The Anaphor Agreement Effect: further evidence against anaphora-as-agreement

Omer Preminger (University of Maryland)

The Anaphor Agreement Effect (henceforth, *AAE*; Rizzi 1990, Woolford 1999)—the avoidance of phi-feature agreement with reflexives—suggests that anaphoric binding and phi-agreement interact with one another in a rather close fashion. This, in turn, is often taken to suggest a reduction of anaphoric binding to the same formal mechanism underpinning phi-agreement (see, for example, Reuland 2011:261–262).

I will argue that, upon closer inspection, the AAE provides evidence <u>against</u> the unification of anaphoric binding with phi-agreement.

The AAE and binding/agreement unification: Let F_{phi} be the formal relation (or process) that values the phi-features on a functional head using the phi-values found on one or more DPs (say, Chomsky's *Agree*, or some variation thereof). If anaphoric binding is to be unified with phi-agreement, it follows that F_{phi} must also be how an anaphor enters into a binding relationship with its antecedent. Antecedents can be, and often are, phi-complete (they can be R-expressions, for example). If the antecedent and the anaphor have entered into F_{phi} together, then—all else being equal—the anaphor should have every opportunity to acquire phi-feature values from the antecedent (and, on the approach that unifies binding with phi-agreement, this is exactly what gives rise to morphologically expressed phi-features on the anaphor). Why, then, would a(nother) functional head (e.g. T^0 or v^0) fail to find phi-values on the anaphor?

One potential answer has to do with "derivational timing": we may suppose that the anaphor has not yet acquired phi-feature values from the antecedent *at the stage where it is probed by* T^0 *or* v^0. This approach, however, is ill-suited to account for the AAE. First, we know that phi-agreement is, to some extent, a countercyclic process (Holmberg & Hroarsdottir 2003), meaning that such strict timing-based approaches to bleeding are a poor fit for phenomena that purportedly reduce to F_{phi}. More challenging yet, reducing binding to F_{phi} requires that F_{phi} be able to transmit feature values <u>downward</u> (Zeijlstra 2012, Wurmbrand 2012, 2014, i.a., *pace* Preminger 2013). If this is so, then F_{phi} must be even more radically countercyclic than supposed above: F_{phi} must be able to alter features on an embedded syntactic object long after that object is no longer near the current root of the structure. If so, it is no longer clear how timing could have anything to do with the AAE: T^0 or v^0 could simply probe the phi-features of the anaphor *after* the latter has acquired its phi-values from the antecedent.

An approach that initially seems more promising is what I'll call "encapsulation": the idea is that anaphors do enter into F_{phi} with their antecedents, but that the part of the anaphor whose phi-features are consequently valued is only a proper sub-constituent of the anaphor in its entirety. Thus, when probed from the outside, the anaphor will appear to lack the relevant phi-feature values. While this may well account for the AAE—see below for a particular implementation of this idea—it means that anaphoric binding (*qua* the mechanism that regulates the distribution of binding indices on anaphors) very clearly does <u>not</u> reduce to F_{phi}. To see why, consider English *−self* anaphors. As the following examples show, the binding index on *himself* must reside on its outermost structural layer, in order to c-command the pronoun in (2) and trigger the attested violation of Condition B. (The contrast with (1) shows that it is indeed the anaphor in (2) that triggers this violation, not the higher R-expression.)

(1) John$_{\mathbf{i}}$ expects Mary to outdo him$_{\mathbf{i/k}}$.

(2) John$_{\mathbf{i}}$ expects himself$_{\mathbf{i}}$ to outdo him$_{\mathbf{k/*i}}$. *[anon., pers. comm.]*

English thus counterexemplifies the predictions of the "encapsulation" hypothesis as it pertains to the indices of anaphoric binding. Therefore, if F_{phi} is supposed to be what regulates those indices (and thus, binding indices and phi-features are to go hand-in-hand), "encapsulation" could not hold for phi-features, either—invalidating "encapsulation" as an account of the AAE.

A related argument can be found in the behavior of reflexive anaphors in Basque. These have the structure in (3), where PRON.GEN is a possessive pronoun, and N is the noun *buru* ('head').

(3) [PRON.GEN N D] — e.g.: *ene buru-a* 'my head-DET' (≡ "myself")

Bound reflexives in object position in Basque trigger 3sg absolutive agreement, regardless of the phi-features of their antecedent, or of PRON.GEN (Rebuschi 2003). On a view that reduces anaphoric binding to F_{phi}, this means that the relevant binding index cannot reside at the level of the entire DP in (3), and instead must reside on its phi-bearing subpart, PRON.GEN. But this is incongruous: the possessor in (3) is pronominal, not anaphoric; it is the same type of element that occurs in cases of deictic, non-reflexive pronominal possession. As such, it is subject to Condition B, not Condition A—

leaving the fact that an expression like *ene burua* (3) requires a local (roughly, clausemate) antecedent entirely mysterious from a binding-theoretic perspective.

For Basque, then, we see that "encapsulation" fails outright as a theory of the distribution of binding indices. But since the phi-features of Basque anaphors are very clearly encapsulated (i.e., even a 1sg anaphor like the one in (3) triggers 3sg absolutive agreement), this constitutes fairly direct evidence against reducing anaphoric binding to phi-agreement.

Whence the AAE? Since accounts of the previous sort—which attempt to derive the AAE from a unification of binding with agreement—seem to fail, we obviously need some other way of deriving the AAE. An account of the AAE must deliver a (perhaps exceptionless) avoidance of the following two conditions holding of one and the same XP:

(4) a. XP bears some anaphoric index *i*
b. XP has been targeted for phi-agreement by some head H^0

It strikes me that a likely story here is one that appeals to substantives: the kind of elements that bear an anaphoric indices (call them, in the interest of perspicuity, *AnaphP*, which may or may not unify in the end with more familiar projections) are simply not the kind of elements that bear phi-features (call them, in the interest of perspicuity, *PhiP*, which, again may or may not unify in the end with more familiar projections). And, moreover, as the case of Basque vividly illustrates, AnaphP is structurally higher than PhiP (when both are present). This state of affairs should sound familiar to any practicing syntactician: it is par for the course to assume that certain substantives cannot "cohabitate" on a single head. For example, syntactic verbs (say, *v* or V) are not—perhaps ever—the kinds of things that bear tense; that is the purview of a separate syntactic object, namely T/Infl. (This is not to be confused with the fact that the two often end up spelled out as part of the same morphological complex; that much, in fact, is also going to be true for many cases of AnaphP and PhiP, too, which is precisely why the attempt to unify binding with agreement was not self-evidently moribund from the get-go.)

Whence phi-matching? The final piece that we may seek to comment on concerns the apparent matching in phi-features between anaphors and their antecedents. If this matching is not delivered by F_{phi}, it must come about in some other fashion. I propose that this comes about via the meaning contribution of phi-features—and in particular, the presuppositional theory of phi-feature meanings (as in Heim & Kratzer 1998). If a DP bears a set of phi-features *S*, the interpretation assigned to that DP must be within a domain restricted by the presuppositions generated by ⟦*S*⟧. Consequently, if an anaphor bears index *i*, the individual *g(i)* (where *g* is the assignment function) must fall within the set identified by ⟦*S*⟧. Thus, if the antecedent bears plural phi-features, and bears index *i*, then any bindee bearing index *i* will only be interpretable if its set of phi-features *S* carves out a set of individuals that includes *g(i)*, i.e., plurals. This generally ensures that pairs of expressions bearing the same syntactic index will bear the same sets of phi-feature values.

A common objection to this line of reasoning concerns phi-features which don't seem to be "interpreted" in the usual sense, such as grammatical gender on inanimates, number on pluralia tantum, etc. But this is not problematic on the present approach: if *S* ={PLURAL}, then *by the definition just given*, ⟦"scissors"⟧ ⊂ ⟦*S*⟧ (and similarly for instances of grammatical gender on inanimates). That is, the set of individuals picked out by an *S* that contains the feature PLURAL is exactly those individuals that the phi-feature PLURAL is compatible with in the language in question. This may seem to run counter to the notion of *interpretation function*, since this instance of plural is not interpreted in the usual way. But the same conclusion is forced on independent grounds:

(5) Speaker A: "Where are the scissors?"
Speaker B: "They are right here."

Consider the expression *they* in Speaker B's utterance. It would be logically incoherent to speak of a syntactic relation (e.g. *Agree*) holding between this expression and the expression *the scissors* in Speaker A's utterance. That is because syntactic relations are grammatical entities; a grammar is, by definition, a mental object; and minds are, by definition, confined to individual speakers. For *they* to refer to the scissors in this instance, it must be the case that *S* ={PLURAL} → ⟦"scissors"⟧ ⊂ ⟦*S*⟧, as just suggested.

Summary: We have seen that the AAE does not speak in favor of a unification of anaphoric binding with phi-agreement, and in fact militates against this unification. Furthermore, phi-feature matching between an antecedent and a bindee requires little more than what is independently necessary to negotiate matching phi-features on pronouns across different utterances (and speakers).