is attached (*no*) as central. Thus, the analysis for *dewe?* extends to sentences with the emphatic *-la*:

(23b) dio cinto samo jo-la (MD)
3 love with 3-EMPH
'He loves himself/ him.'

Finally, turning to bare pronominal forms (*dio?* and *no*), when they are objects of outward acting predicates like *cinto* they also would be expected not to be central:

(23a) dio cinto dio (MD)
 3 love 3
 ‘He loves himself/him.’

The presumption that bare *dioʔ/no* are peripheral implies that they are not coreferential with the actor. This is because, unlike sentences with *dioʔ deweʔ* and *dioʔla*, in the absence of an intensifier or an emphatic particle, the presumption that the theme is peripheral is left unchallenged. If the theme is not central, the theme cannot be the same individual as the actor, because the actor **is** assumed to be central. Thus, the actor and the theme are likely to be different individuals. We emphasize, however, that the presumptions that the subject and theme are or are not the same individual are dischargeable inferences, and are subject to correction when necessary, as we saw with regard to *dioʔ deweʔ* and *dioʔla*.

We shall assume that König and Siemund's analysis for Old English provides an appropriate analysis for TJM. We would like to emphasize that a theory like that of König and Siemund is a (partial) theory of the conceptual-intentional interface rather than a theory of UG

itself. The fact that predicates like ‘love’ are prototypically outward acting and that the actor is interpreted as the center is based on our understanding of what ‘love’ is in the world. It does not derive from specifically linguistic properties of the verb *cinto* ‘love’, nor from properties of language as such, nor, indeed, from UG. Thus, like pragmatic knowledge in general, König and Siemund’s account of pronoun + intensifier is an account of (aspects of) how language interfaces with other faculties and our knowledge of the world, rather than an account of the internal properties of language itself as an autonomous system.

8. Grammaticalization

We have argued that König and Siemund’s functional account of the interaction between intensifiers/emphatics and pronouns provides an explanation for the distribution of these forms in TJM. We will now turn from TJM to JCM. König and Siemund’s theory of the interpretation of pronoun + intensifier would not correctly predict the facts in JCM because the preferences found in TJM are based on dischargeable inferences, but the distribution of pronoun + intensifier is obligatory in JCM. The pattern in JCM can also not be attributed to UG because then no explanation would be provided for the near identity of the patterns in the two dialects.

These difficulties can be overcome if we give up the claim that language specific grammatical rules are epiphenomenal, and instead adopt the traditional position, abandoned by most contemporary Generative linguists, that languages have systems of grammatical rules that are specific to that language. From this perspective, the source of grammatical rules in language is not restricted to the interaction among principles of universal grammar and properties of the vocabulary of the language, although that remains a possible source of grammatical rules. Syntax (and not just the lexicon) is viewed as partially determined “biologically” (determined by

GRAMMAR OF BINDING

the human language faculty, i.e. by UG) and partially “culturally” (by grammatical rules that are learned by speakers in the process of acquiring their language). In the view of language we are adopting, specific grammatical rules can and typically do arise through a process of “grammaticalization” that transforms extra-grammatical tendencies and preferences due to properties of the interface into rules of grammar that are integral parts of the linguistic system itself and which are independent of their diachronic extra-grammatical source. We take the grammatical rules derived via grammaticalization to constitute an aspect of the systematic knowledge of the language that is learnable (in Chomskian terms, learnable by the Language Acquisition Device), and which is passed on from the current generation to the succeeding generation of speakers. We, furthermore, assume that synchronic grammars do not partition whatever rules can be directly attributed to UG (if any) from those peculiar to the language. Thus, once a grammar has been acquired by speakers, no easy diagnostic is available regarding which aspects of the grammar are directly due to UG and which are language specific, nor is it transparent how language specific grammar and universal grammar interact. Under this view, a major aspect of empirical linguistic research is concerned with deconstructing the nature of the interaction between biological and cultural elements in grammars. We shall examine the plausibility of this approach with respect to the relationship between TJM and JCM. We will then discuss some implications and difficulties with this approach.

Returning to Jambi Malay, it will be remembered that the critical difference between the village dialects and the city dialect is that, in the city dialect (as in other languages that appear to obey the classic Binding Theory), forms consisting of a pronominal plus an intensifier are obligatorily, not preferentially, coreferential with (or bound by) a local antecedent. Studies of the historical development of reflexives like Keenan (2002), Kiparsky (2002), König and

GRAMMAR OF BINDING

Siemund (2000) and Haspelmath's frequentist analysis (2008) have shown that similar developments have occurred in the differentiation of reflexives and pronouns in many languages. From a cross-linguistic perspective, the direction of development appears to be consistently from a functionally based system resembling that of TJM to a more rigid system in which pronoun + intensifier is obligatorily treated as a Binding Theoretic reflexive. Thus, the relationship between TJM and JCM appears to be one of what Lehmann (1982/1995) refers to as "obligatorification", a subtype of grammaticalization, which he defines as a unidirectional process in which linguistic structures become increasingly more obligatory in the process of grammaticalization.¹⁶

Grammaticalization explains the development of Binding Theoretic reflexives from intensifiers in Jambi Malay and other languages by positing a general tendency for discourse/pragmatic patterns to be bleached of their discourse/pragmatic content and to be reanalyzed as formal (hence, grammatical) patterns. The occurrence of grammaticalization is well described in a wide variety of languages, but the question persists as to why such a process might exist. We would argue that a plausible explanation, at least in this case, is that obligatorification facilitates processing by reducing the attentional demands on the speaker/hearer. The primary result of converting a discourse distinction to grammar is an increase in automaticity since by definition grammatical requirements apply automatically in specified environments without the need to consider the multitude of factors affecting speaker intent. The advantage of greater automaticity is a decreased need for attention on the part of the

¹⁶ See Campbell and Janda (2001) for discussion of problems related to "grammaticalization theory".

speaker and hearer. That is, we hypothesize that the grammaticalizing of a distinction reduces the processing load on the speaker and hearer since once a construction has been grammaticalized it is not necessary to attend to certain aspects of the intentions of the participants in the discourse. These aspects of speaker intention would require attention in the absence of grammaticalization. In the case of JCM, the grammaticalization of *dio? dewe?* as a Binding Theoretic reflexive allows this phrase to be understood as referring to the same individual as the subject without the need to go through the complex deduction involved in recognizing that the only way *dio?* could be both the object of *cinto* ‘love’, and hence periphery, and at the same time the center, as entailed by its relationship with *dewe?*, is if *dio?* is the same individual as the actor of ‘love’. The need for this complex chain of deductions is eliminated if the conclusion of the deductions—that *dio? dewe?* is the same individual as the subject—is directly associated with the expression *dio? dewe?*. We take it to be a general principle of the conceptual-intentional interface that it is desirable to minimize the processing load imposed by sentences, in this case by the grammaticalization of pronoun + intensifier as a reflexive pronoun. Thus, there is a general tendency for discourse preferences to be converted to grammatical rules.

The diversity of Malayic dialects found in the wider region surrounding Jambi City makes it possible to test whether Traditional Malayic dialects fit the pattern that would be expected if the system of anaphora found in JCM is due to the grammaticalization. As mentioned previously, Keenan, and König and Siemund both point out with respect to the development of English reflexives that pronoun + intensifier became an unambiguous reflexive long before bare pronouns ceased to allow a reflexive interpretation. This sequence is expected because the grammaticalization of pronoun + intensifier creates a system in which co-argument coreferentiality can be expressed in two ways, by pronoun + intensifier or by a bare pronoun. On

GRAMMAR OF BINDING

the assumption that separate forms will over time acquire separate functions, it would be expected that the system would have a tendency to readjust to one in which the overlap between pronoun + intensifier and bare pronouns would be reduced or eliminated. There would, however, be a period in which both the bare form and the pronoun + intensifier could be used for coreferential co-arguments. These stages are illustrated in (46):

(46) Stages in grammaticalization of anaphora

- Stage 1: bare pronouns are undifferentiated anaphors (local coreference and non-coreference possible); likelihood of Binding Theoretic reflexive interpretation higher when intensifier present
- Stage 2: anaphor + intensifier interpreted as Binding Theoretic reflexive but simple bare pronoun can be interpreted as either “reflexive” or “pronoun” (i.e. as an “exempt anaphor”)
- Stage 3: bare pronoun is interpreted only as a Binding Theoretic pronoun

Returning to the Malayic dialects of the Jambi region, on the basis of a grammaticalization account of the facts of Jambi Malay like (46), it would be expected that dialects might occur that illustrate Stage 2, the stage at which Binding Theoretic reflexives exist but Binding Theoretic pronouns do not. A dialect with Binding Theoretic pronouns but not Binding Theoretic reflexives would be expected not to be found in any dialect. The predictions of (46) appear to be correct. In the Malay of the rural area surrounding Sarolangun (Jernih)—several hours drive from Jambi City—the rural varieties of Malay exhibit the pattern predicted in Stage 2 of (46):

GRAMMAR OF BINDING

(47) Sarolangun (Jernih) Malay

- Pronoun + Intensifier (and *diri* + intensifier) in unambiguously interpreted as a reflexive.
- Anaphoric forms **without** the intensifiers are still ambiguous and thus are underspecified.

(48) Bare pronoun interpretable as Binding Theoretic pronoun or reflexive

siti nunton po dalam tipi
Siti N.watch 3SG inside television
'Siti watched herself/him on the television.'

(49) Pronoun + intensifier only interpreted as Binding Theoretic reflexive

bude? tu paket po diwi?
kid that N.sick 3SG INT
'That kid hurt himself/*him.'

(50) Diri + pronoun + intensifier only interpreted as Binding Theoretic reflexive

bude? tu paket diri po diwi?
kid that N.sick self 3SG INT
'That kid hurt himself/*him.'

(51) Diri + intensifier (no pronoun) only interpreted as Binding Theoretic reflexive

adi? man paket diri diwi?
younger.sibling Man N.sick self INT
'Man's younger brother hurts himself/*him.'

According to our analysis, it is not accidental that dialects are found in which dedicated reflexives exist but dedicated pronouns do not occur, but that no dialects are found in which

dedicated pronouns occur but not reflexives. This pattern is not predicted by a theory that takes the existence of pronouns and reflexives to be primitives provided by UG. Thus, we conclude that the distributional patterns we have encountered in traditional Malayic dialects of Jambi Province conform to the predictions of a grammaticalization analysis of anaphora rather than to one that treats the categories of pronoun and reflexive as primitives of Universal Grammar.

9. The *De Novo* Creation of Language: Creole and Signed Languages

The grammaticalization analysis just presented makes additional predictions that appear to solve previously unsolved puzzles. Grammaticalization explains the fact that the grammatical distinction between reflexives and pronouns occurs in very many (and perhaps a large majority) of the world's languages, while at the same time, the distinction cannot be a primitive of UG, for the reasons we have discussed. The problem for linguistic theory that has motivated this study has been why it should be rare for languages not to make a distinction between reflexive and non-reflexive pronouns yet for that state of affairs still to be possible.¹⁷ If, as we have suggested, it is true that the grammaticalization of pronoun + intensifier as a reflexive reduces the

¹⁷ Surprisingly, there has never been a large scale typological study of the distribution of anaphoric systems in the languages of the world that would answer the question of which languages distinguish between reflexive and non-reflexive pronouns. Levinson (2000) states, "There are scattered reports from around the world of languages that do not directly code reflexivity." This would suggest that Levinson believes that such languages are few in number. This is further supported by the fact that Levinson provides a short discussion that appears to list all the languages known to him that fail to distinguish between reflexives and pronouns.

attentional demands on speakers and hearers (a claim that should be investigated using psycholinguistic and neurolinguistic methodologies), it is not surprising that stages in which these forms are not grammaticalized are typically short-lived and, hence, rare in languages. The grammaticalization analysis predicts, however, that they should be detectable as occurring, in appropriate circumstances. We shall discuss briefly some such circumstances. One situation in which undifferentiated anaphoric forms might be expected to be observed frequently is when languages are formed *de novo*. The *de novo* formation of languages is a very infrequent event, but there are circumstances in which it is claimed to occur (or is at least to be approximated). A test case that comes to mind immediately is the early stages of creole languages. While it appears to be an exaggeration that the grammar of the substrate language is completely obliterated in creoles, it does seem to be correct that radical simplification occurs and much of the substrate grammar is eliminated (the so-called creole bottleneck). As a result, creoles as a group are among the simplest languages.

It has, thus, been argued that creoles should be viewed as reflecting the requirements of UG more directly than do languages with thousands of years of change induced complexity (a proposal suggested by Bickerton, 1981; 1983; 1984; 1988; and 1991). To the extent that creole languages can be viewed as creations *de novo*, or at least as radical simplification to a linguistic minimum, they constitute environments in which the grammaticalization hypothesis would lead to the expectation that at least in the early stages of the creole it is very likely that there will be no distinction between pronouns and reflexives. In contrast, the hypothesis that the distinction between reflexives and pronouns is stipulated by UG (and perhaps obscured by language change) would predict that a distinction between pronouns and reflexives would occur just as frequently (perhaps more frequently) in creole languages as in non-creoles. These conflicting predictions

are discussed by Levinson (2000), who notes that it was his impression that up to half the pidgins and creoles do not distinguish between pronouns and reflexives, and “creoles that do seem to have anaphors also seem to have passed through [...] a stage with no anaphors...” If this description of the typology of reflexives in creole languages is factually correct, the facts would appear to follow from a theory predicting the existence of a stage in which there are no Binding Theoretic reflexives or pronouns in the genesis of a new language. That stage, however, may be short-lived, presumably for processing reasons.

An empirical test of the distribution of reflexives in creole languages can be carried out by examining the types of reflexives occurring in the creole languages examined by the Max Plank Institute for Evolutionary Anthropology project entitled “The Atlas of Pidgin and Creole Language Structures (APiCS),” Michaelis, Maurer, Haspelmath, and Huber (2013). According to Chapter 87, “Reflexive Constructions,” by Martin Haspelmath and the APiCS consortium, of the 76 creole languages examined, in 18 of the languages an ordinary anaphoric pronoun is used for reflexives. This number considerably exceeds our expectations for languages in general (although our expectations are based on the personal experience of the authors of this paper rather than on a formal survey). The results seem to confirm Levinson’s general impression that the use of an ordinary pronoun for a reflexive is over-represented in creoles. It is of interest, however, that the percentage of languages found in the APiCS sample in which reflexives are expressed by an ordinary anaphoric pronoun is considerably less than half the pidgins and creoles. In fact, just under a quarter of the creole languages included in the APiCS survey made use of ordinary pronouns to express reflexive coreference. While this number does confirm that many more creoles use ordinary pronouns than do languages in general, the question remains why a much higher percentage of the creole languages do not use ordinary pronouns for reflexive

purposes. An answer that comes to mind is that the processing demands on the speaker in languages using ordinary pronouns as reflexives are sufficiently great as to lead to a rapid differentiation between pronouns and reflexives in a majority of languages. This hypothesis can be tested empirically by examining earlier stages of creole languages that now distinguish between reflexives and pronouns. The expectation is that the distinction between pronouns and reflexives will not be present in the earliest stages of many of these languages.

Similar predictions would be made for other sorts of newly created languages, among them newly created sign languages like Nicaraguan Sign Language, now referred to as Idioma de Señas de Nicaragua (hereafter, ISN; see Kegl (2008) and (in progress)), and Russian Sign Language (RSL, Kimmelman 2011). In both ISN and RSL ordinary pronouns can be used in contexts in which a reflexive would be expected in European languages. This state of affairs is expected if the occurrence of reflexives is due to diachronic processes like grammaticalization: In the earliest stages of a new language, it is predicted that no dedicated reflexive will exist. At a somewhat later stage, dedicated reflexives will be introduced, often as a result of the grammaticalization of emphatics that also have non-reflexive uses. Pronouns, however, will continue to be unspecified with regard to the distinction between reflexives and pronouns. That is, our analysis predicts 1) that new languages will not grammaticalize the distinction between pronouns and reflexives initially; 2) that the direction of change will go from pragmatic to grammatical and not the other way around; and 3) that reflexives will grammaticalize before pronouns and not the reverse. These predictions appear to be confirmed by the data from sign languages.

10. Alternative Binding Theories

In this section we will discuss briefly whether a different and more up-to-date theory of binding than the Classical Binding Theory might make better predictions than those of the Classical Binding Theory. The main alternative to the Binding Theory within Generative Grammar is the binding theory of Reinhart and Reuland (1993), the so-called Reflexivity analysis. According to Reinhart and Reuland, the Binding theoretical requirements of UG constitute restrictions on the marking of reflexive predicates (defined as predicates with two co-indexed arguments) rather than restrictions on coreference among arguments. According to the Reflexivity analysis, the following conditions must hold:

- (52) a. A reflexive-marked predicate is reflexive.
- b. A reflexive predicate is reflexive-marked.

- (53) a. A predicate is reflexive iff two of its arguments are coindexed.
- b. A predicate (P) is reflexive-marked iff either P is lexically reflexive or one of P's arguments is a SELF-anaphor [thereby reflexive-marking P].

Simplifying somewhat, the Reflexivity analysis makes the claim that reflexive predicates will be specially marked as such by the presence of a SELF-anaphor (53b), and that any predicate marked by the presence of a SELF-anaphor will be a reflexive predicate (53a). (The core of the theory is restricted to co-argument reflexives, and, hence, does not cover coreference between an argument of a predicate and a non-argument in the same clause, e.g. *John bought the cake for himself*, where *himself* is not an argument of *buy*.) Note that typically reflexive

GRAMMAR OF BINDING

predicates like verbs of grooming do not require a SELF-anaphor, according to Reinhart and Reuland's version of the Binding Theory. This is because verbs of grooming and other inherently reflexive predicates are considered to be lexically reflexive marked.

We will not explore here the many difference in the predictions of the Classical Binding Theory and the Reflexivity theory, but will restrict our attention to the predictions of the Reflexivity analysis about the occurrence of languages in which ordinary pronouns are used as reflexives. Such examples appear to violate the Reflexivity analysis to the same extent that they violate the Classical Binding Theory since they involve reflexive predicates that by hypothesis are not reflexive-marked. In a recent MA thesis and a related conference paper from Utrecht University, Franssen (2010) and Everaert and Franssen (2012) have suggested that the occurrence of ordinary pronouns as reflexives in Austronesian languages may have an explanation compatible with the Reflexivity theory of binding. As we just noted, according to the Reflexivity theory, a reflexive predicate does not need to be reflexive marked if the predicate is lexically reflexive, e.g. verbs of grooming etc. The purpose of Franssen's MA thesis and Everaert and Franssen's ICAL paper is to examine whether the use of ordinary pronouns as reflexives in Austronesian can be explained on this basis. (The thesis also deals with reflexives in Australian languages. We will not discuss that aspect of the thesis here.) Our discussion is based on Franssen (2010) since we only have an abstract and a handout for Everaert and Franssen (2012).

Franssen's results are stated to be preliminary. Because the thesis is based entirely on published grammars, the results are influenced by the choice of published sources on the languages considered, the topics that were of interest to the authors of the grammars and the extent to which the categorization made by the grammar writers is correct. This can seriously

bias the results of the study. For instance, we would question whether, in SMI, Franssen's sources are correct in categorizing *diri* as a pronoun in examples like the following:

- (54) Siti bunuh diri.
 Siti kill self
 'Siti killed herself.'

Evidence is presented in e.g. Sneddon (1996) and Gil (2001) that the distribution of *diri* used without a following pronoun is different from that of [*diri* + pronoun] and [*diri* + pronoun + *sendiri*]. The correct analysis of *diri* alone would seem to be roughly as an adverb (conjunctive operator for Gil) that, in appropriate circumstances, has (roughly) the effect of converting the verb that it modifies to a reflexive predicate. See Gil (2001) for details. In contrast to V + *diri*, the nominal anaphoric system in SMI consists of a Binding Theory compatible pronoun (*dia*, *ia*, or *-nya*, in the third person (*ia* is restricted to subjects; *-nya* is a clitic), and, in partially overlapping syntactic environments, a Binding Theory Compatible reflexive nominal (*diri* + pronoun + *sendiri*) and a Binding Theory exempt anaphor, *diri* + pronoun. This anaphoric system has essentially the same characteristics as the system occurring in Peranakan Javanese *awake dheen*, discussed earlier in the paper. The anaphoric system of SMI is described in detail in Cole and Hermon (1998, 2005). Thus, due to his choices with respect to the sources he employs, the picture of SMI provided by Franssen is, in our view, inaccurate in ways that directly affect the hypotheses he is examining. In the case of SMI, this problem could be corrected if Franssen had employed different but widely available published sources, but for many of the languages that are counted the literature is much more sparse than is the case for SMI. Thus, gaps in the descriptions given in the source grammars could, and, in our view,

probably did, affect the results materially. For these reasons, we do not believe that the results of Franssen's survey should be taken as providing support for the Reflexivity account of Binding.

Despite the problems with Franssen's survey data, the hypotheses he proposes are directly relevant to the issues we discuss here, and Reflexivity could, if the predictions it makes hold up to examination, provide an alternative account of the facts discussed in earlier sections. The Franssen/Everaert approach would predict that a close examination of e.g. Peranakan Javanese and of Jambi Malay would show that the use of Binding Theory exempt anaphors as reflexives would be limited to inherently reflexive predicates. This is a testable prediction. Turning first to Peranakan Javanese, we see that those predictions are not supported by the data: the examples given earlier in this article do not involve inherently reflexive predicates:

- (21b) [Gurue Tono]_i ketok **awake dheen**_{i/j/k} nggon kaca.
 teacher.3 Tono see body.3 3SG in mirror
 'Tono's teacher saw himself/him/her in the mirror.'

Additional examples using different predicates are provided below:

- (55) a. Dhewe_i benci mbek awake (dheen)_{i,j}
 1PL hate with body.3 3SG
 'We hate ourselves.'
- b. Mary_i ngeki medali emas nggo awake dheen_{i,j}
 Mary N.give medal gold for body.3 3SG
 'Mary gave the gold medal to herself/him/her.'
- c. Mary_i ndudoh-i John_j awake dheen_{i,j,k}
 Mary N.show-APPL John body.3 3SG
 'Mary showed John herself/himself/him/her.'

GRAMMAR OF BINDING

- d. Tono_i mukul awake dheen_{i,j}
Tono N.hit body.3 3SG
'Tono hit himself/him/her.'

As the examples of (55) show, a pronominal form that can be used in environments in which ordinary pronouns occur can also be used in reflexive contexts. This is a general property of the anaphoric system, and is not, as predicted by Franssen and Everaert and Franssen, restricted in any way to inherently reflexive predicates. Thus, the Reinhart and Reuland Reflexivity theory of Binding does not provide an explanation for the facts in Peranakan Javanese. (The same pattern is observed in SMI, so the same conclusion can be drawn about SMI as well.)

Turning next to Jambi Malay (TJM dialects), we can see a pattern similar to that just observed in Peranakan Javanese. The use of the bare pronominal form with reflexive predicates is not prevented when the predicate is not inherently reflexive: for example, *cinto* 'love', or *bacakap samo* 'speak with' are not inherently reflexive but bare pronominal forms are still well formed. This is illustrated in the following examples:

- (56) a. arna pikir yanti cinto dio?
Arna think Yanti love 3
'Arna thinks Yanti loves herself/him.'
- b. yanti neŋo? dio di tipi
Yanti N.look 3 LOC TV
'Yanti saw herself/him on TV.'
- c. mama? nik ba-cakap samo ana?-e, laki-e jugo samo dio?-e
mother Nick INTR-talk with child-3 husband-3 also with 3-3
'Nick's mother loves her child, her husband and also her/herself.'

Thus, we conclude that the Reflexivity theory of Binding and the Classical Binding Theory make essentially the same incorrect predictions about the languages discussed in this paper.

11. Implications and Conclusions

We have argued that the facts of Binding in the languages we have discussed are inconsistent with the view that all languages share the same grammatical rules and constraints, and that the seeming differences among rules in different languages are attributable to properties of the vocabulary items found in each individual language rather than to variations in the rule systems themselves. Rather, the evidence seems to support the view that the typical division of anaphora into two classes, reflexives and pronouns, is due to the grammaticalization of functionally based tendencies related to the facilitation of processing, etc. The final issue we would like to discuss is whether the evidence we have presented is also persuasive if it is assumed that universals of Binding (and, perhaps, of syntax generally) are applicable only to the initial state of the language learner, and that, pace Chomsky's recent views, the initial state can be modified in significant ways on the basis of experience. We acknowledge that this weaker position is much more difficult to argue against than are Chomsky's current views. It is possible to imagine a route for diachronic change by means of which one could modify an initial (UG-based) grammar in which anaphora are divided into two complementary classes so that the resulting state would be a single class of anaphors. Thus, the mere existence of the languages discussed in this article would not constitute a serious analytical problem for this weaker view. There is, however, evidence that even the weaker view is incorrect. Note that according to the weaker view discussed in this section, the direction of linguistic change must be from an initial state of affairs with two distinct classes of anaphors (specified by UG) to one with only one such

class. In fact, the direction of change seems to be in the opposite direction. For instance, as we discussed above, a single indeterminate anaphor in Old and Middle English has divided into separate reflexives and pronouns. The same direction of change appears to have occurred in the Malay varieties discussed in detail above. It is notable that we cannot find a single example in the literature reporting an instance of historical change in which the direction of change is from distinct reflexives and pronouns to a single class of indeterminate anaphors. Thus, the direction of diachronic change suggests that change must always be from a single indeterminate anaphor to two classes of anaphora, and not the opposite direction. This state of affairs is inconsistent with the proposal that UG specifies a state of affairs roughly like that specified by the classical Binding Theory, a state of affairs obscured by later historical change.

The facts regarding creole languages and signed languages also constitute counter evidence to the view that UG specifies two classes of anaphora. If the existence of separate classes of reflexives and pronouns is specified by UG as the initial state of language, instances of *de novo* creation of language, which many presume creoles and new signed languages to be, should uniformly manifest these two classes of anaphora, and the division into these classes should be less frequent in non-creole/non-signed languages. This, however, is not the case. As we saw in our earlier discussion, the existence of a single class of anaphor in the early stages of creoles and signed languages seems to be much more frequent than in language generally. This state of affairs is also inconsistent with the view that the initial state of language is one in which UG specifies something along the general line of the classical Binding Theory.¹⁸

¹⁸ While most Generative syntacticians seem to regard Binding as determined at least in large part by UG, in recent work Chomsky acknowledges the possibility that the patterns observed in

These considerations lead us to the conclusion that much of syntax is in fact learned though it does not argue against the possibility that other aspects are in fact innate. Regardless of whether e.g. the basic principles of sentence construction (the principles constraining “Merge” and “Move”, in Minimalist parlance) are innate, speaker must have the tools to learn complex syntactic patterns that were previously thought by Generative syntacticians to be innate. Thus, a key issue is how such patterns can be acquired in the absence of negative data. While many functionalist linguists believe that the answer is to be found in the stochastic learning abilities of the child (e.g. Saffran, Aslin, & Newport, 1996 and many later works), Generative linguists remain skeptical about the ability of stochastic learning to adequately account for the complex facts of syntactic systems. Thus, the explanation of how it is possible for the language learner to acquire the complexities of grammar remains a central research issue for the future.

Acknowledgment

We would like to thank the many friends and colleagues who helped us to understand the issues surrounding anaphora in Indonesian, Javanese and in the colloquial Malay varieties of central Sumatra. Special thanks go to Bernard Comrie, David Gil, Martin Haspelmath, Jaklin Kornfilt,

Binding as well as in some other areas that were traditionally viewed as core areas of syntax must be seen as outside the scope of UG:

“Conditions on representations—those of binding theory, Case theory, ϕ -theory, and so on—hold only at the interface, **and are motivated by properties of the interface**, perhaps properly understood as modes of interpretation by performance systems” (Chomsky, 1995, p. 170-171, emphasis added).

Timothy Mckinnon, Susanne Michaelis and Uri Tadmor for their many thoughtful suggestions and challenging criticisms. Brad Taylor provided us with the computational tools we needed to organize and find the patterns in the data. Thanks to Judy Kegl for information on reflexives (or rather the absence of reflexives) in Nicaraguan Sign Language. We would like to offer special thanks to our reviewers, whose comments and criticisms forced us to sharpen our views and to confront important issues that we had evaded in the paper as originally submitted.

Much of the research we report on here was carried out in the rural areas surrounding Jambi City, Indonesia, at the Jakarta Field Station of the Max Planck Institute for Evolutionary Anthropology Institute (MPI EVA), and at the Department of Linguistics of the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany. MPI EVA has also provided us with many sorts of logistic and computational help in carrying out our project.

This project was funded in part by the National Science Foundation (BCS-0444649 and BCS-1126149), the Max Planck Institute for Evolutionary Anthropology, the College of Arts and Sciences of the University of Delaware, and the American Institute for Indonesian Studies. We thank them all for their generous assistance.

References

- Besnier, N. (2000). *Tuvaluan: A Polynesian language of the Central Pacific*. London: Routledge.
- Bickerton, D. (1981). *Roots of language*, Ann Arbor: Karoma Publishers, Inc.
- Bickerton, D. (1983). Creole languages. *Scientific American*, 249 (8), 116–122. doi: 10.1038/scientificamerican0783-116

- Bickerton, D. (1984). The language bioprogram hypothesis, *The Behavioral and Brain Sciences*, 7(2), 173-188. doi: 10.1017/S0140525X00044149
- Bickerton, D. (1988). Creole languages and the bioprogram, in F.J. Newmeyer (Ed.), *Linguistics: The Cambridge survey: 2, Linguistic theory: Extensions and implication* (pp. 268-284). Cambridge: Cambridge University Press.
- Bickerton, D. (1991), On the supposed 'gradualness' of creole development, *Journal of Pidgin and Creole Languages*, 6(1), 25-58.
- Campbell, L. & Janda, R. (2001). Introduction: Conceptions of grammaticalization and their problems. *Language Sciences*, 23(2-3), 93-112.
- Chomsky, N. (1986). *Knowledge of language: Its nature, origins and use*. New York: Praeger.
- Chomsky, N. (1995). *The Minimalist Program*. Cambridge, Mass.: The MIT Press.
- Cole, P. & Hermon, G. (1998a). Long distance reflexives in Singapore Malay: An apparent typological anomaly, *Linguistic Typology*, 2(1), 57-98.
- Cole, P. & Hermon, G. (1998b). VP ellipsis and Malay reflexives. In A.Z. Wyner (Ed.), *IATL 5: The Proceedings of the Thirteenth Annual Conference of the Israeli Association for Theoretical Linguistics* (pp. 39-54).
- Cole, P. & Hermon, G. (2005). The typology of Malay reflexives, *Lingua*, 115, 627-644.
- Cole, P., Hermon, G., & Huang, C.-T.J. (2006). Long-distance binding in Asian languages. In M. Everaert & H.V. Reimsdijk (Eds.), *The Blackwell Companion to Syntax*, Vol. 3, ch.39. Oxford: Blackwell.
- Cole, P., Hermon, G., Tjung, Y., Sim, C-Y., & Kim, C. (2007). *Anaphoric expressions in the Peranakan Javanese of Semarang*. Lincom Studies in Asian Linguistics 72, Lincom Europa (LINCOM GmbH, Gmunder Str. 3 5, D-81379 München)

- Cole, P., Hermon, G., Tjung, Y., Sim, C-Y., & Kim, C. (2008). A Binding Theory exempt anaphor. In E. König & V. Gast (Eds.), *Reciprocals and Reflexives: Theoretical and Typological Explorations*. Trends in Linguistics. Studies and Monographs [TiLSM] 192. Berlin: Mouton de Gruyter.
- Cole, P., Hermon, G., & Yanti. (2010). Anaphora in Traditional Jambi Malay. In R. Mercado, E. Potsdam and L. Travis (Eds.), *Austronesian and Theoretical Linguistics* (pp. 327-343). Amsterdam/Philadelphia: John Benjamins.
- Crain, S., Khlentzos, D., & Thornton, R. (2010). Universal Grammar versus language diversity, *Lingua*, 120(120), 2668-2672.
- Evans, N. & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science, *Behavioral and Brain Sciences*, 32, 429-492. Doi: 10.1017/S0140525X0999094X.
- Everaert, M. & Franssen, F. (2012). Reflexivization in the Austronesian language family. Abstract for conference presentation at 12 ICAL: Twelfth International Conference on Austronesian Linguistics, 2-6 July 2012, Bali, Indonesia.
- Frank, R. (2002). *Phrase Structure Composition and Syntactic Dependencies*. Cambridge, Mass: MIT Press.
- Franssen, F. (2010). Australian and Austronesian anaphora: An archival approach. Unpublished Master Thesis in Linguistics: The Study of the Language Faculty, Utrecht University. Retrieved from: <http://igitur-archive.library.uu.nl/student-theses/2010-0908-200241/AaAA%20-%20AAA.pdf>

- Gil, D. (2001). Reflexive anaphor of conjunctive operator: Riau Indonesian *sendiri*. In P. Cole, G. Hermon, & C.-T. J. Huang (Eds.), *Syntax and Semantics: Long Distance Reflexives* (pp. 83-117). New York: Academic Press.
- Haspelmath, M. (2008). A frequentist explanation of some universals of reflexive marking. *Linguistics Discovery*, 6(1), 40-63.
- Huang, C. -T. J., Li, Y.-H. A., & Li, Y. (2009). *The Syntax of Chinese*. Cambridge: Cambridge University Press.
- Joos, M. (1958). *Readings in linguistics: The development of descriptive linguistics in America Since 1925*. (Ed.) Washington: ACLS, P. 96.
- Keenan, E. (2002). Explaining the creation of reflexive pronouns in English. In D. Minkova & R. Stockwell (Eds.), *Studies in the history of English: A Millennial Perspective* (pp. 325–355). Berlin/New York: Mouton-De Gruyter.
- Kegl, J. (2008). The case of signed languages in the context of pidgin and creole studies. In J. Singler & S. Kouwenberg (Eds.), *The handbook of pidgin and creole studies* (pp. 491–511). London: Blackwell's Publishers.
- Kegl, J. (in progress). Reflexive marking in Nicaraguan sign language.
- Kimmelman, V. (2011). Binding theory and sign languages: the role of signing space. Venice FEAST Colloquium Formal and Experimental Advances in Sign language Theory. Università Ca' Foscari, Venice, June 20-22, 2011.
- Kiparsky, P. (2002). Disjoint reference and the typology of pronouns. In I. Kaufmann & B. Stiebels (Eds.), *More than words*. Studia Grammatica 53 (pp. 179-226). Berlin: Akademie Verlag, 2002.

- König, E. and Gast, V. (2006). Focused assertion of identity: A typology of intensifiers, *Linguistic Typology*, 10, 223-276.
- König, E. & Siemund, P. (2000). The development of complex reflexives and intensifiers in English, *Diachronica*, XVII(1), 39–84.
- Lehmann, C. (1982/1995). Thoughts on grammaticalization: A programmatic sketch, Vol. 1. *Arbeiten des Kölner Universalienprojekts* 49 (1982) [Department of Linguistics, University of Cologne, Cologne]. Revised as *Thoughts on Grammaticalization*, LINCOM Studies in Theoretical Linguistics, 1. LINCOM Europa, Munich, 1995.
- Levinson, S.C. (2000). Presumptive meanings: The theory of generalized conversational implicature. Cambridge, MA: MIT Press.
- Haspelmath, M. & the APiCS Consortium. (2013). Reflexive constructions. In Michaelis, S.M., Maurer, P., Haspelmath, M., & Huber, M. (Eds.), *The Atlas of Pidgin and Creole Language Structures*, (pp. 346-349). Oxford: Oxford University Press.
- Reinhart, T & Reuland, E. (1993). Reflexivity, *Linguistic Inquiry*, 24, 657-720.
- Reuland, E. (2011). Anaphora and Language Design. Linguistic Inquiry Monographs. Cambridge, MA: Mit Press.
- Saffran, J.R., Aslin, R.N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274 (5294), 1926–1928.
- Sneddon, J.N. (1996). *Indonesian reference grammar*. New South Wales: Allen and Unwin.
- Sneddon, J. N. (2006). *Colloquial Jakarta Indonesian*, Canberra, Australia: Pacific Linguistics, Research School of Pacific and Asian Studies, the Australian National University.
- Safir, K. (2004). The Syntax of Anaphora. Oxford: Oxford University Press.

GRAMMAR OF BINDING

- Tjung, Y. (2006). The formation of relative clauses in Jakarta Indonesian: a subject-object asymmetry, Ph.D. dissertation, University of Delaware.
- Wouk, F. (1999). Dialect contact and koineization in Jakarta, Indonesia. *Language Sciences*, 21, 61-86.
- Yanti. (2010). A reference grammar of Jambi Malay, Ph.D dissertation, University of Delaware.