An Analysis of Current Theories of Anaphora:
The delineation of the binding of English 'HIMSELF' into syntax and discourse

Many current theories of anaphora appeal both to syntax *and* discourse, where older theories, such as Chomsky's Government-Binding (hereafter 'GB'), simply called upon syntax. This "move" to discourse, so to speak, endeavored to account for grammatical instances of 'HIMSELF' in environments producing ungrammatical judgments by GB. While these theories are alike in their dual appeal, they differ in where they draw the line between what role syntax plays and what role discourse plays in determining the grammaticality of 'HIMSELF'. My goal here is to analyze where the respective lines are drawn and offer (arguably) more plausible[1] alternatives.

0. Introduction

In the last decade, many theories of anaphora have been formulated which in one way or another incorporate discourse with syntax to explain grammaticality of English SELF anaphora. The three very different theories I chose to analyze are the ones presented in Reinhart & Reuland (1991, (hereafter 'R&R')), Safir (1992), and Zribi-Hertz (1995, (hereafter 'ZH')).

However, before delving into theories of anaphora, it would be useful to define 'anaphor', characterize the function of binding, and explain why I limit my analysis to 'HIMSELF'. The definition of 'anaphor' that will be adopted here derives essentially from R&R (1991): An anaphor is a referentially defective NP (such as 'HIMSELF'). That is, anaphors do not explicitly refer to anyone or anything.[2] To use them

---

[1] For an alternative to be "more plausible", I mean the following: a) It should reflect native speaker intuition. b) It should arise from independently motivated principles and/or well-documented phenomena.

[2] Anaphors are referentially defective because of certain attributes of the "semantic atoms" they are composed of. These atoms are for the most part relational, meaning that they have an argument structure. For example, the SELF atom as in English 'himself' takes two arguments: 1) the pronoun 'him', and 2) an external argument such as 'John' in **John$_i$ hit himself$_i$**. So, for an anaphor to successfully refer, both of its arguments must be saturated. For a more in-depth analysis, see Safir (1996) and/or R&R (1991).

grammatically, a reference must be fixed or *bound* to them via a process called binding. Different theories have different things to say about binding, e.g. whether it is strictly a syntactic process, or one that is manifested in both syntax and discourse, but that question is set aside for future discussion.

Although the essential issue here is the delineation of binding into syntax and discourse in theories of *anaphora*, I have limited my analysis to English SELF anaphora. There are three reasons behind this. Firstly, English is the only language I am fluent in, and as such the only one I feel knowledgeable enough about to analyze. Secondly, I believe many of the current theories of anaphora commit what could be a grave oversight: conflating autonomous atoms of anaphora[3] in an effort to develop a theory of binding that is all-encompassing, i.e. one that accounts for SELF, SAME, OWN, OTHER, *and* SE anaphors. While the results of such an account would seemingly be complete, it would run the risk of potentially ignoring the semantic content of each atom in relation to its distribution. This is important, since a theory that *describes* all phenomena is not as desirable as a theory that *explains* all phenomena. One that ignores the semantic content of anaphors cannot be said to explain very much at all.[4] Thirdly, although an account of English SELF anaphora is not exhaustive in that it does not account for other anaphoric atoms or cross-linguistic data, it is still a perfectly interesting one: It would, in theory, explain the difference between the distributions of English HIM (which will be termed a 'pronominal') and HIMSELF, intuitively predicting the grammatically of all environments in which the SELF anaphor would appear.

---

[3] I assume Safir (1996) to be essentially correct, at least in its semantic views. The syntactic (binding) claims are omitted from my analysis.

[4] This is not to say that a theory that covers all atoms of anaphora must be ignorant of semantic content, but to say that it would be a rather easy thing to take for granted when trying to amalgamate these distinct atoms under one set of principles.

1. THE GB THEORY: ITS CONTENT AND FAILURE

One of the pioneering theories of anaphora which attempted to explain the difference between the distributions of HIM and HIMSELF was Chomsky's Government-Binding theory. Crucial notions of this account are as follows:

Principle A - An anaphor α must be bound in its binding domain.

Principle B - A pronominal α must be free in its binding domain.

Binding domain - The minimal maximal category containing α, its governor, and an accessible subject[5].

Binding - For α to be bound, it must be c-commanded by a coindexed β. α is said to be free if it is not bound.

Principles A and B are structured in a way such that anaphors and pronominals should be in complementary distribution; In a given environment, if an anaphor is grammatical then a pronominal isn't, and vice versa. This follows from that fact that bound and free elements are mutually exclusive, by definition. Theoretically speaking, this would entail that there is no environment where both 'HIM' and 'HIMSELF' may be used grammatically. Hence, if there exists data which exhibits non-complementarity between a pronominal and an anaphor, it would directly contradict this version of GB[6].

(1)a. The grotesque photo of (her_i/$herself_i$) in the paper bothered $Mary_i$.
  b. $John_i$ said there was a picture of (him_i/$himself_i$) hanging in the post office.
  c. Mary said that a picture of (you_i/$yourself_i$/me_i/$myself_i$) would be nice on the wall[7].
  d. $John_i$ didn't hear the snake behind (him_i/$himself_i$).

---

[5] A subject called "accessible" is one which does not violate the *i within i* condition - a topic beyond the scope of this discussion.

[6] Note that there are numerous syntactic fixes and revisions of this very theory. However, I find some of them to miss the point.

[7] Examples b,c taken from ZH '89.

e. Max$_i$ enjoyed the jokes about (him$_i$/himself$_i$).
f. John$_i$ said that he$_i$ would never allow his daughter to even consider
   marrying a man similar to (him$_i$/himself$_i$).
g. Milton$_i$ warned Marsha that she shouldn't trust anyone other than
   (him$_i$/himself$_i$)[8].

In (1), the environments in which the pronominals/anaphors appear all display non-complementary distributions. Thus, all of these examples contradict the predictions GB makes. The data in (1) is interesting, however, in that it is comprised of 4 distinct types of categorizable GB violations: "psych" predicates (1a), "picture NPs" (1b,c), adjunct PPs (1d,e), and contrastive predicates[9] (1f,g). While these four constructions which systematically violate GB are not exhaustive of its failures, they are sufficient to prove that (the above formulation of) GB is by itself insufficient as a theory of anaphora.

Furthermore, one of they key goals in developing a theory within linguistics is to capture native speakers' intuitions; In layman's terms, "Things happen for reasons that make sense." GB's reliance upon the notion of c-command in its formulation of binding is crucial: without c-command, there is no binding. And with no binding, there is no theory. The point I am getting at is this: "What exactly does c-command *mean* or *represent*?" It is essentially a precedence relation, and that in itself does little to "make sense" of binding. This is more than a personal pet peeve however: Tanya Reinhart, the inventor of the notion of c-command in 1975, abandoned its usage in her theory just one year after inventing it and subsequently turned to developing a more intuitive theory[10]. In

---

[8] Examples f,g taken from Safir '92.
[9] "Contrastive predicate" is a phrase used in Safir '92 which subsumes similarity, comparative, and exclusion predicates. More information on this can be found in section 2.2 of this paper, where I analyze Safir '92 in detail (with regard to the discourse/syntax distinction).
[10] For more information on R&R's latest theory of anaphora (which avoids c-command entirely), see R&R '93.

addition to that, Head-driven Phrase Structure Grammar (hereafter 'HPSG'), an alternative syntactic model to GB, abhors the usage of c-command in its theory.

<p style="text-align:center">2.0 ALTERNATIVE THEORIES OF ANAPHORA</p>

GB and HPSG together do not comprise the whole of theories of anaphora. In fact, there are many other very interesting theories that are quite distinct from these. The focus of this section is examining and analyzing how a few of these interesting theories delineate binding of English SELF anaphora into syntax and discourse. Together, they shed some light on what the specific job of syntax should and should not be, and suggest strongly that discourse *must* be involved in a creditable theory of anaphora.

<p style="text-align:center">2.1 THE ARGUMENT STRUCTURE PERSPECTIVE</p>

R&R begin this paper by pointing out flaws in the GB theory, which was the standard for quite some time. GB is far too restrictive in trying to develop a theory of anaphora which predicts strict complementarity in local domains. Realizing that so many SELF anaphors fall outside of this local domain, R&R propose that discourse licenses them as something they call logophors[11]. Hence, there are two types of SELF anaphors: locally bound anaphors and logophors. Both types of anaphors share their morphology, and part of this morphology is the SELF element. R&R argue that the SELF element is a relation between some x and some y, x being the pronominal "attached" to SELF, and y

---

[11] R&R use this term rather loosely, but essentially relate logophoric anaphora to perspective or point-of-view: something shown insufficient for licensing anaphora (Baker 1995). I present more in-depth information on this in section 2.3.

being some external argument held in the identity relation to x (and in fact this is rather well-documented outside of R&R, e.g. Safir 1996). For example, in the sentence from footnote 2 "*John$_i$ loves himself$_i$*"; 'self' is the SELF morpheme, 'him' is the x argument of SELF, and 'John' is the (external) y argument of SELF. It is important to note that complementarity is for the most part maintained, but in R&R's theory it holds between pronominals and locally bound anaphors. The significance of complementarity is, again, due to the linguist's goal to discover a system which explains why one uses HIM instead of HIMSELF, or vice versa. Without the complementarity, it would be an easy move for one to simply claim that free variation exists.

R&R incorporate the notion of movement into their theory; They propose that the SELF element of a locally-bound anaphor moves to V°, effectively reflexivizing the predicate. As a loose definition of reflexivizing: the subject and object of a predicate both refer to the same entity. This movement illustrated in terms of the "*John$_i$ loves himself$_i$*" example yields "*John$_i$ SELF-loves him$_i$*". Formulated this way, (locally bound) anaphors can be viewed as restrictions on predicates.

According to their theory, the binding domain for anaphora should be more local than it was determined to be in GB, i.e it should be restricted to arguments on a θ-grid. This move makes significant improvements on GB: incorporating thematic roles into binding and essentially changing the focus of binding not on the non-intuitive c-command relation but on the relations of predicates, e.g. reflexivity[12]. Along with these improvements, Principles A and B are also reformulated to reflect the change from c-command to reflexivity.

---

[12] However, there are some self-admitted (potential) problems with a theory such as this: ECM structures and picture NPs. R&R defend that those problems have either been misanalyzed or can be said to fall under logophoricity. ECM structures tend to not show complementarity, so that is a definite problem. ECM structures within a theory built upon this are discussed in detail in R&R '93.

Principle A - SELF occurring on a grid position of fully-assigned predicate P reflexivizes P.

Principle B - A predicate is reflexive (i.e. has coindexed coarguments) if and only if it is reflexive marked (either by a SELF morpheme or by its intrinsic morphology).

Principle B effectively rules out all reflexive predicates that are not marked as being so. For instance, in "*John$_i$ loves him$_i$*", the predicate "loves" is reflexive, because it has coindexed coarguments ("John" and "him"), yet there is no SELF morpheme to move to V° to reflexive mark the predicate, nor is "loves" intrinsically reflexive. Hence, that sentence is given an ungrammatical judgment by Principle B, a judgment matched by GB's Principle B in this environment. Principle A also functions like GB's Principle A in that it functions to filter out free occurrences of non-logophoric anaphora in argument positions.[13] In "*John$_i$ thinks Mary$_j$ loves himself$_i$*", 'himself' falls upon the argument grid for "loves", but doesn't mark the predicate as reflexive, as the external argument is "John."

With locally-bound anaphora accounted for, R&R then must turn to detailing their views on logophoricity. In section 6.1.2, there seems to be a contradiction. Earlier in their paper, they concede that local anaphors allow logophoric uses. However, in section 6.1.2, they start by noting that logophoricity is only possible in environments where non-complementarity holds. Sentences like "*John$_i$ loves myself$_i$*", where the SELF morpheme is put under contrastive stress, show that an anaphor occupying a position on a thematic grid, can in fact be used logophorically (unless, of course R&R contend that this

---

[13] This is not say that logophoric anaphora never appear in an argument position; Indeed they do, and they violate Principle A. Presumably though, the SELF anaphors in those environments are stressed, indicating some feature of discourse, which frees them from being susceptible to the syntactic principles.

contrastive stress is not related to logophoricity at all and that there are yet even more discourse conditions factoring in with syntactic binding domains).

Then what is logophoricity? R&R believe logophoricity to be about point of view. Since all (according to R&R) logophoric environments show non-complementarity, then that means when a speaker utters a sentence with one of these environments, he is presented with a choice: whether to use a logophor or pronominal. The choice of logophor over pronominal is made to "mark" the morpheme as representing some point of view. R&R's account of point of view, however, is poorly neglected and underdeveloped, as "marking" a morpheme doesn't necessarily have to indicate point of view, i.e. it could indicate prominence, contrastiveness, and a host of other discourse phenomena[14]. Regardless of what discourse phenomenon licenses logophors, R&R still have to say something about how they are bound. After all, logophors are still SELF anaphora which need to pick up their reference from somewhere.

R&R note the existence of a "centre" that is associated with every utterance. That is, every utterance is made by someone to someone, at some time and some place. The speaker, addressee, time, and place are all entities that may be referred to deictically, also. This helps expain example (1c) above. Terms like 'here' and 'now', and more importantly 'yourself' and 'myself' may be used in a discourse without prior mention of who they refer to, as they automatically refer by virtue *of* that discourse. So using a logophor like 'himself' involves a relation between that expression the so-called centre, marking the logophor as reflecting point of view (in R&R's view, at least).

2.2 IMPLIED NON-COREFERENCE

---

[14] ZH '89 details the POV theory thoroughly and along with Baker '92 mentions some of the other discourse phenomena mentioned.

While Safir assumes GB principles A and B to be correct, he also points out and develops a theory concerning a large subset of GB-violating data. One of the four types of environments (cited in section 1f,g) that systematically violate GB's principles is contrastive predicates. These include predicates that deal with similarity, exclusion, and comparison: ones that contrast their subjects with their objects. Safir's main question in this paper is "Why do contrastive predicates allow their anaphoric complements to have long-distance antecedents?"

2. (a) These men$_i$ believe that Mary would never consider marrying a man (less wealthy than themselves$_i$/*angry at themselves$_i$).

(b) Milton$_i$ warned Marsha that she shouldn't trust anyone (other than himself$_i$/*in love with himself$_i$).

Taking the minimal pairs in (2a,b) which illustrate that the grammatical contrastive structures are parallel with ungrammatical non-contrastive ones, it is clear that a structural difference can not explain why contrastive predicates allow long-distance antecedents.

Semantically, this class of predicates share some property, e.g. (mentioned above) they make a contrast between their subjects and objects, in terms of identity. Safir calls this relation 'Implied Non-Coreference.' In the predicates 'x other than y' (exclusion), 'x like y' (similarity), and 'x taller than y' (comparative), x and y are implied to corefer: It would seem silly to talk of someone other than himself, like himself, or taller than himself. In fact, to utter a sentence that implies one of those relations would be to utter something contradictory, in the cases of comparative and exclusion, or tautologous, in the case of similarity. Contradictions and tautologies share a special logical distinction, in

that one needs no world knowledge in order to know the truth values of these statements.
In these environments then, the binding domain of anaphors is raised, in effect making
the subjects of contrastive predicates transparent to binding. This move relies upon
discourse (or pragmatics, a la Safir): It entails that a predicate implied non-coreferent has
a subject transparent to binding *because* no real world knowledge is required to
understand that it is impossible for the subject to be a coreferent antecedent to the
complement anaphor.

With the discourse condition fleshed out, the syntactic component of Safir's
theory needs to be explained. While discourse allows the antecedent to be outside of the
contrastive predicate, there is yet nothing to determine the relevant binding domain for
anaphor complements. This is handled by the Potential Antecedent Restriction (hereafter
'PAR'):

> PAR - If the minimal complete functional complex containing x (an anaphor
> or pronominal) also contains a c-commanding antecedent implied non-
> coreferent for x, then the binding domain for K will be the binding domain
> for x.

The PAR is important in that it interacts with the discourse condition of Implied Non-
Coreference in a way that limits antecedents from occuring "too far" away from the
complement. This is effective overall, but does fail in similarity predicates similar to
(1f). It appears that antecedents for similarity predicate antecedents can routinely occur
"too far" away from the complement.

Overall, the implied non-coreference theory is one that seems intuitive in a certain
sense, i.e. the discourse condition allowing binding domains to be extended. There are a
few limitations that arise from it, however. Firstly, and most obviously, it is a theory that

is concerned only with a limited subset of GB violations, albeit a significant one, and sheds little light on the other subsets. Secondly, it fails to explain why similarity predicates can license such long-distance antecedents.

<div align="center">2.3 The Intensive / Reflexive Analysis</div>

The reflexive/intensive analysis is one that derives from three papers: ZH '89, Baker '95, and ZH '95. All three papers focus on how discourse conditions work to license anaphora.

ZH analyzes anaphors much like Safir (1996) and R&R (1991) in that they're comprised of a pronoun and a lexical adjunct (such as 'self', 'same', or 'own'). ZH then points out a fact known to all theories of anaphora; There appear to be locally bound reflexives (or anaphors. When referring to 'HIMSELF', they are virtually interchangeable.) in addition to locally free reflexives. In locally bound reflexives, the lexical adjunct portion of the anaphor functions to cancel disjoint reference constaints. So if 'x loves y', then it is presumed that x and y are disjoint in reference, but if y is actually an anaphor, then that disjoint reference restriction is nullified. That kind of analysis seems backward, though. It is not the case that a SELF form cancels any disjoint reference constraint, but reflexivizes a predicate which turns out to be felicitous if the subject and complement corefer. Locally free reflexives, ones not locally bound, occur outside the minimal subject and tense. In these environments, non-complementarity holds, hence the lexical adjunct is optional. This point was one touched upon and defended by R&R (91).

Chomsky's Principle A is assumed by ZH to be responsible for licensing locally bound reflexives. Locally free reflexives, i.e. anaphors in non-complementary distribution, abide by discourse constraints dealing with logophoricity, contrastiveness, and prominence. To explain these discourse conditions, it would be helpful to describe how ZH (1995) developed from ZH (1989) and Baker (1995).

In her 1989 paper, ZH develops an account in which all occurrences of locally free 'HIMSELF' are logophoric: dealing with point of view and 'subject of consciousness'. Baker, using a huge corpus of literary[15] data, provides ample examples to show that point of view is not the only relevant discourse condition on logophoric anaphora. Baker goes on to analyze 'HIMSELF' not strictly as a reflexive, but also as an intensive[16], such as 'HIM HIMSELF'. These intensives serve to pick out a figure in a discourse and both mark him as more prominent than another and to contrast him with a less prominent figure. Intensives, then, abide by two discourse conditions:

Contrastiveness Condition - Intensives are appropriate only in contexts in which emphasis or contrast is desired

Prominence Condition - Intensives can only be used to mark a character in a sentence or discourse who is relatively more prominent or central than other characters.

Upon further examination of the Prominence Condition, it is obvious that the property of logophoricity is subsumed. That is, for anything to be logophoric, it is also prominent. This removes the condition that all locally free anaphors must have the property of logophoricity, while adding more thrust to the theory. For instance, the anaphor in the

---

[15] While the data Baker presents is vast, it is for the most part all taken from 19th century British novels. As written language, its grammaticality judgments may be marginally different than those in spoken language. This argument appears in Pollard & Sag (1992), specifically pp. 278-279.

[16] Baker doesn't explain exactly how or why it is possible to have one rather simple morpheme performing two very different jobs: 'HIMSELF' functioning as a reflexive and an intensive. In section 3 of this paper, I mention a theory which endeavors to explain just that.

sentence *"Give it to the King$_i$ himself$_i$!"* can now be accounted for by appealing to discourse prominence, as it seems to have nothing to do with point of view[17].

ZH assumes most of Baker's analysis to be right, and concludes the following way:

English 'HIMSELF' has two effects which follow from SELF adjunction. First, it can serve as a marking device (or intensive), as in non-complementary environments, where the speaker has a choice of whether to use an anaphor or pronominal. Second, it is linked to the de-stressing of 'HIM', which leads to the bound anaphoric interpretation. As an intensive, it must be licensed by one or both discourse conditions brought to light in Baker's theory. As a reflexive, it must be licensed by GB's Principle A. How exactly discourse and syntax interact, however, is left open.

3.0 CONCLUSION

In summary, what I set out to do is present first a theory of anaphora which didn't include discourse at all (GB), then show several theories that did incorporate discourse to some success. While Safir (1992) and ZH (95) both assume basic GB to be correct for the syntactic portions of their theories, I believe R&R (91) to have the most intuitive and explanatory theory of syntax (a la anaphora). That is, I believe R&R's syntax does what it is supposed to do without overstepping its bounds or trying to do too much.

With regard to discourse, I believe the ZH/Baker analysis to be the most persuasive: with bound anaphors taking on the reflexive role and free ones taking on the

---

[17] *"Give it to himself!"* where 'himself' refers to the king and is accompanied by an ostensive gesture, however, is still unaccounted for.

intensive one. This raises a potentially disturbing question, though. How can one simple morpheme, such as 'HIMSELF', function in two entirely different ways? The answer to that question may lie in the past. According to Keenan (1994), the 'HIMSELF' form originally came from two morphemes which were used in ways similar to intensives. So, the reflexive form, in effect, developed out of the intensive form: a diachronic solution to a synchronic problem.

REFERENCES

Baker, C.L.: 1995, 'Contrast, Discourse Prominence, and Intensification, with Special Reference to Locally Free Reflexives in British English', *Language* **71**, 63-99.

Keenan, Edward L.: 1994, 'Creating Anaphors: An Historic Study of the English Reflexive Pronouns', unpublished ms., UCLA, 1-40.

Koster, Jan and Eric Reuland (eds.): 1991, *Long-Distance Anaphora*, Cambridge University Press, Cambridge.

Pollard, Carl and Ivan A. Sag: 1992, 'Anaphors in English and the Scope of Binding Theory', *Linguistic Inquiry* **23**, 261-303.

Reinhart, Tanya and Eric Reuland: 1991, 'Anaphors and Logophors: An Argument Structure Perspective', in Koster and Reuland (eds.), pp. 283-321.

Reinhart, Tanya and Eric Reuland: 1993, 'Reflexivity', *Linguistic Inquiry* **24**, 657-720.

Rooryck, J., Pierre Pica, and H. Bennis (eds.): (To appear), *Atomism and Binding*, Foris, Dordrecht.

Safir, Ken: 1992, 'Implied Non-Coreference and the Pattern of Anaphora', *Linguistics and Philosophy* **15**, 1-52.

Safir, Ken: 19??, 'Symmetry and Unity in the Theory of Anaphora', in Rooryck, Pica and Bennis (eds.), 1-39.

Safir, Ken: 1996, 'Semantic Atoms of Anaphora', *Natural Language and Linguistic Theory* **14**, 545-589.

Zribi-Hertz, Anne: 1989, 'Anaphor Binding and Narrative Point of View: English Reflexive Pronouns in Sentence and Discourse', *Language* **65**, 695-727.

Zribi-Hertz, Anne: 1995, 'Emphatic or reflexive? On the endophoric character of French *lui-même* and similar complex pronouns', *Linguistics* **31**, 333-374.