

# *The Anaphor Agreement Effect: further evidence against binding-as-agreement*

Omer Preminger\*  
January 2019

---

## **Abstract**

This paper is about the Anaphor Agreement Effect (AAE): the ban on  $\varphi$ -agreement with an anaphor. While the AAE has sometimes been taken to support a reduction of anaphoric binding to syntactic  $\varphi$ -feature agreement, and the latter has supported an account of the AAE in terms of derivational timing, closer inspection reveals the opposite picture. The AAE actually provides evidence against such reductionism. Furthermore, the AAE arises not because of the relative timing of binding and  $\varphi$ -agreement, but because the  $\varphi$ -features of the anaphor are (always) encapsulated within a larger anaphoric structure. Encapsulation not only renders reductionism superfluous in accounting for the AAE, but demonstrates that reductionism cannot be right, in the first place.

## **1. Introduction**

The *Anaphor Agreement Effect* (AAE; Rizzi 1990a, Woolford 1999, *a.o.*) refers a cross-linguistic ban on  $\varphi$ -feature agreement with anaphors. At first glance, the AAE seems to suggest a rather tight interaction between syntactic  $\varphi$ -feature agreement on the one hand, and binding on the other. This has led some (e.g. Reuland 2011) to take the AAE as evidence in support of a reduction of binding to  $\varphi$ -agreement.

I will show that, upon closer inspection, the AAE provides fairly strong evidence against the reduction of binding to syntactic  $\varphi$ -feature agreement. That is because the reductionist view requires, somewhat paradoxically, assumptions about  $\varphi$ -agreement, and about the structure of anaphoric expressions, which break their compatibility with the very mechanism of binding-as- $\varphi$ -agreement.

I will instead propose that the AAE arises due to what I call *encapsulation*: the binding index associated with anaphoric binding resides on a separate projection, higher than the one where valued  $\varphi$ -features reside. In addition to accounting for the AAE, we will see that the encapsulation hypothesis enjoys broad cross-linguistic morphological support (Middleton 2018).

Finally, I will offer a proposal on the nature of  $\varphi$ -feature matching between anaphors and their binders, given that syntactic  $\varphi$ -feature agreement cannot be what is responsible. I will show that independent facts necessitate a mechanism that, even in the absence of any syntactic relation between binder and bindee, ensures  $\varphi$ -feature matching between the two (e.g. in cases of Donkey Anaphora). This is not restricted to semantically interpreted features; the same applies to, e.g., grammatical gender on inanimates. Crucially, the same non-syntactic mechanism is then sufficient to ensure  $\varphi$ -matching between an anaphor and its binder, with no involvement of syntactic  $\varphi$ -feature agreement.

---

\*Thanks to Rafael Abramovitz, Karlos Arregi, Seth Cable, Winfried Lechner, Aitor Lizardi Ituarte, Jeffrey Lidz, Jason Merchant, Gurujegan Murugesan, Eric Reuland, Giorgos Spathas, Sandhya Sundaresan, Juan Uriagereka, and the audiences at the 1st Annual Thematic Workshop of Linguistics in Göttingen (LinG1) and at UMD's S-Lab, for helpful comments.

I will also discuss potential challenges for the account, including the proper treatment of languages that appear to violate the AAE.

## 2. An overview of the AAE

The AAE (Rizzi 1990a, Woolford 1999, *a.o.*) can be thought of as a typological gap in the joint distribution of anaphors, on the one hand, and noun phrases being targeted for morpho-phonologically overt  $\varphi$ -agreement, on the other. It can be variously described in ways that imply causation (e.g. an avoidance of  $\varphi$ -agreement with anaphors, or an avoidance of anaphors in positions targeted for  $\varphi$ -agreement); but without imputing causation quite yet, a first approximation of the effect is given in (1):

- (1) ANAPHOR AGREEMENT EFFECT (AAE), *first version*

$$* \text{H}^0 \dots \text{DP}_{\text{ANAPH}}$$

$\varphi\text{-agr.}$

As we will see, some refinement of (1) will ultimately be necessary. First, however, let us see some representative cases illustrating the effect.

As noted by Thráinsson (1976, 1979), Maling (1984), and others, Icelandic has a long-distance anaphor, *sig/sér/sín* (ACC/DAT/GEN, respectively; I will gloss this element as “SIG” henceforth).<sup>1</sup> As such, SIG can take as its antecedent the subject of a superordinate clause:<sup>2</sup>

- (2) Jón<sub>i</sub> heldur [ að þú hatir sig<sub>i</sub> ] Icelandic  
 John believes that you hate.SBJV.2sg SIG.ACC  
 ‘John<sub>i</sub> believes that you hate him<sub>i</sub>.’ [Thráinsson 2007:467]

In (2), the embedded verb *hatir* (“hate.SBJV.2sg”) shows overt  $\varphi$ -agreement with its local subject, *þú* (“you(sg.)”). The anaphor SIG, on the other hand, is not in a position that has been targeted for  $\varphi$ -agreement. The occurrence of SIG in (2) is therefore in accordance with the condition in (1).

Contrast this with (3): here, the embedded verb *líki* (“likes.SBJV.3sg”) is a quirky-subject verb, assigning dative case to *mér* (ME.DAT). Consequently, the object of this verb bears nominative case, rather than the usual accusative (see Marantz 1991, Yip, Maling & Jackendoff 1987, Zaenen, Maling & Thráinsson 1985, *a.o.*, for details). In Icelandic, finite agreement tracks case, not grammatical function (Sigurðsson 1993, *et seq.*).<sup>3</sup> Consequently, in the embedded clause in (3), it is

<sup>1</sup>It has been argued, by Thráinsson (1979), Maling (1984), and others, that SIG is in fact *logophoric* in nature (cf. Clements 1975 and subsequent literature), and as such, its distribution is subject to discourse-oriented restrictions. While this is certainly an interesting facet of SIG’s distribution, it is ultimately at right angles to our present concerns. That is because it turns out that there is no discourse configuration, however unorthodox, that allows SIG to occur in a position violating (1). The latter is clearly not a discourse-related fact, and it is this facet of SIG’s distribution that is of interest here. This does illustrate, of course, that (1) should be taken as a necessary-but-not-sufficient condition on the distribution of anaphors—exactly as one would expect. See Maling (1984), Sells (1987), Thráinsson (1991, 1992), among others, for further discussion.

<sup>2</sup>Glossing conventions used in this paper: 1: first person; 2: second person; 3: third person;  $\sqrt{\quad}$ : lexical root; ABS: absolutive; ACC: accusative; ANAPH: anaphor; ART: article; CL: clitic; COMP: complementizer; DAT: dative; ERG: ergative; GEN: genitive; IMPF: imperfective; INTS: intensive/emphatic; LOC: locative; NOM: nominative; pl: plural; PRON: pronoun; SBJV: subjunctive; sg: singular.

<sup>3</sup>As is probably the case in all languages (Bobaljik 2008).

the object, not the subject, that controls the form of the finite verb. In accordance with the condition in (1), SIG cannot occur in this environment. (Notice that the attempted structural configuration in (3) mirrors that of (2), i.e., an embedded object SIG bound by the matrix subject.)

- (3) Sigga<sub>i</sub> telur [ að mér líki hún<sub>i</sub>/\*sig<sub>i</sub> ]  
 Sigga thinks that me.DAT likes.SBJV.3sg she.NOM/\*SIG  
 ‘Sigga<sub>i</sub> thinks that I like her<sub>i</sub>.’ [Maling 1984:217]

Finally, note that this is not a general ban on SIG occurring as the argument of a quirky-case verb. Ex. (4) also features a quirky-subject verb, *vantaði* (“lacked.SBJV.3sg”), again standing in a  $\varphi$ -agreement relation with its direct object. Here, however, SIG is in the subject position, and thus marked with (quirky) accusative case. Consequently, SIG is not targeted for  $\varphi$ -agreement, and therefore—once again in accordance with (1)—its occurrence is licit.

- (4) Hún<sub>i</sub> sagði [ að sig<sub>i</sub> vantaði peninga ].  
 she.NOM said that SIG.ACC lacked.SBJV.3sg money  
 ‘She<sub>i</sub> said that she<sub>i</sub> lacked money.’ [Maling 1984:216]

As noted throughout, finite  $\varphi$ -agreement in Icelandic tracks nominative case. It is therefore possible to construe the behavior in (2–4) not as a ban on SIG in agreeing positions, but rather, as a ban on SIG in *nominative* positions. (Or, perhaps, as a missing nominative cell in the case paradigm of SIG.) Internal to Icelandic, it might be hard to disentangle these two options. But as Woolford (1999) points out, the latter, case-based characterization fails to generalize cross-linguistically. That is because it is common for languages with overt case-marking but no overt  $\varphi$ -agreement to *allow* nominative anaphors. Woolford illustrates this using examples from Korean and Japanese:

- (5) sensei<sub>i</sub>-ni(-wa) zibun<sub>i</sub>-ga wakar-ani-i (Japanese)  
 teacher-DAT(-TOPIC) self-NOM understand-NEG-PRES  
 ‘The teacher<sub>i</sub> does not understand themselves<sub>i</sub>.’ [Shibatani 1977:800, via Woolford 1999:263]

A case-theoretic characterization of the AAE can explain facts like (2–4), but leaves facts like (5) unexplained. An agreement-based characterization, as in (1), can capture the Icelandic and Japanese facts alike.

This is not the end of the story, however, when it comes to the AAE. As Woolford already notes, there are exceptions to the AAE *as formulated in (1)*. In some languages, anaphors occur in positions targeted for agreement, but the result is so-called “default agreement”: there is no co-variation in  $\varphi$ -features between the anaphor (or its antecedent) and the finite verb. Albanian is a case in point:

- (6) Vetja më dhimset. (Albanian)  
 self.NOM CL.1sg.DAT feel.sorry.for.3sg.PRES.NONACTIVE  
 ‘I feel sorry for myself.’ [Woolford 1999:271, citing Everaert 1991 and Hubbard 1985:91]

Being the nominative Theme of a dative-Experiencer verb, the anaphor *vetja* in (6) occurs in a position that would normally control finite agreement. Nevertheless, the verb fails to register the  $\varphi$ -features of the antecedent (and, presumably, the bindee), in this case 1sg, and instead surfaces with invariant 3sg morphology.

A superficially distinct case involves languages where it is not the anaphor in its entirety that co-varies in  $\varphi$ -features with the antecedent, but a proper subpart of the anaphor. Basque instantiates

this pattern. In Basque, reflexives are formed using the structure in (7), where: PRON.GEN is a (strong or weak) possessive pronoun; N is a designated noun, *buru*, meaning “head” (in line with a cross-linguistically common strategy of forming reflexives from possessed body-part nouns); and D is the determiner/article. (See *Artiagoitia 2003* for a detailed discussion.)

(7) structure of Basque reflexives: [PRON.GEN N D]

Now, crucially, the expression in (7) behaves as far as the external syntax (e.g. when probed by heads outside of the brackets) as any instance of a possessed (singular) common noun would. In particular, it triggers 3sg agreement on the finite verb:

(8)  $\langle pro_{2pl.ERG} \rangle$  [ *zuen buru-a* ] *saldu d-φ-u-zue* (Basque)  
 2pl(weak).GEN head-ART<sub>sg</sub>(ABS) sold 3.ABS-SG.ABS-√-2pl.ERG  
 ‘Y’ all have given yourselves away’ [*Artiagoitia 2003:620*]

In (8), the binder is 2pl (as can be seen in the ergative agreement morphology on the sentence-final finite auxiliary). And, accordingly, the features on the possessive pronoun within the (bracketed) reflexive are 2pl, as well. However, the entire reflexive *zuen buru-a* (“2pl(weak).GEN head-ART<sub>sg</sub>(ABS)”) triggers the agreement morphology one would expect from a possessed instance of *buru* (“head”), namely 3sg absolutive agreement (again, visible in the form of the finite auxiliary).

To foreshadow the analysis argued for in this paper (specifically, in section 8), I will argue that Basque realizes transparently what is also going on, albeit less transparently, in Albanian.

But for our present purposes, we must ask ourselves—as *Woolford 1999* does—what the fate of examples like (6) and (8) is with respect to the AAE, as formulated in (1):

(1) ANAPHOR AGREEMENT EFFECT (AAE), *first version*

$$* \underbrace{H^0 \dots DP_{ANAPH}}_{\varphi\text{-agr.}}$$

Because (1) places no restrictions the results of the  $\varphi$ -agreement relation in question, a case where  $H^0$  shows up with invariable 3sg agreement morphology (regardless of the  $\varphi$ -features of the antecedent/bindee) still meets the structural description of the constraint. Therefore, cases like (6) and (8) violate the constraint as formulated above.

Let us therefore propose the following alternative formulation:

(9) ANAPHOR AGREEMENT EFFECT (AAE), *revised version*

$$* \underbrace{H^0 \dots DP_{ANAPH}}_{\mathcal{R}}, \text{ where } \mathcal{R} \text{ is a nontrivial } \varphi\text{-agreement relation}$$

(10) A  $\varphi$ -agreement relation between  $\alpha$  and  $\beta$  is nontrivial if there are at least two sets of feature-values,  $F$  and  $F'$ , such that  $\alpha$  bears  $F$  **iff**  $\beta$  bears  $F$ , and  $\alpha$  bears  $F'$  **iff**  $\beta$  bears  $F'$ .

Under this revised definition, cases like the Albanian (6) and the Basque (8) no longer constitute violations of the AAE. That is because, when the agreement target is an anaphoric DP, there is only one possible set of feature-values involved in the  $\varphi$ -agreement relation (3sg).

This leaves only a handful of languages, such as Tamil, Standard Gujarati, Archi, and Ingush (Murugesan 2018, *in prep.*), which appear to genuinely violate the AAE (even under this revised definition). I return to this issue, and the proper treatment of these cases, in section 9.

The final point I would like to make in this section concerns not an exception to the AAE, but an observation regarding the scope of its application. (This, too, was already noted by Woolford 1999:283.) Consider once more the distinction between the Icelandic (3) and the Japanese (5), both repeated here:

- (3) Sigga<sub>i</sub> telur [ að mér líki hún<sub>i</sub>/\*sig<sub>i</sub> ]  
 Sigga thinks that me.DAT likes.SBJV.3sg she.NOM/\*SIG  
 ‘Sigga<sub>i</sub> thinks that I like her<sub>i</sub>.’ [Maling 1984:217]
- (5) sensei<sub>i</sub>-ni(-wa) zibun<sub>i</sub>-ga wakar-ani-i (Japanese)  
 teacher-DAT(-TOPIC) self-NOM understand-NEG-PRES  
 ‘The teacher<sub>i</sub> does not understand themselves<sub>i</sub>.’ [Shibatani 1977:800, via Woolford 1999:263]

In order to maintain our account of (3) without incorrectly ruling out (5), it is necessary for at least one of the following conditions to hold: (i) Korean, Japanese, and all other languages like them do not have an abstract, morpho-phonologically unexpressed version of the agreement relation in (3); or (ii) the AAE is a morphological constraint, rather than syntactic one. To see why, suppose both (i) and (ii) were false. What would distinguish the nominative anaphor in (3) from its counterpart in (5)? It cannot be the presence of overt agreement with the anaphor in (3), since syntax is modularly encapsulated from morpho-phonology, and thus cannot “query” the phonological content of a term to see whether agreement will or will not be overtly expounded. And since Japanese is allowed to have a syntactically extant but morpho-phonologically unexpounded version of the agreement relation in (3), there would be no explanation for why the anaphor in (5) is licit while the one in (3) is not.

Now let us ask: is (ii)—the idea that the AAE is morphological in nature—a tenable position? I argue that it is not. The reason is as follows. The relation that the AAE prohibits is (nontrivial)  $\varphi$ -agreement. Thus, no module of grammar could detect and block violations of the AAE unless it was able to hold both the probe ( $H^0$ ) and the anaphor ( $DP_{ANAPH}$ ) within a single domain of computation, to examine whether the former has entered into (nontrivial)  $\varphi$ -agreement with the latter. Crucially, however,  $\varphi$ -agreement is unbounded: there is no upper bound on the amount of structure or linear distance that this relation can span.<sup>4</sup> For a particularly vivid demonstration of this fact, I refer the reader to Keine (2017), who shows that long-distance agreement in Hindi can span across an unbounded number of transitive verb phrases. Consequently, that which the AAE must “guard against” (i.e., prohibit from happening) involves an unbounded relation. And if ‘morphology’ differs from ‘syntax’ in any contentful way, surely unbounded dependencies are the purview of the latter, not the former. Therefore, whatever underpins the AAE must be operative in syntax, not in morphology. Put another way: for the AAE to be “morphological” in nature, it would have to be the case that there is no upper bound on the size of a “morphological domain”—a position that I take to be self-evidently indefensible.

---

<sup>4</sup>That is not to say that there is no structure that would block  $\varphi$ -agreement, of course. Much like *wh*-movement is unbounded but is stopped by certain structures (syntactic islands), so too is  $\varphi$ -agreement stopped by certain structures (finite CP boundaries). None of this changes the fact that both relations are unbounded, in the sense that there is no limit on the number of phrases they can span.

If at least one of (i)–(ii) must be true, and (ii) is false, it follows that (i) is necessarily true: languages like Japanese must lack any finite  $\varphi$ -agreement relation whatsoever. In particular, they do not have a morpho-phonologically unexpressed counterpart of the  $\varphi$ -agreement relation that one finds overtly expressed in, say, Icelandic. This is a welcome result, given that it converges with independent arguments adduced from the distribution and nature of Person Case Constraint (PCC) effects, supporting the same conclusion (Preminger 2019).

### 3. Reductionist theories and the AAE

The AAE seems to suggest that anaphoric binding and  $\varphi$ -agreement interact with each other quite closely. That, in turn, has been taken as support for the view that anaphoric binding *reduces* to  $\varphi$ -agreement—see, e.g., Reuland (2011:261–262). (For other reductionist proposals, see Heinat 2008, Kratzer 2009, Rooryck & Vanden Wyngaerd 2011, among others.) On this view, binding of an anaphor consists in valuing the  $\varphi$ -features on the (initially  $\varphi$ -deficient) anaphor using the values found on its binder.

Now, there is plenty of evidence unrelated to the AAE suggesting that reductionist approaches are off track. While it is not the purpose of the present paper to rehash these arguments, I will briefly mention two of them here. One source of evidence concerns *directionality*. It is well established at this point that  $\varphi$ -agreement (*viz.* finite agreement between a predicate and one or more of its nominal arguments) transmits values upward in the structure: from a c-commanded goal to a c-commanding probe (see Keine & Dash 2018, Polinsky & Preminger 2017, Preminger 2013, Preminger & Polinsky 2015, *a.o.*).<sup>5</sup> Anaphoric binding, in contrast, appears to operate in the other direction, transmitting values from a c-commanding antecedent to a c-commanded bindee. Thus, the idea that anaphoric binding is underpinned by  $\varphi$ -agreement is antithetical to the apparent structural properties of the two relations.<sup>6</sup>

Another source of evidence against reductionist theories has already been mentioned here, and concerns the ban on morpho-phonologically null agreement (Preminger 2019). As noted earlier, such a ban is required if we are to maintain the AAE, in light of languages that lack overt  $\varphi$ -agreement and appear to allow anaphors in any argument position. Otherwise, we lose our explanation for why nominative anaphors are banned in Icelandic, for example, but allowed in Japanese. But, crucially, it is a truism that anaphoric binding exists even in languages that lack overt  $\varphi$ -agreement (and, by hypothesis, lack syntactic  $\varphi$ -agreement as well). Japanese, after all, does have anaphors. Thus, the idea that anaphoric binding is underpinned by  $\varphi$ -agreement is antithetical to the ban on null agreement, which is required for an account of the AAE in the first place.<sup>7</sup>

---

<sup>5</sup>See also Bjorkman & Zeilstra (2018), who, by making explicit the vast and problematic array of assumptions that would be required to support a theory in which  $\varphi$ -feature values were transmitted downward in the structure, inadvertently provide a fairly strong argument against such a theory. For example, Bjorkman & Zeilstra’s theory must assume that all  $\varphi$ -agreement include a *checking* component alongside *valuation*—an assumption that is known to be false (Preminger 2011a, 2014).

<sup>6</sup>Not *every* reductionist approach runs afoul of this directionality consideration. The aforementioned work by Reuland (2011), for example, does not—as it involves only agreement relations that are well-behaved from a directionality perspective (i.e., relations where the recipient of values c-commands the supplier of values).

<sup>7</sup>Sandhya Sundaresan (p.c.) points out that it is technically possible to maintain the ban on morpho-phonologically null finite  $\varphi$ -agreement, while assuming that  $\varphi$ -agreement between a DP and another DP (*viz.* binding) is not subject to the same ban. While I agree that this is technically possible, it seems ad hoc, and in any event, it decreases the amount of ‘reduction’ that reductionist theories can actually claim to have achieved, thus lessening their appeal in the first place.

All that being said, the focus of the present paper is different. I aim to show that once we take seriously what the AAE entails for reductionist theories, we see that far from providing support for such theories, it actually provides strong evidence against them.

Let us start by adopting some utility definitions:

- (11) Let  $F_\varphi$  be the formal process or relation that values the  $\varphi$ -features on a functional head using the  $\varphi$ -feature values found on one or more DPs (Chomsky's 2001 Agree, or Preminger's 2014 FIND $_\varphi$ ).
- (12) Let DP[A] be the anaphor, and let  $\varphi(\text{DP}[A])$  be the  $\varphi$ -features borne by DP[A] at a given stage of the derivation.
- (13) Let DP[B] be a potential binder, and let  $\varphi(\text{DP}[B])$  be the  $\varphi$ -features borne by DP[B] at a given stage of the derivation.
- (14) Let  $H^0$  be the  $\varphi$ -probe that putatively agrees with DP[A] (i.e., the  $\varphi$ -probe that, if DP[A] were replaced with a non-anaphoric nominal  $\alpha$ , would enter into  $F_\varphi$  with  $\alpha$ ).

Given these definitions, we can state the reductionist hypothesis quite straightforwardly as follows:

- (15) THE REDUCTIONIST HYPOTHESIS  
Binding of DP[A] by DP[B] arises via  $F_\varphi(\text{DP}[B], \text{DP}[A])$ .

For a reductionist theory to capture the AAE, it must be the case that at whatever point in the derivation DP[A] is targeted for  $\varphi$ -agreement, it does not bear the  $\varphi$ -features of DP[B]. In other words, such a theory requires that  $\varphi(\text{DP}[A]) \neq \varphi(\text{DP}[B])$  when  $F_\varphi(H^0, \text{DP}[A])$  applies (and possibly afterwards, as well). I will refer to this hypothesis as the  **$\varphi$ -deficiency of anaphors**.

There are two variants of the  $\varphi$ -deficiency of anaphors to consider. One is that anaphors are categorically unable to carry  $\varphi$ -features (i.e.,  $\varphi(\text{DP}[A]) \neq \varphi(\text{DP}[B])$  throughout the course of the derivation). While logically possible, it seems clear that we can dismiss this possibility outright, as it is the case that anaphors in many languages show the full range of  $\varphi$ -feature distinctions available to other DPs in the language. (One caveat to this conclusion concerns the possibility that anaphors carry  $\varphi$ -features, but these are categorically inaccessible from the outside; cf. the Basque example in (8), and the surrounding discussion. For reasons I discuss in section 6, this does not salvage the reductionist approach.)

At this juncture, it is worth noting that the inference just drawn (against theories where anaphors are categorically  $\varphi$ -deficient)—as well as other inferences I will draw in this paper—only goes through if ‘anaphor’ is a natural class. Otherwise, it would be logically unsound to carry over conclusions from one ‘anaphor’ to another. And there is indeed a growing body of work suggesting that there are, in fact, different types of anaphors (see, e.g., Sundaesan to appear). That said, different (sub-)types do not yet imply that ‘anaphor’ is not a natural class (compare, for example, ‘plosive’—which has different sub-types, but is nevertheless a natural class). In fact, if ‘anaphor’ is not a natural class, then the AAE is (definitionally) an epiphenomenon. Since I find it unlikely to be an epiphenomenon, I will assume, as a working hypothesis, that ‘anaphor’ does indeed carve out a natural class, one whose members can be pre-theoretically identified as follows:

- (16) ANAPHOR: WORKING DEFINITION  
Those nominal elements whose reference is necessarily dependent on another nominal in the sentence.

Thus, even if it is true that *some* anaphors are categorically unable to carry (some)  $\varphi$ -features, this fails as an explanation of the AAE, inasmuch as the AAE applies to the class ‘anaphors’, and that class undeniably includes expressions that show the full range of  $\varphi$ -distinctions available in their respective languages.

This brings us to the second variant of the hypothesis concerning the  $\varphi$ -deficiency of anaphors. On this view, anaphors lack  $\varphi$ -feature values *at the derivational stage* at which they are targeted for  $\varphi$ -agreement. Given the failure of the other, categorical version of  $\varphi$ -deficiency, reductionist theories are forced into this timing-based approach to the AAE. Therefore, the next section is devoted to examining in some detail what a timing-based approach to the AAE would have to look like under a reductionist theory.

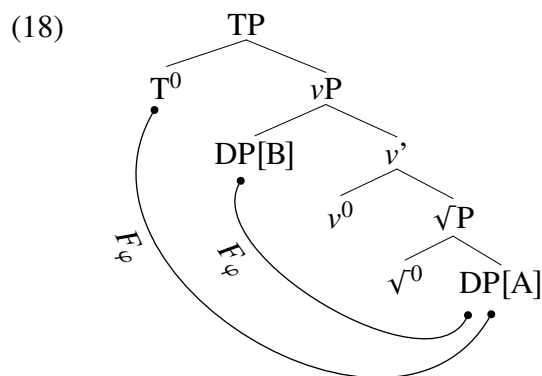
#### 4. Timing and reductionism

As noted in the previous section, a timing-based approach to the AAE requires the following to hold:

- (17) NECESSARY CONDITION FOR TIMING-BASED ACCOUNTS OF THE AAE  
 $\varphi(\text{DP}[A]) \neq \varphi(\text{DP}[B])$  when  $F_\varphi(\text{H}^0, \text{DP}[A])$  applies

In this section, I will show that (17) is in fact incompatible with the reductionist hypothesis (given in (15), above).

For expository purposes, let us consider a fairly straightforward anaphoric binding scenario: a transitive clause where the subject binds a reflexive in direct object position. Let us assume furthermore that the language in question is an ergative-absolutive language, where the absolutive object is targeted for agreement by  $T^0$  (a “HIGH-ABS” language in Coon, Mateo Pedro & Preminger’s 2014 parlance, or an “ABS=NOM” one in Legate’s 2008 parlance). Languages that fit this bill include Georgian, Basque, Q’anjob’al, Kaqchikel, and many others.<sup>8</sup> It is also, importantly, the structure that arises when one attempts to bind a direct-object SIG in a clause whose subject bears quirky case (which, as the AAE predicts, is also illicit). The scenario in question is schematized in (18). (Other local binding scenarios will be discussed later in this section.)



Because we are evaluating the reductionist hypothesis, both relations in (18) are indicated as relying on  $F_\varphi$ . Note that (18) also assumes that the copy of DP[B] relevant to anaphoric binding of DP[A]

<sup>8</sup>Interestingly, Basque does not fit Legate’s (2008) and Coon, Mateo Pedro & Preminger’s (2014) case-theoretic diagnostics for a “HIGH-ABS”/“ABS=NOM” language. Nevertheless, it is clear that the locus of absolutive agreement in Basque is  $T^0$  (see Arregi & Nevins 2012 for extensive evidence of this).

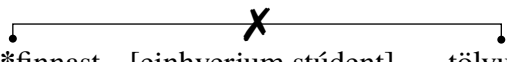


is the one in [Spec,vP] (i.e., its base position), and not some higher position that DP[B] might later move into. The reasons for this are as follows. First, we know that A-bar movement cannot create new antecedents for binding, and therefore subsequent A-bar movement of DP[B] would be irrelevant here. As for A-movement: we know that even if a DP has undergone A-movement, its lower copy (or copies) can still bind. Consider (19):

- (19) a. The children seem to her<sub>j</sub>/\*<sub>i</sub> to have amused Mary<sub>i</sub>.  
 b. The children seem to me to have amused each other.

As is well known and demonstrated in (19a), the Experiencer argument of *seem* can bind into its infinitival complement clause. Given this, the Experiencer *me* in (19b) is a potential binder, meaning that the copy of *the children* relevant to local binding of *each other* cannot be the one in the matrix clause. The binder must be a lower copy of *the children*, presumably one located within the embedded infinitive itself. This shows that lower copies in an A-chain can be binders, in turn supporting the decision to treat the base-generated copy of DP[B] in (18) as the relevant one for binding of DP[A].

Another aspect of (18) that merits discussion is the status of the relation between T<sup>0</sup> and DP[A] vis-à-vis minimality (Abels 2012, Boeckx & Jeong 2004, Frampton 1991, Rizzi 1990b, 2001, Starke 2001, *a.o.*). Specifically, one might have expected DP[B] to intervene in the relation between T<sup>0</sup> and the (absolutive) DP[A]. However, as shown by Anagnostopoulou (2003), Holmberg & Hróarsdóttir (2003), and others, traces (or lower copies) of A-movement do not intervene in  $\varphi$ -agreement:

- (20) það finnst/\*finnast [einhverjum stúdent] tölvurnar ljótar. (Icelandic)  
 EXPL find.SG/\*find.PL some student.DAT computers.the.NOM ugly  
 ‘Some student finds the computers ugly.’
- 

- (21) [Einhverjum stúdent]<sub>1</sub> finnast t<sub>1</sub> tölvurnar ljótar  
 some student.DAT find.PL computers.the.NOM ugly  
 ‘Some student finds the computers ugly.’ [Holmberg & Hróarsdóttir 2003:999–1000]

This presents an obstacle in and of itself for theories that aim to reduce binding to  $\varphi$ -agreement, given that lower copies of A-chains *do* intervene for the purposes of anaphoric binding—in (22), for example, the lower copy of *the children* prevents the Experiencer *me* from locally binding the embedded reflexive:

- (22) \* The children seem to me to have amused myself

Setting aside, however, the question of why binding and  $\varphi$ -agreement differ in this manner, this property of  $\varphi$ -agreement is sufficient to solve the putative minimality violation in (18): A-movement of DP[B] past T<sup>0</sup> would ameliorate any minimality problem with the relation between T<sup>0</sup> and DP[A]. (See Preminger 2014:77ff. for an analysis along these very lines of absolutive agreement in transitives in the K’ichean languages.)

Having established the basic properties of the scenario in (18), two conclusions now follow. First, there is no possible stage in the derivation of (18) where the  $\varphi$ -probe, T<sup>0</sup>, has already entered the derivation, but DP[A] has not yet been bound by DP[B]. Thus, given the reductionist hypothesis repeated below, T<sup>0</sup>, *qua*  $\varphi$ -probe, would come upon a fully valued DP[A] as its goal. This would

result in full-fledged (=nontrivial)  $\varphi$ -agreement with the anaphor—exactly the opposite of what an account of the AAE should deliver.

(15) THE REDUCTIONIST HYPOTHESIS

Binding of DP[A] by DP[B] arises via  $F_\varphi(\text{DP[B]}, \text{DP[A]})$ .

Second, an appeal to phases and/or spellout domains cannot figure in an account of the AAE. If phases or spellout domains were to prevent  $T^0$  in (18) from accessing DP[A] (or prevent  $T^0$  from doing so prior to the spellout of the structure that contains it), the same would be true of  $T^0$  and the direct object in sentences with *non*-anaphoric arguments. But that is simply not the case in the languages in question (Georgian, Basque, etc.): these languages all have (nontrivial)  $\varphi$ -agreement between  $T^0$  and the direct object. Whether this is because  $vP$  is never a phase / spellout domain (Keine 2017), or because the object moves out of the  $vP$  phase—it does not matter. Whatever facilitates such  $\varphi$ -agreement in normal transitive clauses will do so when the object is anaphoric, as well.

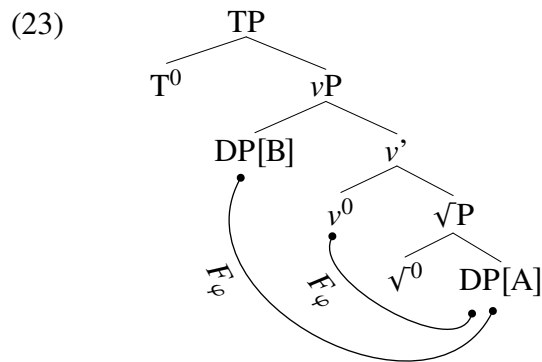
Crucially, since languages of this type do obey the AAE (see the discussion surrounding (8), above), the arguments presented so far suffice to show that there can be no general account of the AAE that is both timing-based and reductionist. Any such account could, at most, apply to languages whose clause-structure is *not* as in (18), and so the fact that (18)-type languages also obey the AAE would essentially amount to an accident. (In section 8, I will present a general account of the AAE that is neither reductionist nor timing-based, in which all AAE effects emerge in a uniform fashion.)

We have thus delivered that which was promised at the outset of this section: a demonstration that (17)—a necessary condition for timing-based accounts of the AAE—is incompatible with reductionism.

(17) NECESSARY CONDITION FOR TIMING-BASED ACCOUNTS OF THE AAE

$\varphi(\text{DP[A]}) \neq \varphi(\text{DP[B]})$  when  $F_\varphi(H^0, \text{DP[A]})$  applies

Nevertheless, the reader may wonder how a reductionist, timing-based account fares in a language with different clause structure; and the answer is: not much better. Let us now consider a language in which the operative  $\varphi$ -probe is  $v^0$ , rather than  $T^0$ . (In other words, a “LOW-ABS”/“ABS=DEF” language.) The clause structure for such a language is given in (23):

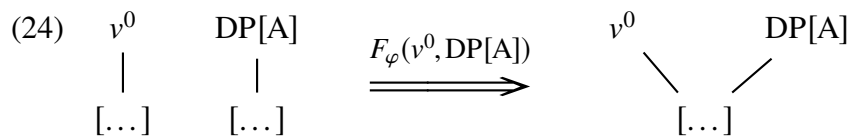


Let us note one more consequence of the reductionist view, not mentioned so far. Because the reductionist view takes  $F_\varphi$  to underpin anaphoric binding, it must be the case that  $F_\varphi$  can apply

countercyclically, changing (valuing) the features on DP[A] (the anaphor) long after DP[A] is no longer near the root of the syntactic tree. At most,  $F_\varphi$  could be subject to a weaker kind of cyclicity, one that demands that at least *one* of its operands (the recipient of values, or the supplier) be near the root of the tree.

Now, importantly, reductionism entails that *at some point* before the end of the derivation, DP[A] would have its  $\varphi$ -features valued by the binder, DP[B]. But given that  $F_\varphi$  is countercyclic,  $F_\varphi(v^0, DP[A])$  could apply *at that point*. This would yield full, nontrivial  $\varphi$ -agreement with the anaphor, in violation of the AAE. Adding the assumption that  $F_\varphi$  is weakly cyclic, in that one of its operands must be near the root of the tree, would change nothing. That is because it is well established that the head of a projection (even a phasal one) can trigger syntactic operations even after a specifier has been merged to it. Richards (2001), for example, has shown that even after  $C^0$  has attracted one phrase to its specifier, features on  $C^0$  can continue to drive further operations at the root.<sup>9</sup> Thus,  $v^0$  (like  $C^0$ ) can probe even after something (in this case, DP[B]) has already been merged in its specifier, and before completion of the  $vP$  phase. The result would again be full, nontrivial  $\varphi$ -agreement with DP[A], exactly the opposite of what an account of the AAE must deliver.

In fact, things get even worse for the reductionist view. If  $F_\varphi$  is supposed to be the relation underpinning  $\varphi$ -agreement, then there are a few more things we know about it. In particular, we know that  $F_\varphi$  does not “check features” (Preminger 2011a, 2014). Nor does it “copy” feature-values from one element to another; instead,  $F_\varphi$  creates a feature-sharing structure (Andrews 1971, Frampton & Gutmann 2000, 2006, Gazdar et al. 1985, Pesetsky & Torrego 2007, Pollard & Sag 1994, Preminger 2017, *a.o.*). The output of (a successful application of)  $F_\varphi$  involves multiple syntactic nodes being linked to a single feature-structure:



This means that *regardless* of the relative timing of anaphoric binding and  $\varphi$ -agreement, if the former values the  $\varphi$ -features of DP[A], then the result will ipso facto be reflected on the  $\varphi$ -agreement probe. Phases and/or spellout domains provide no relief here, for the reasons outlined above—Richards’ (2001) work has already shown that the phase is not complete until the needs of the head have been addressed, that at least some needs of the head can be addressed *after* something has been merged in the specifier position of that head. Therefore,  $\varphi$ -probing by  $v^0$  will occur in the same phase / spellout domain as the binding of DP[A] by DP[B].

In sum, reductionist timing-based approaches to the AAE do not fare much better in languages where  $v^0$  agrees with the object than they do in languages where  $T^0$  does so.

Finally, Sandhya Sundaesan (p.c.) has suggested an alternative way of implementing a timing-based approach, where the issue consists not in cyclicity/countercyclicity per se, but in an inability to revise a previously-established agreement relation. On this view, the AAE arises because the  $\varphi$ -probe establishes an agreement relation with DP[A] first, and the results of this relation cannot

<sup>9</sup>This is perhaps unsurprising given Bare Phrase Structure assumptions (Chomsky 1994), which take all projection levels to be nothing more than additional occurrences of the head. On this view, the object at the root of the tree at any given stage in the derivation *is* (an occurrence of) the head of the relevant projection.

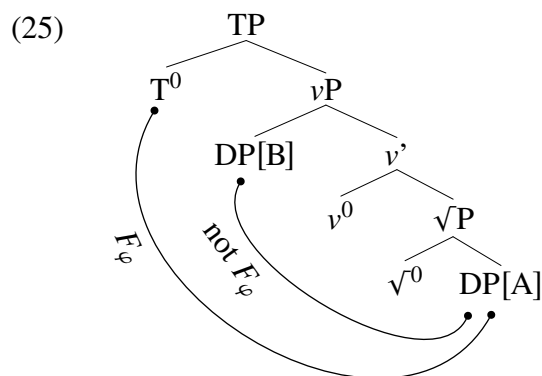
be revised later in the derivation (*viz.* after DP[B] has entered the derivation and valued the  $\varphi$ -features on DP[A]). However, this cannot provide a general explanation of the AAE. For one thing, given (24), there is literally no way to alter the features of DP[A] without also altering the features on the  $\varphi$ -probe; they are literally one and the same object. But more importantly, this could not possibly serve as an explanation for how the AAE arises in cases like (18), where the  $\varphi$ -probe ( $T^0$ ) is merged *after* the binder (DP[B]).

At this juncture, the reader may notice that I have not said a word so far about the AAE in subject-agreement languages, where  $T^0$  targets the subject for morpho-phonologically overt  $\varphi$ -agreement. I conclude this section with a few comments on such languages.

It should be clear that even if a reductionist, timing-based account of the AAE worked flawlessly for such languages, it would have no hope of achieving true generality, in light of its failures as it regards ergative-absolutive languages (as in (18) and (23), above). The reason I have avoided talking about subject-agreement languages specifically is that one can be far less certain of the role of phases there. If the anaphor, DP[A], is a subject in [Spec,TP], it may very well be the case that there will always be a phase boundary (*viz.* CP) separating it from its antecedent. This would make it impossible, internal to subject-agreement languages, to disentangle the role of phases from the role of other factors in giving rise to the AAE. Examining direct-object anaphors in ergative-absolutive languages allowed us to argue directly against phases playing a role in the AAE. In subject-agreement languages, on the other hand, the AAE may be overdetermined—caused both by the dynamics of phases, and by whatever it is that causes the AAE in ergative-absolutive languages. (See section 8 for a positive proposal concerning the latter.)

## 5. Timing without reductionism

The previous section made the case that the reductionist hypothesis, in (15), cannot be maintained together with timing-based approaches to the AAE. We can, however, ask a separate question, which is whether timing-based approaches can be maintained *without* reductionism. In other words, if we accept that anaphoric binding and  $\varphi$ -agreement are formally separate mechanisms, can a timing-based approach be salvaged? The answer, it seems to me, is once again no, due again to “HIGH-ABS”/“ABS=NOM” languages, as in (25):



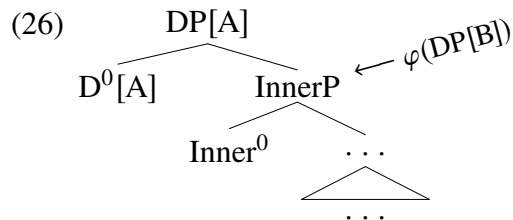
In such languages, DP[B] (the binder) enters the derivation before  $T^0$  (the  $\varphi$ -probe). It therefore does not matter whether binding of DP[A] by DP[B] arises by way of  $F_\phi$ , or by some other operation; DP[A] will have been bound by the time  $T^0$  probes it. Consequently, the only way to derive the

AAE here would be to stipulate that DP[A] never acquires any  $\varphi$ -features whatsoever. This was the approach referred to in section 3 as the *categorical  $\varphi$ -deficiency of anaphors*—and it does not accord with the facts, for the reasons discussed there. (See section 6 for further discussion.)

In any event, an approach based on a categorical  $\varphi$ -deficiency of anaphors would not be a timing-based approach, in the first place. Therefore, the overall conclusion is that timing-based approaches have no hope insofar as our goal is to furnish a general account of the AAE cross-linguistically.

## 6. Encapsulation and reductionism

Given that we cannot maintain a general, timing-based account of the AAE, we could at this juncture abandon any hope of a fully general account of the AAE. But let us consider an alternative. Suppose for a moment that the anaphor, DP[A], *did* enter into an  $F_\varphi$  relation with the binder, DP[B]—as per reductionism—but that the constituent whose features were consequently valued was not DP[A] in its entirety. Instead, the bearer of valued  $\varphi$ -features on the anaphor side was a proper *subpart* of DP[A], which I will label InnerP for now:



Now suppose we add the assumption that the  $D^0$ [A]/DP[A] layer halts probing for  $\varphi$ -features when probed from the outside. This could be because DP[A] is a locality boundary (e.g. a phase), or it could be because DP[A] bears its own set of  $\varphi$ -features. (The latter seems less likely in this case, since it is hard to imagine what  $\varphi$ -features an anaphor would bear other than those which match its binder.) Either way, a probe entering into  $\varphi$ -agreement with DP[A] would now fail to access the valued  $\varphi$ -features corresponding to DP[B], since they are located on InnerP and are inaccessible from the outside. Let us refer to this type of approach as “encapsulation” (because the features matching the binder are encapsulated in additional structure, within the anaphor).

Recall, now, the reductionist tenet: anaphoric binding *consists in*  $\varphi$ -feature agreement. Taken seriously, this means that not only is InnerP what bears the anaphor’s valued  $\varphi$ -features in (26); InnerP must also be the constituent that carries the *binding index*. (Otherwise, one cannot seriously claim any reduction of binding to  $\varphi$ -agreement.) It then follows that the binding index carried by anaphors should not c-command out of DP[A] (since the c-command domain of InnerP is, at most, the entirety of DP[A] and nothing more.) Happily, this is a testable prediction. Consider the following examples:

(27) a. John<sub>i</sub> expects Mary to outdo him<sub>i/k</sub>.

b. John<sub>i</sub> expects himself<sub>i</sub> to outdo him<sub>k/\*i</sub>.

[Norvin Richards, *p.c.*]

Notice first that (27a) does not give rise to a disjoint-reference effect between *John* and the pronoun *him*, meaning *John* is too far away from the pronoun, structurally speaking, for the two to enter into a binding relation. Crucially, this means that the cause of the disjoint-reference effect observed in (27b) must be the anaphor, *himself*. But this could only be the case if the binding

index resided on the outermost projection of the anaphor, in contradiction to the predictions of the reductionist hypothesis when coupled with (26).

It is possible, of course, that English happens not to abide by encapsulation (i.e., it does not accord with the schema in (26)). But recall that what we are in search of here is a *general* theory of the AAE; insofar as encapsulation offers any hope of realizing such a theory, it too would have to be general. The evidence in (27), however, only speaks against encapsulation *if coupled with reductionism*. If we liberate ourselves from the latter, there is no longer anything in (27) that speaks (directly) for or against encapsulation.

Thus, while the encapsulation approach may ultimately be correct—I will argue below that it is—this is only possible if  $\varphi$ -features and binding indices are decoupled. In other words: only if the reductionist hypothesis is false.

Let us consider another argument to the same effect, this time from Basque. As already noted in section 2, Basque reflexives have the structure in (7), where: PRON.GEN is a (strong or weak) possessive pronoun; N is a designated noun, *buru*, meaning “head”; and D is the determiner/article.

(7) structure of Basque reflexives: [PRON.GEN N D]

The possessor in (7) is a bona fide pronominal element, not just diachronically (cf. English), but synchronically as well. Initial evidence from this comes from the fact that the possessor can take the form of a weak *or* a strong pronoun (which, in Basque, are morphologically distinguished):<sup>10</sup>

- (28) a. <pro<sub>2pl.ERG</sub>> [ **zuen** buru-a ] saldu d- $\phi$ -u-zue  
**2pl(weak).GEN** head-ART<sub>sg</sub>(ABS) sold 3.ABS-SG.ABS- $\sqrt{\text{V}}$ -2pl.ERG  
 ‘Y’all have given yourselves away’ (Basque) [= (8)]
- b. <pro<sub>2pl.ERG</sub>> [ **zeuen** buru-a ] saldu d- $\phi$ -u-zue  
**2pl(strong).GEN** head-ART<sub>sg</sub>(ABS) sold 3.ABS-SG.ABS- $\sqrt{\text{V}}$ -2pl.ERG  
 ‘Y’all have given yourselves away’ [Artiagoitia 2003:620]

As these examples demonstrate, the agreement morphology that arises when the anaphor is targeted is *trivial* (in the sense defined in section 2). This means that the outermost layer of the anaphor (*viz.* the DP that is an extended projection of *buru* “head”) is invariably 3rd person and singular.

Once again, the reductionist view—whereby  $F_\varphi$  underpins anaphoric binding—entails, if taken seriously, that the binding index associated with this reflexive cannot reside at the level of the outermost DP. Importantly, the  $\varphi$ -features of the antecedent *are* tracked by the Basque reflexive; but not at the level of the outermost DP. On the reductionist view, then, the binding index must reside on the  $\varphi$ -bearing subpart of the reflexive—which, in this case, is the possessive pronoun (PRON.GEN). This leads to a contradiction. To see this, let us ask: what kind of element, binding-theoretically, is this index-bearing possessive pronoun? Consider the following examples (indicated judgments reflect western varieties of Basque):<sup>11</sup>

<sup>10</sup>Both variants of (28) are also possible with a plural head noun (*buru-ak* “head-ART<sub>pl</sub>(ABS)”; Artiagoitia 2003:621). This fact is reminiscent of the phenomenon of dependent plurals (cf. the English *The plastic surgeons gave each other a new nose / new noses*).

<sup>11</sup>Thanks to Karlos Arregi, Aitor Lizardi Ituarte, and Juan Uriagereka, for helpful discussion.

- (29) *Mirandek<sub>i</sub>* [ **berē**□ *buru-a* ]□ *hil z-φ-u-en*  
 Mirande.ERG 3sg.GEN head-ART<sub>sg</sub>(ABS) killed 3.ABS.3sg.ERG-sg.ABS-√-PAST  
 ‘Mirande killed himself.’  
*Cannot mean:* ‘Mirande killed <someone else>.’
- (30) *Peio<sub>k<sub>i</sub></sub>* *erran d-φ-u-φ* [ **berē<sub>i</sub>** *zakurra hil dela* ]  
 Peio.ERG say 3.ABS-sg.ABS-√-3sg.ERG 3sg.GEN dog.ART<sub>sg</sub>(ABS) die AUX.that  
 ‘Peio<sub>i</sub> said [that his<sub>i</sub> dog died].’
- (31) A: *Ezagu-tzen al d-φ-u-zu Peru Arrieta?*  
 know-IMPF Q 3.ABS-sg.ABS-√-2sg.ERG Peru Arrieta  
 ‘Do you(sg.) know Peru Arrieta?’
- B: *Bai horixe! Sarritan izan n-aiz bere etxe-a-n.*  
 yes this.INTS often 1sg.ABS-√ 3sg.GEN house-ART<sub>sg</sub>(ABS)-LOC  
 ‘Yes, of course! I have often been to his house.’
- [western dialects; Artiagoitia 2003:621, 626]*

Suppose the anaphoric index in Basque reflexives indeed resided on the possessive pronoun—*berē* (“3sg.GEN”), in this case. How would we explain, then, the fact that *berē* in (29) requires a local (clausemate) antecedent, whereas *berē* in (30–31) does not? It is logically possible, of course, for a given element to be systematically ambiguous(=homophonous) between an anaphor and a pronoun; but that is crucially not what is going on here. If it was, then (29) would be predicted to have both a coreferential and a non-coreferential reading (cf. (31B)), contrary to fact.<sup>12</sup> Instead, what is going on (in western dialects; see *Artiagoitia 2003:625–626* for discussion of the relevant dialectal variation) is that *berē* (“3sg.GEN”) requires a local antecedent *only* when it is the possessor of *buru* (“head”). But if the binding index is on the possessor, then the possessum—whether its contribution is a grammaticalized one or not—is outside the scope of the binding index, and is ipso facto inaccessible to any compositional account of that index’s binding-theoretic interpretation. We have thus arrived at a contradiction for the reductionist view.

The alternative, as suggested above, is to concede that the  $\varphi$ -varying (sub-)constituent of a reflexive anaphor need not be where the binding index resides. Specifically, the anaphoric index would reside on the entirety of *berē buru-a* (“3sg.GEN head-ART<sub>sg</sub>(ABS)”), similar to what was shown for English in (27), while, crucially, maintaining that this is *not* the constituent whose  $\varphi$ -features co-vary with the antecedent (and cannot be, given the agreement facts in (28)). This is anathema to the reductionist view, since it illustrates very clearly that binding indices and  $\varphi$ -features do not transit together.

Coming to terms with such dissociations between the locus of binding indices and the locus of valued  $\varphi$ -features also allows us to make sense of other long-standing puzzles. Consider the behavior of *to*-Experiencers in English:

- (32) The children seem [PP *to* [DP *her<sub>j</sub>/\*<sub>i</sub>*]] to have amused Mary<sub>i</sub>. *[=(19a)]*

As is well known, these experiencers behave, binding-theoretically, as though the element bearing the index was the entire PP headed by *to* (as the disjoint-reference effect in (32) demonstrates). However, PPs in English do not typically behave as bearers of  $\varphi$ -features; it stands to reason that

<sup>12</sup>See Amiridze (2003) for related observations about Georgian.

the bearer of  $\varphi$ -features, when it comes to the Experiencer argument, is the DP complement of *to*. One possible response to this mismatch is in terms of *feature percolation* (Chomsky 1973, Cowper 1987, Gazdar et al. 1985, Grimshaw 2000, Kayne 1983, Webelhuth 1992; but see Cable 2007, 2010, Heck 2004 for general arguments against such a mechanism). On this view, the  $\varphi$ -features of *her* in (32) percolate to the PP layer, together with the binding index, explaining the binding-theoretic behavior observed in (32).

Internal to English, I see no way to reason for, or against, this feature-percolation approach. Crucially, however, we already know that this approach does not generalize. Consider once more the Basque cases in (28a–b), above. If  $\varphi$ -feature percolation (from the unmistakably  $\varphi$ -feature-bearing possessor, to the entire anaphoric expression) were what underpinned these cases, we would see these percolated  $\varphi$ -features reflected in absolutive agreement with the anaphor—i.e., we would see nontrivial agreement with the anaphor—and that is simply not what we see. In other words, in languages where we can actually *test* whether  $\varphi$ -features have percolated in the manner one might hypothesize for (32), the verdict is that they have not. (See Amiridze 2003 for converging evidence from Georgian.) This renders the percolation explanation of (32) not only untestable but, from a cross-linguistic perspective, *ad hoc*. The alternative is, once again, to acknowledge that binding indices do not travel in tandem with  $\varphi$ -features. In other words: to reject reductionism.

## 7. Interim Summary

In previous sections, we have seen that the Anaphor Agreement Effect (AAE) speaks against reductionist theories: those that seek to reduce anaphoric binding to syntactic agreement in  $\varphi$ -features. Reductionist theories require a timing-based account of the AAE (as discussed in section 3); but timing-based approaches are in direct conflict with reductionism (section 4). Above and beyond this, timing-based approaches to the AAE turn out to be untenable *regardless* of whether one adopts reductionism or not (section 5).

An alternative to this stagnant line of reasoning involves the idea of encapsulation (section 6): the  $\varphi$ -features of the anaphor are enclosed in a larger structure, which is opaque for  $\varphi$ -probing, giving rise to the AAE. Some languages—such as Basque and Georgian—manifest encapsulation quite transparently. Moreover, taking the data from these languages seriously involves giving up on reductionism once and for all, since they are, upon careful examination, incompatible with the idea that binding indices and  $\varphi$ -features travel in tandem (see, again, section 6).

With this in place, we are now in a position to develop a general account of the AAE, one that eschews the reductionist view, and is based solely on encapsulation. That account is the topic of the next section.

## 8. Proposal: encapsulation without reductionism

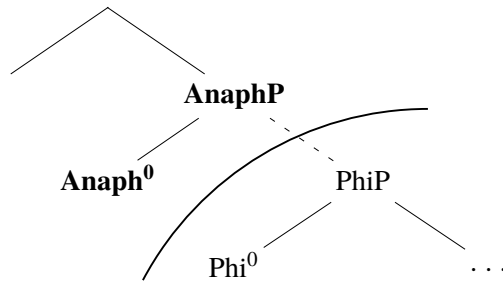
### 8.1. The encapsulation hypothesis once more

The encapsulation approach put forth in section 6 takes the  $\varphi$ -bearing layer of anaphors to be distinct from, and internal to, their outermost structural layer. Since we have now moved away from reductionism, and the idea that  $F_\varphi$  between DP[A] and DP[B] is implicated in anaphoric binding, let us also move away from the DP[A]/DP[B] labels used in previous sections. (This leaves open, of course, the question of how  $\varphi$ -feature-matching between the binder and the anaphor *does* arise;



I set aside this question temporarily, returning to it in section 8.3.) Let us re-label the projections implicated in the encapsulation approach as in (33):

(33) THE ENCAPSULATION HYPOTHESIS



AnaphP stands for whatever the outermost layer for the anaphor is, while PhiP stands for whatever projection within the anaphor hosts valued  $\varphi$ -features.<sup>13</sup> To the extent that the interior of AnaphP is consistently opaque to  $\varphi$ -probing from the outside (as indicated graphically in (33)), the result would be the absence of nontrivial agreement with anaphors, i.e., the AAE.<sup>14</sup>

Even setting aside the issue of AnaphP’s opacity, though, this can only be taken as an account of the AAE if the (partial) structure in (33) is universal. We have already seen that some languages realize the structure in (33) quite transparently (see section 6). But if other languages could “opt out” of this structure, we would then predict that they could also “opt out” of the AAE—which does not seem to be the case.

Therefore, I devote section 8.2 to a review of recent work by Middleton (2018), who shows that anaphors indeed adhere to a universal articulated structure, of which (33) is proper subpart. Section 8.3 is then devoted to the issue of  $\varphi$ -feature-matching between binder and anaphor, and specifically to the question of how it is to be handled given a rejection of reductionism. In section 8.4, I return to the status of the assumption regarding the syntactic opacity of AnaphP.

## 8.2. Evidence for the universality of encapsulation: Middleton 2018

Middleton (2018) has conducted a survey of 86 languages, from 13 language families, looking specifically at the forms used to express the most-embedded nominal in each of the following four target meanings (the labels used for the different categories are Middleton’s):

<sup>13</sup>Technically, it need not be the case that all valued  $\varphi$ -features on the anaphor are hosted on a single projection (cf. Ritter 1991, 1992 and much related work). All that is crucial is that even the highest of the relevant projections is still lower than AnaphP.

<sup>14</sup>One issue that the structure in (33) does not address is the fate of attempted agreement with an anaphor—whether it results in a non-varying default, or in outright ungrammaticality. I return to this issue in section 9.2.

- (34) a. “ANAPHOR”  
Diana  $\lambda x(x$  thinks that only Charles  $\lambda y(y$  loves  $\underline{y}$ ))
- b. “DIAPHOR”  
only Diana  $\lambda x(x$  thinks that Charles  $\lambda y(y$  loves  $\underline{x}$ ))
- c. “EXOPHOR”  
only Diana  $\lambda x(x$  thinks that Charles  $\lambda y(y$  loves  $\underline{z}$ )), where  $z =$  Diana
- d. “PRONOUN”  
only Diana  $\lambda x(x$  thinks that Charles  $\lambda y(y$  loves  $\underline{z}$ )), where  $z \neq$  Diana

Middleton shows that, when lined up in this order, the forms each language employs for the four most-embedded nominals in these target meanings exhibit what is called a *ban on discontinuous syncretism* (cf. Bobaljik 2012, Bobaljik & Sauerland 2018, Caha 2017). To understand what this means, suppose we list the forms a given language uses in each of these four cases, starting from (34d) and proceeding upwards in the list. And suppose we visualized the paradigm by assigning a new capital letter every time a new form enters the paradigm. By this token, we could say that English exhibits an “**AAAB**” pattern (e.g. *them, them, them, themselves*). Other languages behave differently—Malayalam, for example, exhibits an “**AABC**” pattern (as do Icelandic, and some dialects of Mandarin):

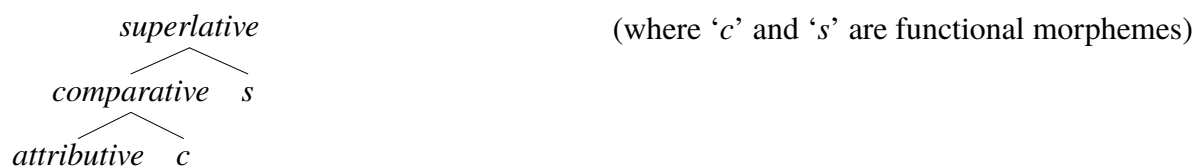
- (35) a. kanga piglet-inə ma:trame: **avanavan**-e iʃta:ɳeɳ:ə ɟintik:juɳu (Malayalam)  
Kanga Piglet-DAT only AVANAVAN-ACC like.COMP thinks  
Kanga  $\lambda x(x$  thinks that only Piglet  $\lambda y(y$  loves  $y$ ))
- b. piglet ma:tre: viɟa:rik:ʃiɳlu kanga-k:ʃe **tan**:-e iʃta:ɳeɳ:ə  
Piglet only thinks Kanga-DAT TAN-ACC like.COMP  
only Piglet  $\lambda x(x$  thinks that Kanga  $\lambda y(y$  loves  $x$ ))
- c. piglet ma:tre: viɟa:rik:ʃiɳlu kanga-k:ʃe **avan**-e iʃta:ɳeɳ:ə  
Piglet only thinks Kanga-DAT AVAN-ACC like.COMP  
only Piglet  $\lambda x(x$  thinks that Kanga  $\lambda y(y$  loves  $z$ )), where  $z =$  Diana  
only Piglet  $\lambda x(x$  thinks that Kanga  $\lambda y(y$  loves  $z$ )), where  $z \neq$  Diana

[Jane Middleton, p.c.]

What is excluded, however, are cases where two non-adjacent elements in the quartet share a given form but one or more elements in between them do not share that same form (e.g. **\*AABA**, **\*ABCA**, **\*ABCB**, **\*ABAC**, and so forth).

Bobaljik (2012), based on a similar ban on discontinuous syncretism found in the domain of attributive-comparative-superlative adjectival paradigms (i.e., **\*ABA**), argues for a containment relation between these three expressions. Namely: the comparative is built from, and structurally contains, the attributive; and the superlative is built from, and structurally contains, the comparative. This is schematized in (36):

(36) ADJECTIVAL CONTAINMENT HYPOTHESIS (Bobaljik 2012)



Here is how the structure in (36) derives the ban on discontinuous syncretism: if the comparative is created by affixing the functional morpheme *c* to the attributive, and the superlative properly contains the comparative, then *c* will be present in the superlative, as well. Thus, if *c* is a trigger for contextual allomorphy of the attributive stem, that allomorphy will also be triggered in the superlative (since the trigger, *c*, is present in the latter, as well). This would result in an **ABB** pattern. Of course, *s* could itself be an allomorphy trigger, in which case it would give rise to a third form for the superlative, i.e., an **ABC** pattern.

On these assumptions, the only way to create an **ABA** pattern is as a particular subcase of **ABC** in which **C** is accidentally homophonous with **A** (but both remain distinct from **B**). Bobaljik stipulates that there is a weak type of *anti-homophony* constraint at play: one that does not ban accidental homophony outright—that would obviously be too strong—but which bans a context-sensitive allomorph of a given morpheme being homophonous with an ‘elsewhere’ allomorph of the same morpheme:

- (37) A context-sensitive allomorph of  $\mu$  cannot be homophonous with the default exponent of  $\mu$ .  
[Bobaljik 2012:35]

In the pathological derivation of **ABC** sketched above, the shape of the morpheme corresponding to *attributive* in (36) has to be homophonous with the allomorph of that morpheme that arises in the context of *s*, in violation of (37). This completes Bobaljik’s (2012) derivation of the ban on discontinuous allomorphy (i.e., **\*ABA**)—though this is a severely abridged version, and the reader is strongly encouraged to consult the source for a far more detailed version. (See also Caha 2009, 2013.)

As noted by Caha (2017), however, a ban on discontinuous syncretism is not an unambiguous indicator of containment (see also Bobaljik & Sauerland 2018). Focusing, for simplicity, on the three-cell paradigm (where the ban on discontinuous syncretism amounts to a ban on **ABA** patterns), the same ban could arise in the following fashion. Let  $\alpha$ ,  $\beta$ , and  $\gamma$  be three cells in a paradigm which, cross-linguistically, consistently obeys the ban on **ABA** patterns. Caha shows that alongside the containment-based derivation proposed by Bobaljik 2012 ( $\gamma$  containing  $\beta$ , which in turn contains  $\alpha$ ; or  $\alpha$  containing  $\beta$ , which in turn contains  $\gamma$ ), there exists another way to derive such facts. Suppose that the three categories in question are each composed of one or more of a pair of features *f* and *g*, as shown in (38):

(38)

CATEGORY	FEATURAL MAKE-UP
$\alpha$	[ <i>f</i> ]
$\beta$	[ <i>f g</i> ]
$\gamma$	[ <i>g</i> ]

If one adopts a slightly modified version of *anti-homophony*—one which prohibits two identical allomorphs of a morpheme triggered by disjoint featural contexts—then the same results will obtain here. The exponent of  $\beta$  could be an  $f$ -triggered allomorph of the default exponent of  $g$ ; or it could be a  $g$ -triggered allomorph of the default exponent of  $f$ . But the only way for it to be simultaneously distinct from both  $\alpha$  and  $\gamma$ , and have  $\alpha$  and  $\gamma$  themselves be identical, would be to have accidentally-homophonous spellouts for  $f$  and for  $g$  in the absence of the other feature, in violation of the (modified) anti-homophony constraint. Caha (2017) discusses a real-world example where the state of affairs shown in (38) is exceedingly likely to be at play, involving the relation between dative, allative, and locative morphology. Importantly, there are languages that transparently expone the hypothesized feature structures:

(39) Tigrinya (Afro-Asiatic)

CATEGORY	EXPONENT
DATIVE	ne
ALLATIVE	ne + ab
LOCATIVE	ab

However, just as there are languages (like Tigrinya) which transparently realize a Caha-style, featural overlap derivation of the dative-allative-locative paradigm, so too there are language that transparently realize a Bobaljik-style, containment-based derivation of the attributive-comparative-superlative paradigm. Persian is an example of this:

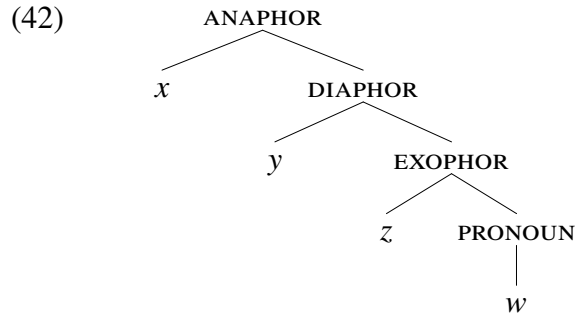
(40) *bozorg* “big” (Persian)  
*bozorg-tar* “bigger”  
*bozorg-tar-in* “biggest”

As Bobaljik & Sauerland (2018) and Middleton (2018) discuss, data like (39) and (40) disambiguate between the two competing explanations of \*ABA in a given empirical domain, in favor of featural overlap and containment, respectively.

So now, returning at long last to our domain of interest, anaphora—and specifically, to the findings of a ban on discontinuous syncretism in this domain—the question is whether a pattern like (39) or (40) is ever observed here. And Middleton provides the answer: there are languages that realize transparent containment in this domain. In the Peranakan Javanese of Semarang (PJS), one finds the following realization of the expressions in Middleton’s taxonomy:

(41) ANAPHOR *awake dheen dhewe* (PJS)  
 DIAPHOR/EXOPHOR *awake dheen*  
 PRONOUN *dheen* [Cole et al. 2007, via Middleton 2018]

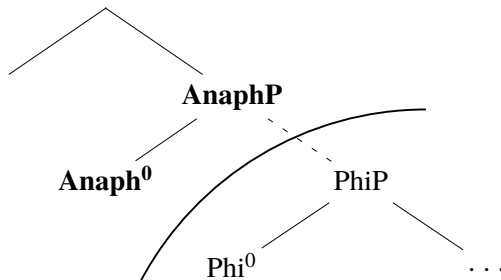
As Middleton (2018) discusses in greater detail, this pattern decisively disambiguates between the two possible explanations in favor of containment, and specifically the following structure:



There are two specific aspects of (42) that I would now like to turn the reader’s attention to. First, it is necessarily the case that (42) represents a *universal*, cross-linguistically invariant structure for the expressions in question. If we relax this assumption, the explanation for why there is not a single example of discontinuous syncretism in Middleton’s (fairly vast) survey is lost.

Second, recall the (partial) structure put forth in section 8.1 to explain the AAE, repeated here:

(33) THE ENCAPSULATION HYPOTHESIS



We know that, cross-linguistically, pronouns are perfectly good bearers of  $\varphi$ -features. We might even say that they are the *canonical* bearers of  $\varphi$ -features, as they are the only expressions that exhibit a nontrivial range of PERSON values, alongside NUMBER and GENDER. Thus, if we were to map the categories in (33) onto Middleton’s categories, PhiP must be at least as low as PRONOUN. AnaphP, on the other hand, must be—at a minimum—high enough to contain Middleton’s complete structure for anaphors. In other words, AnaphP is at least as high as ANAPHOR.

We thus have, in Middleton’s (2018) results, evidence for the universality of the structure in (33): PhiP properly contained within AnaphP. Consequently, an account of the AAE can make reference to (33) as an invariant universal, as the account put forth in section 8.1 does. (It should be noted that these results still do not speak to the syntactic opacity of AnaphP; see section 8.4.)

### 8.3. Whence $\varphi$ -matching?

Earlier, in section 6, we saw data from Basque (for which close parallels exist in Georgian), showing that encapsulation is flatly incompatible with the assumption that binding indices travel in concert with  $\varphi$ -features. That is, encapsulation is incompatible with the reductionist hypothesis, which takes anaphoric binding to rely on the same relation ( $F_\varphi$ ) that underpins  $\varphi$ -agreement. Since languages like Basque still exhibit the AAE, there is no hope for a reductionism-based account of the AAE to achieve true generality. We have also seen how an alternative account of the AAE, one that is based on encapsulation but eschews reductionism, would work (section 8.1). Finally, we saw evidence that encapsulation is not restricted to languages like Basque (and Georgian), which “wear it on

their sleeve.” As discussed in detail in section 8.2, encapsulation is a universal, cross-linguistically stable property of anaphors. Anything else would leave Middleton’s (2018) results unexplained.

This leaves us with two main questions. One concerns the syntactic opacity of the anaphoric layer (AnaphP), required for the encapsulation-based account of the AAE to work. I continue to defer discussion of this matter until section 8.4. In the current subsection, I address a different question, namely, how anaphors come to match their binders in  $\varphi$ -features, given that the relation responsible for  $\varphi$ -agreement cannot be responsible for this.

The answer here is a fairly simple one, and amounts to the following: we know that the matching of  $\varphi$ -features is enforced even in scenarios where syntactic  $\varphi$ -agreement (i.e.,  $F_\varphi$ ) could not possibly be involved. And we know that this is so even when it comes to  $\varphi$ -features that are not interpreted, such as grammatical gender on inanimates. (I will discuss two such scenarios shortly.) Given this, there is no choice but to posit some distinct, non-syntactic mechanism—call it  $NSM_\varphi$ —responsible for the latter kind of case. And finally, given that  $NSM_\varphi$  is also sufficient to ensure  $\varphi$ -matching under anaphoric binding, there can be no argument based on  $\varphi$ -matching in favor of reductionism. In other words, an account like the present one, which eschews any role for  $F_\varphi$  in anaphoric binding, loses no ground as far as  $\varphi$ -matching is concerned. The non-syntactic  $NSM_\varphi$  must necessarily exist, and can do the same work in the relevant cases.

Let us now turn to those phenomena which motivate the existence of  $NSM_\varphi$ . The first involves  $\varphi$ -matching in the absence of c-command; consider (43):

(43) No linguist who has purple pants looks silly in them/\*it.

The example in (43) is an instance of Donkey Anaphora: on the intended readings, the underlined expressions covary. This, despite the absence of c-command (in either direction) between the covarying expressions; as well as the fact that the antecedent clause is buried inside a Complex NP Island (of the relative-clause variety), which is itself situated *inside* a Subject Island. Nevertheless, the features of the pronoun must match the (apparently uninterpreted) features on the antecedent.

Lest one think that this is specific to instances of pluralia tantum (for which a semantically-contentful approach may not be out of the question; cf. Acquaviva 2008, *a.o.*), the same happens with grammatical gender on inanimates, as the following Hebrew examples demonstrate:

- (44) a. kol exad je-yej l-o maxberet je-ya-sim ot-a/\*ot-o  
 every one that-EXIST DAT-3sgM notebook<F> that-3sgM.FUT-put ACC-3sgF/\*ACC-3sgM  
 ba-tik (Hebrew)  
 in.the-case  
 ‘Everyone who has a notebook<F>, put it.F/\*it.M in your bag ’
- b. kol exad je-yej l-o maxfevon je-ya-sim ot-o/\*ot-a  
 every one that-EXIST DAT-3sgM calculator<M> that-3sgM.FUT-put ACC-3sgM/\*ACC-3sgF  
 ba-tik  
 in.the-case  
 ‘Everyone who has a calculator<M>, put it.M/\*it.F in your bag ’

Clearly, a syntax that can relate two expressions in the absence of c-command *and* in a manner that disregards islands is hopelessly unrestrictive; it is really no syntax at all. Consequently, cases like (43–44) are proof positive that even  $\varphi$ -features that are not interpreted, such as the grammatical

gender features of inanimates, can be transmitted by a *non-syntactic* mechanism—what was referred to above as  $NSM_\varphi$ .

There are candidates in the literature for what this mechanism might be. For Elbourne (2013), pronouns are (hidden) definite descriptions, and so the matching requirements evidenced in (43–44) dissolve into whatever pragmatic forces favor coherence between a definite description and the expression(s) used to introduce the described entity earlier in the discourse. For Merchant (2014), pronouns like these are the pronounced residue of NP ellipsis, and so the matching requirements evidenced in (43–44) dissolve into whatever enforces identity of form between ellipses and their antecedent (cf. *\*John is no longer a bachelor, and Bill did <get married> too*).

For the present purposes, however, it is not crucial that either of these approaches be correct or without problems (see also the discussion below, concerning unanteceded deixis); all that matters is the reductio showing that the mechanism underpinning  $\varphi$ -matching in (43–44) cannot be a *syntactic* one.

Also relevant are instances of deixis that lacks any utterance-internal antecedent (henceforth, *unanteceded deixis*). These cases involve  $\varphi$ -matching across different utterances, and even across different speakers:

(45) A: Where are the scissors?

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 (be it Agree or something else) 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.

As before, the same holds of grammatical gender. In Kinyarwanda (Bantu), for example, if one is pointing to a pair of saloon-style doors, one can say (46):

(46) ir-a-kingu-ye

4SUBJ-PAST-OPEN-PRFV

‘They are open.’ (‘They have been opened.’)

Crucially, the subject-agreement marker in (46) is *ir-* because (plural) “doors” are a member of noun-class 4 (rather than 2, 6, 8, and so forth). But these facts concerning (46) hold even in the absence of any linguistic antecedent whatsoever (i.e., even if the earlier reference to the doors was entirely non-linguistic). Therefore, the mechanism that enforces class 4 morphology here could not possibly be syntactic.

Before returning to the implications of these facts for our account of the AAE, it is worth pausing to discuss a putative swing of the pendulum all the way in the other direction. Dowty & Jacobson (1988), discussing a related set of facts, hint (at the very least) at the possibility that *all* cases of  $\varphi$ -matching, even those that are referred to here as  $\varphi$ -agreement, be handled by a non-syntactic mechanism. (For Dowty & Jacobson, the mechanism in question would be the semantic-pragmatic component, but that is not crucial for our current discussion.) It is certainly tempting to consider a theory where  $F_\varphi$  is entirely subsumed by  $NSM_\varphi$ . One operation is always better than two, and given that  $NSM_\varphi$  cannot be subsumed by  $F_\varphi$  (as just shown), the only one-operation theory that

is still on the table is one that would use  $NSM_\varphi$  alone. There is, however, good evidence that  $\varphi$ -agreement—that is,  $\varphi$ -feature agreement between a predicate and its argument(s)—must be handled syntactically, i.e., by  $F_\varphi$ .

Consider a structural configuration like (47), where ‘ $\gg$ ’ represents c-command, and H is some functional head:

(47) H  $\gg$  DP<sub>1</sub>  $\gg$  DP<sub>2</sub>

As has been documented extensively by Abels (2012), Boeckx & Jeong 2004, Frampton 1991, Preminger 2014, Rizzi 1990b, 2001, Starke 2001, and others, H cannot enter into  $\varphi$ -agreement with DP<sub>2</sub> across DP<sub>1</sub> in a configuration like this.<sup>15</sup>

Now, there are superficially similar effects in binding, and even in association with focus. But the case involving  $\varphi$ -agreement is quite different. As has been shown in the work cited above, the intervention effect in (47) is purely *structural*. It cares about case (including structural case, which cannot be reduced to semantics), and it can be fixed by A-moving DP<sub>1</sub> out of the way, as already shown in section 4 (see (20–21) and the surrounding discussion).

Binding, on the other hand, does not care about case. (In fact, the ability of dative subjects to bind subject-oriented anaphors in Icelandic is one of the most striking pieces of evidence that they are indeed subjects; see Zaenen, Maling & Thráinsson 1985 and related literature). And binding-related minimality violations cannot be repaired by movement, as also shown in section 4 (see (19, 22) and the surrounding discussion).

Perhaps most importantly,  $\varphi$ -agreement never does anything like what we see in (43–44). Verbs—or more accurately, the functional heads that verbs typically occur with—can agree with DPs that are not their arguments (see Artiagoitia 2001, Rezac, Albizu & Etxepare 2014, among many others); they can agree with DPs in other (lower) clauses (see Polinsky 2003, Preminger 2011b, among many others); but they cannot blatantly ignore the contours of syntactic structure (c-command, islands, etc.) in the manner shown in (43–44).

What this all means is that a Dowty & Jacobson-inspired swing in the other direction, whereby even  $\varphi$ -agreement would be handled by  $NSM_\varphi$ , does not accord with the facts.

To summarize, we have seen in this subsection that  $\varphi$ -matching occurs in cases where it could not possibly be underpinned by a syntactic relation (under Donkey Anaphora, and in unantecedent deixis). In moving away from a reductionist theory—as we have done in the preceding sections—we lose the ability to attribute  $\varphi$ -matching under anaphora to  $F_\varphi$ . What the current subsection has shown is that this is not a real loss of coverage; the grammar is clearly equipped with other mechanisms, which achieve the same result of  $\varphi$ -feature matching between two expressions, and could not possibly be based on syntactic  $\varphi$ -agreement. Those mechanisms could just as well apply to the binder-anaphor relation, and yield  $\varphi$ -matching there.

---

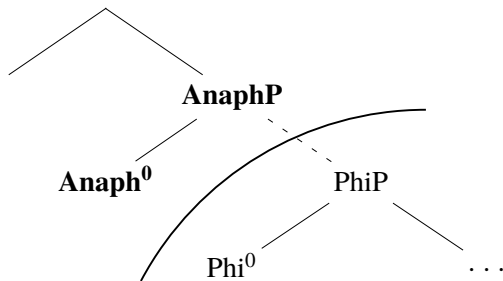
<sup>15</sup>This phenomenon of intervention in  $\varphi$ -agreement is not to be confused with the phenomenon of dative intervention in *A-movement*, which Bruening (2014) and, more recently, Branan (2018), have challenged the very existence of. The discussion here concerns cases where DP<sub>2</sub> is not a candidate to A-move, regardless of whether DP<sub>1</sub> is there or not, and only *agreement* between H and DP<sub>2</sub> is at stake (e.g. in Basque or Icelandic). Bruening and Branan are mum on such cases.



#### 8.4. Anaphor Opacity: a one-postulate account of the AAE

The last remaining element of the account put forth in section 8.1, which has not yet received independent support, is the syntactic opacity of AnaphP:

(48) ANAPHOR OPACITY



[=(33)]

At the present, I have no independent support to offer for this opacity assumption. The fair thing, therefore, is to treat it as nothing more, and nothing less, than a stipulation.

However, since the structure involving PhiP properly contained in AnaphP is independently supported (section 8.2), all that needs to be stipulated here is the property of opacity. This property is known to vary among different phrases. Some phrases are locality boundaries (e.g. Chomsky’s 2001 “phases”), some are not. Postulating that AnaphP is syntactically opaque, then, represents tangible progress in explaining the AAE. Stipulating the latter directly would involve stipulating *sui generis* restrictions on agreement with anaphors, of a kind not found in other syntactic contexts (cf. (9–10), repeated below); whereas stipulating that a given category is opaque has well-established parallels elsewhere in grammar.

(9) ANAPHOR AGREEMENT EFFECT (AAE), *revised version*

\*  $H^0 \dots DP_{ANAPH}$ , where  $\mathcal{R}$  is a nontrivial  $\varphi$ -agreement relation

└───┬───┘  
 $\mathcal{R}$

(10) A  $\varphi$ -agreement relation between  $\alpha$  and  $\beta$  is nontrivial if there are at least two sets of feature-values,  $F$  and  $F'$ , such that  $\alpha$  bears  $F$  **iff**  $\beta$  bears  $F$ , and  $\alpha$  bears  $F'$  **iff**  $\beta$  bears  $F'$ .

Compare this one-postulate account with the alternatives. A timing-based account of the AAE requires its own stipulation, namely, that anaphors begin their lives in a  $\varphi$ -deficient state (section 3). But as show in sections 4–5, it cannot serve as a truly general account of the AAE, given the existence of “HIGH-ABS”/“ABS=NOM” languages. Reductionism, while much more appealing from the perspective of theoretical economy, does not work either, as transparent encapsulation languages demonstrate; see section 6. (Another issue that merits our attention is cross-linguistic differences in the fate of attempts to agree with anaphors; this issue is taken up in section 9.2, but, to foreshadow that discussion, it does not seem to distinguish among the different accounts currently available.)

It should be clear, then, that the anaphor-opacity account in (48) is the best available account of the AAE. Moreover, it suggests a clear direction for future research, namely, figuring out why it is that AnaphP is opaque. It is not clear—to me at least—that we currently have a satisfactory answer for this question as it regards other categories, like CP, either. Chomsky (2000:106–107) entertains the idea that phases can be defined in terms of *convergence*. But the logic of convergence is based on a theory of feature-checking and derivational time-bombs, a model that is at any rate incorrect

(Preminger 2014). The other option entertained by Chomsky involves the notion of a phase being *propositional*, but this definition is too vague to be of any use. PPs, for example, behave as phases (Abels 2003), and if PPs are “propositional” in the relevant sense, one begins to wonder what it would take for a category *not* to qualify as propositional.

None of this means, of course, that the question of why some XP is a phase is not answerable in principle. Moreover, it is conceivable that we would have a viable answer for this as it regards AnaphP *before* we have a viable answer as it regards CP. But I leave these questions for future research.

## 9. Challenge for the analysis

### 9.1. Non-AAE-obeying languages

The account presented in the previous section derives the AAE as a universal property of all anaphors in all languages. Therefore, to the extent that one finds languages that *violate* the AAE, such languages pose a problem for this analysis.

Murugesan (2018, in prep.) gives data from several languages showing exactly this. Here I will concentrate on Murugesan’s Tamil data, as it poses what is *prima facie* the most acute challenge to the current proposal.

Tamil (Dravidian) is a nominative-accusative language, where only the nominative controls agreement on the verb:

- (49) Meena Kohli.y-ai paar-t-aal (Tamil)  
 Meena(NOM) Kohli-ACC see-PAST-3sgF  
 ‘Meena saw Kohli.’ [Murugesan 2018:(13)]

Tamil also has dative-subject verbs, whose object then surfaces in the nominative:

- (50) Kohli-ukku Meena kidai-t-aal  
 Kohli-DAT Meena(NOM) see-PAST-3sgF  
 ‘Kohli got Meena.’ [Murugesan 2018:(14)]

As one might expect, the nominative is then the controller of agreement in such cases. (Note that Kohli is masculine, and Meena is feminine.) Further demonstrating the nominative-only restriction on agreement, verbs with no nominative arguments at all must surface within 3sg neuter agreement, regardless of the  $\varphi$ -features of their arguments:

- (51) a. Kohli-ukku pasi-t-atu/\*aan  
 Kohli-DAT hungry-PAST-3sgN/\*3sgM  
 ‘Kohli was hungry.’  
 b. Kohli-ukku Meena.v-ai pidi-t-atu/\*aan/\*aal  
 Kohli-DAT Meena-ACC like-PAST-3sgN/\*3sgM/\*3sgF  
 ‘Kohli liked Meena.’ [Murugesan 2018:(16–17)]

We can now place an anaphor in the object position of a dative-subject verb, confident that this anaphor would be the only possible source of non-3sgN agreement in the sentence. And the fact is that such cases do give rise to nontrivial agreement:<sup>16</sup>

- (52) a. Kohli-ukku taan                      tirumba kidai-t-aan  
       Kohli-DAT REFL(NOM) again    got-PAST-3sgM  
       ‘Kohli got himself back again.’  
       b. Meena-ukku taan                      tirumba kidai-t-aal  
       Meena-DAT REFL(NOM) again    got-PAST-3sgF  
       ‘Meena got herself back again.’

[Murugesan 2018:(15a–b)]

In light of (50–51), we can be quite confident that the source of non-3sgN agreement in (52a–b) is *not* the antecedent. These data therefore pose a genuine challenge to the encapsulation-*cum*-opacity analysis put forth in section 8.

Murugesan’s (2018, in prep.) analysis of these facts appeals to the relative structural height of the  $\varphi$ -probe relative to the antecedent. He proposes that the AAE arises in those languages where the  $\varphi$ -probe is located below the antecedent, and therefore enters the structure earlier, at a point when the anaphor has not yet been bound. In those languages where the  $\varphi$ -probe is located above the antecedent, the anaphor has already been bound, and thus the AAE does not arise. Tamil, on his account, is a language of the latter type, assuming that the  $\varphi$ -probe is located on T<sup>0</sup> and the subject is base-generated lower, e.g. in [Spec, $\nu$ P].

Notice that this account is a timing-based account, and furthermore, that it requires reductionism (because the difference between a not-yet-bound anaphor and a bound one consists, on this account, in the difference between not having valued  $\varphi$ -features and having valued  $\varphi$ -features). We have already seen that neither assumption is tenable. Timing-based approaches do not work (section 5), nor does reductionism (section 6). And they certainly do not work when combined with one another (section 4).

Nevertheless, much of the empirical burden in arguing against these approaches was carried by languages like Basque, Georgian, etc., where anaphors are transparently analytic, involving a possessor-possessum structure in which the possessum is a body-part noun. Could we retreat from the conclusions drawn from such data, and reinstate a timing-based account of the kind Murugesan envisions? The answer, I think, is negative. The reason is as follows: recall that our explanandum here is the AAE, *viz.* the ban on nontrivial agreement with anaphors. If we endorse the retreat in question, languages like Basque and Georgian would then stand as a twofold coincidence. First, we would have to say that the reason languages like Basque and Georgian exhibit the AAE is unrelated to (and, in fact, disjoint from) the reason why other languages exhibit the AAE. Recall that the absolutive  $\varphi$ -probe in Basque is located on T<sup>0</sup> (Arregi & Nevins 2012), above both the anaphor and the base position of the antecedent, in normal transitive clauses with a reflexive object. So it is precisely the kind of language that is predicted *not* to exhibit the AAE, on Murugesan’s account. On this view, the reason Basque (as well as Georgian) exhibits the AAE is the possessed-body-part structure of its reflexives, which is entirely unrelated to the reason other languages exhibit the AAE (the antecedent being higher than the  $\varphi$ -probe). And this brings us to the second coincidence: this purportedly idiosyncratic structure of reflexives, which causes the AAE in languages like Basque and Georgian, is in fact what Middleton (2018) has shown holds *universally* (whether it

<sup>16</sup>Jeffrey Lidz (p.c.) points out that the same is true of Kannada (also a Dravidian language).

happens to be transparently detectable in a particular language or not): anaphors involve additional structural layers encapsulating the  $\varphi$ -feature-bearing layer of structure. To recapitulate, if it were possible to deviate from this schema on a language-specific basis, Middleton’s results, concerning the universal ban on discontinuous syncretism in anaphoric expressions, would be left unexplained (see section 8.2). Thus, far from being an exception, these languages are transparent exemplars of the universal structure of anaphors, yet their behavior, on Murugesan’s account, would have to be cast as some sort of outlier. Note also that the behavior of languages like Basque and Georgian—where reflexives yield “default agreement”—is not unique to languages with possessed-body-part reflexives. The Albanian reflexive *vetja*, discussed in section 2, is a simplex anaphor (Franks 2013), and yet it gives rise to the same, non-varying 3sg agreement found in Basque and Georgian.

Finally, even if we were to tolerate these coincidences, it remains the case that the timing-based account is technically unsound even in those languages where  $v^0$  is the probe, i.e., even in those languages that are subject to a timing-based analysis on Murugesan’s view. These problems were already discussed in some detail in section 4 (see (23) and the surrounding discussion).

Before sketching an alternative, non-reductionist, non-timing-based analysis of Tamil (and languages like it), let me mention two additional facets of agreement with anaphors in Tamil. First, there are cases that appear, at first glance, to instantiate agreement with anaphors in *subject* position (in contrast to (52a–b), which involved agreement with nominative-object anaphors). Consider agreement on the embedded verbs in (53a–b):

- (53) a. Mani<sub>i</sub> [taan<sub>i</sub> sathat-ai sapi-t-**aan**-nnu]      son-n-aan  
 Mani<sub>M</sub> self rice-ACC eat-PAST-**3sgM**-COMP say-PAST-3sgM  
 ‘Mani<sub>i</sub> said that self<sub>i</sub> ate the rice.’
- b. Banu<sub>i</sub> [taan<sub>i</sub> sathat-ai sapi-t-**aal**-nnu]      son-n-aal  
 Banu<sub>F</sub> self rice-ACC eat-PAST-**3sgF**-COMP say-PAST-3sgF  
 ‘Banu<sub>i</sub> said that self<sub>i</sub> ate the rice.’ [Murugesan 2018:(46a–b)]

These data certainly appear to instantiate agreement between a verb (in the embedded clause) and its subject, which is itself anaphorically bound by an argument in the matrix clause. However, as Murugesan points out, Sundaesan (2016) has already shown that this is actually a case of agreement with a perspective-holder and, as such, need not target one of the core arguments at all. In (54), for example, the same verb exhibits 1sg agreement, despite the absence of any 1sg argument:

- (54) Banu<sub>i</sub> [taan<sub>i/\*j</sub> saatat-ai sapi-t-**een**-nnu]      so-n-aal  
 Banu<sub>F</sub> self rice-ACC eat-PAST-**1sg**-COMP] say-PAST-3sgF  
 ‘Banu<sub>i</sub> said that self<sub>i/\*j</sub> ate the rice.’ [Murugesan 2018:(48)]

Presumably, an example like (54) involves agreement with a phonologically-null, structurally realized perspective holder somewhere in the left periphery. (Given the fact that the verb and complementizer seem to form a single morphological unit, we can hypothesize that the verb has indeed moved high enough in the embedded clause to target such a syntactically realized, left-peripheral perspective holder.) Crucially, this is then sufficient to account for (53a–b) as well, without the need to assume that the anaphor has agreed with the subject here. In these examples, the perspective holder is presumably the Agent argument of the matrix verb-of-saying, and thus,  $\varphi$ -featurally identical to the anaphor; but the anaphor is not the syntactic object being targeted

for agreement. This furnishes an alternative explanation for (53a–b)—but not for the monoclausal dative-subject cases in (52a–b), of course.

Second, as pointed out to me by Rafael Abramovitz, the anaphor *taan* simply cannot be bound by 1st/2nd person antecedents in Tamil. For anaphoric readings where the antecedent is 1st/2nd person, Tamil instead uses forms that are indistinguishable from the 1st/2nd person pronominals in the language (behavior that is reminiscent of the reflexive in Romance):

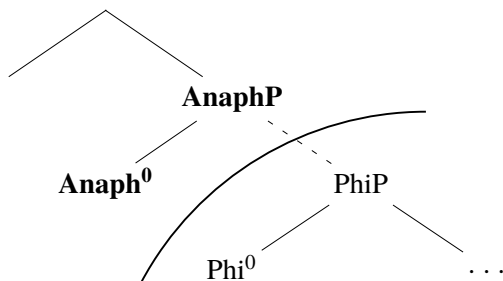
- (55) En-akku<sub>i</sub> naan<sub>i</sub>                      tirumba kidai-tt-*een*  
       1sg-DAT 1sg.NOM(PRON) again    get-PAST-1sg  
       ‘I<sub>i</sub> got myself<sub>i</sub> back again.’ [Sandhya Sundaresan, p.c.]

Thus, while agreement with anaphors in Tamil does not qualify as *nontrivial*—given, e.g., the bona fide variance in gender features seen in (52a–b)—it would also be imprecise to characterize it as full-fledged, given that the only true anaphor is restricted to 3rd person  $\varphi$ -features. One could therefore imagine a further attenuation of how we define the AAE such that nontrivial agreement *in person features* is the crucial element, and cases involving Tamil *taan* cease to be counterexamples at all.

Suppose, however, that we did not pursue such an attenuation of the AAE, and that therefore, the cases in (52a–b) continued to stand as counterexamples. Let us consider what each competing account (Murugesan’s, and the one proposed here) would then have to do to accommodate the data that remains recalcitrant for that account. We have already discussed the pitfalls of excluding languages like Basque and Georgian—languages in which the  $\varphi$ -probe is situated higher than the typical antecedent—from the purview of an AAE account. First, it would mean that these languages exhibit a ban on nontrivial agreement with anaphors for reasons that are essentially a coincidence, when viewed from the perspective of the proffered account. Second, the coincidence in question (structural encapsulation of the  $\varphi$ -bearing portion of the anaphor) actually holds universally, and crucially, not just in these “exceptional” languages (Middleton 2018). Additionally, as discussed in section 4, the timing-based approach does not actually deliver the desired results even when the probe is low (e.g. on  $v^0$ ).

The alternative is to retain the present account, repeated below, and assume that in languages like Tamil, AnaphP is *not* syntactically opaque (e.g. not a phase), and that this is a marked option entertained by the child only in the face of positive evidence.

(33) THE ENCAPSULATION HYPOTHESIS

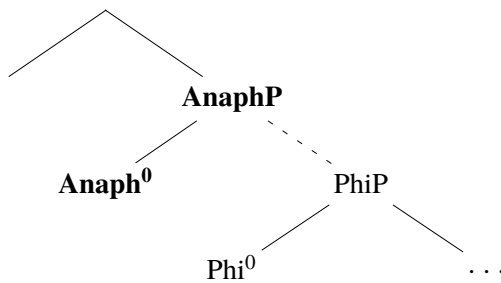


This view is not without precedent: Abels (2003) argues that this is precisely what is behind cross-linguistic variation in preposition-stranding. PP is a phase in most languages, and due to the ban on anti-locality and the inability to move out of a phase without passing through its edge, complements of phase heads are rendered syntactically immobile. This results in a ban on preposition-stranding.

However, PP is *not* a phase in a language like English, presumably because the child has ample positive evidence (in the form of stranded prepositions) to adopt this marked, non-phasal option.

The status of AnaphP would mirror this exactly. The default setting would be that AnaphP is opaque (phasal). Only overt nontrivial agreement with anaphors in the language would trigger its marked setting as non-opaque (non-phasal), as in (56). In such a language, a  $\varphi$ -probe outside of AnaphP would be able to probe into it and access the valued  $\varphi$ -features on PhiP. See also Preminger (2019), for an argument from an unrelated empirical domain that children are sensitive to overt morpho-phonological covariance (i.e., nontrivial agreement) in setting the parameters of their language. It is also suggestive, in this regard, that AAE-violating languages like Tamil are roughly as common as preposition-stranding language like English—that is, very rare (modulo the usual caveats on the pitfalls of counting languages).

(56) THE ENCAPSULATION HYPOTHESIS



It is thus much less problematic for the encapsulation-based account to accommodate languages like Tamil (non-AAE-obeying languages), than it is for a timing-based account to accommodate languages like Basque and Georgian (high-probe languages that exhibit the AAE). This joins the independent problems with a timing-based account, even for low-probe languages (section 4), as a further argument against such accounts.

## 9.2. The fate of attempted agreement with anaphors: defaults vs. ungrammaticality

One issue I have not discussed so far is the fate of attempted agreement with anaphors. Recall that what the encapsulation hypothesis derives is a ban on nontrivial agreement with anaphors:

(9) ANAPHOR AGREEMENT EFFECT (AAE), *revised version*

\*  $H^0 \dots DP_{ANAPH}$ , where  $\mathcal{R}$  is a nontrivial  $\varphi$ -agreement relation

└───┬───┘  
 $\mathcal{R}$

(10) A  $\varphi$ -agreement relation between  $\alpha$  and  $\beta$  is nontrivial if there are at least two sets of feature-values,  $F$  and  $F'$ , such that  $\alpha$  bears  $F$  **iff**  $\beta$  bears  $F$ , and  $\alpha$  bears  $F'$  **iff**  $\beta$  bears  $F'$ .

By itself, this tells us nothing about what happens if, in a given language, one nonetheless attempts to establish an agreement relation with an anaphor.

Cross-linguistically, one finds two different outcomes in this scenario. As we have seen, in languages like Icelandic the result is outright grammaticality; whereas in Albanian, the result is non-varying, “default” agreement:

- (3) Sigga<sub>i</sub> telur [ að mér líki hún<sub>i</sub>/\*sig<sub>i</sub> ] (Icelandic)  
 Sigga thinks that me.DAT likes.SBJV.3sg she.NOM/\*SIG  
 ‘Sigga<sub>i</sub> thinks that I like her<sub>i</sub>.’ [Maling 1984:217]
- (6) Vetja më dhimset. (Albanian)  
 self.NOM CL.1sg.DAT feel.sorry.for.3sg.PRES.NONACTIVE  
 ‘I feel sorry for myself.’ [Woolford 1999:271, citing Everaert 1991 and Hubbard 1985:91]

This does not seem to reduce to any other obvious other properties of the languages or reflexives in question. Albanian is a nominative-accusative language, and *vetja* is a simplex anaphor (Franks 2013). Basque, which is an ergative-absolutive language and whose reflexives are possessed-body-part reflexives, behaves exactly like Albanian in this respect.

Furthermore, as pointed out by Murugesan (2018, in prep.), the Icelandic pattern—where SIG in agreeing positions results in outright grammaticality—is surprising in and of itself. That is because, as shown in (57a–b), actual agreement with nominative objects in Icelandic exhibits optionality. (See also Hornstein 2018.)

- (57) a. Henni leiddist þeir  
 she.DAT was.bored.by.3sg they.NOM  
 ‘She was bored with them.’
- b. Henni leiddust þeir  
 she.DAT was.bored.by.3pl they.NOM  
 ‘She was bored with them.’ [Taraldsen 1995:307]

Importantly, while the encapsulation hypothesis provides no particular insight into whether an attempt to agree with anaphors in a given language will result in ungrammaticality or in trivial, non-varying agreement, neither do approaches based on timing and/or reductionism. Consider: if one assumed that agreement with an anaphor that has not yet been bound—or whose  $\varphi$ -features have not yet valued—gives rise to ungrammaticality, then Albanian and languages like it would remain unexplained. If one assumed that such agreement gives rise to default, 3sg agreement, languages like Icelandic, and English, would remain unexplained.

As it concerns English, one could appeal to a morphological gap in the inventory of reflexives, which lack nominative forms (though note that this would not extend to *itself*, since the accusative and nominative forms of *it* are syncretic):

- (58) a. herself/\*sheself  
 b. \* Mary knows that sheself is ready.

Since Icelandic SIG alternates based on case (e.g. *sig* “SIG.ACC”, *sér* “SIG.DAT”), one could imagine extending the morphological gap account to this case, as well. But even if this works, it seems like little more than a restatement of the problem. The Albanian reflexive also alternates for case (e.g. *vetja* “self.NOM”, *veten* “self.ACC”), and clearly does not have a comparable gap in the nominative.

The proper account of this difference between languages like Icelandic and languages like Albanian is therefore left for future research. What is crucial for our present purposes is that this issue—the cross-linguistically different fates of attempts to agree with anaphors—does not distinguish accounts based on timing and/or reductionism from the present, encapsulation-based account.

## 10. Conclusion

In this paper, we have seen that the Anaphor Agreement Effect (AAE), sometimes taken to support a reduction of anaphoric binding to  $\varphi$ -agreement, actually speaks rather strongly against such a reduction. We have seen that reductionism invites a timing-based approach to the AAE (section 3), but the assumptions a timing-based account must make lead to a contradiction once reductionism is adopted (section 4). Furthermore, we have seen that even when decoupled from one another, neither a timing-based approach to the AAE (section 5) nor a reductionist approach (section 6) is viable.

I have argued that the AAE instead arises as a result of *encapsulation*: the valued  $\varphi$ -features contained in the anaphor are enclosed in additional structure, whose opacity to syntactic probing yields the ban on nontrivial agreement with anaphors (sections 8.1 and 8.4). Reviewing work by Middleton (2018), we saw that this hypothesis—that the  $\varphi$ -bearing portion of anaphors is enclosed in additional structure—receives independent cross-linguistic support, and in fact must be universal (section 8.2). I then offered a proposal on the nature of  $\varphi$ -feature matching between anaphors and their binders, in light of the fact that syntactic  $\varphi$ -feature agreement cannot be what relates the two (i.e., given the failure of reductionism). I argued that nothing more is required to achieve  $\varphi$ -matching between an anaphor and its antecedent than is independently required to achieve  $\varphi$ -matching in cases where syntactic  $\varphi$ -agreement is self-evidently not involved, and that this extends even to cases of pluralia tantum and grammatical gender on inanimates (section 8.3).

Finally, I addressed two potential challenges to the proposed account (section 9): the existence of languages that do *not* obey the AAE, and cross-linguistic differences in the fate of attempted agreement with anaphors (ungrammaticality vs. a non-varying “default”). Regarding the former, I argued that on balance, the encapsulation-based account is better poised to handle such variation than its competitors. Regarding the latter, I argued that it does not distinguish the different accounts on offer from one another.

To summarize: we have seen that the AAE provides fairly strong evidence against reductionism, and that the best available account for it involves structural encapsulation, not derivational timing.

## References

- Abels, Klaus. 2003. *Successive cyclicity, anti-locality, and adposition stranding*. Doctoral dissertation, Storrs, CT: University of Connecticut.
- Abels, Klaus. 2012. *Phases: an essay on cyclicity in syntax*. Linguistische Arbeiten 543, Berlin: de Gruyter.
- Acquaviva, Paolo. 2008. *Lexical plurals: a morphosemantic approach*. Oxford: Oxford University Press.
- Amiridze, Nino. 2003. The anaphor agreement effect and Georgian anaphors. In *Empirical issues in formal syntax and semantics*, eds. Claire Beyssade, Olivier Bonami, Patricia Cabredo Hofherr & Francis Corblin, 4, 99–114. Paris: Presses de l’Université Paris-Sorbonne.
- Anagnostopoulou, Elena. 2003. *The syntax of ditransitives: evidence from clitics*. Berlin: Mouton de Gruyter.
- Andrews, Avery D. 1971. Case agreement of predicate modifiers in Ancient Greek. *Linguistic Inquiry* 2:127–151.
- Arregi, Karlos & Andrew Ira Nevins. 2012. *Morphotactics: Basque auxiliaries and the structure of spellout*. Studies in Natural Language and Linguistic Theory 86, Dordrecht: Springer.



- Artiagoitia, Xabier. 2001. Seemingly ergative and ergatively seeming. In *Features and interfaces in Romance: essays in honor of Heles Contreras*, eds. Julia Herschensohn, Enrique Mallen & Karen Zagona, 1–22. Amsterdam: John Benjamins.
- Artiagoitia, Xabier. 2003. Reciprocal and reflexive constructions. In *A grammar of Basque*, eds. Jose Ignacio Hualde & Jon Ortiz de Urbina, 607–632. Berlin: Mouton de Gruyter.
- Bjorkman, Bronwyn M. & Hedde Zeilstra. 2018. Checking up on ( $\varphi$ -)agree. *Linguistic Inquiry*, doi: <10.1162/ling\_a\_00319>.
- Bobaljik, Jonathan David. 2008. Where's phi? Agreement as a post-syntactic operation. In *Phi Theory: phi-features across interfaces and modules*, eds. Daniel Harbour, David Adger & Susana Béjar, 295–328. Oxford: Oxford University Press.
- Bobaljik, Jonathan David. 2012. *Universals in comparative morphology: suppletion, superlatives, and the structure of words*. Cambridge, MA: MIT Press.
- Bobaljik, Jonathan David & Uli Sauerland. 2018. ABA and the combinatorics of morphological features. *Glossa* 3:15, doi: <10.5334/gjgl.345>.
- Boeckx, Cedric & Youngmi Jeong. 2004. The fine structure of intervention in syntax. In *Issues in current linguistic theory: a festschrift for Hong Bae Lee*, eds. Chungja Kwon & Wonbin Lee, 83–116. Seoul: Kyungjin.
- Branan, Kenyon. 2018. *Relationship preservation*. Doctoral dissertation, Cambridge, MA: MIT.  
URL: <<https://ling.auf.net/lingbuzz/004191>>.
- Bruening, Benjamin. 2014. Defects of defective intervention. *Linguistic Inquiry* 45:707–719, doi: <10.1162/LING\_a\_00171>.
- Cable, Seth. 2007. *The grammar of Q: Q-particles and the nature of wh-fronting*. Doctoral dissertation, Cambridge, MA: MIT.
- Cable, Seth. 2010. Against the existence of pied-piping: evidence from Tlingit. *Linguistic Inquiry* 41:563–594.
- Caha, Pavel. 2009. *The nanosyntax of case*. Doctoral dissertation, Tromsø: University of Tromsø.
- Caha, Pavel. 2013. Explaining the structure of case paradigms by the mechanisms of nanosyntax. *Natural Language & Linguistic Theory* 31:1015–1066, doi: <10.1007/s11049-013-9206-8>.
- Caha, Pavel. 2017. How (not) to derive a \*ABA: the case of Blansitt's generalisation. *Glossa* 2:84, doi: <10.5334/gjgl.348>.
- Chomsky, Noam. 1973. Conditions on transformations. In *A festschrift for Morris Halle*, eds. Stephen R. Anderson & Paul Kiparsky, 232–286. New York, NY: Academic Press.
- Chomsky, Noam. 1994. *Bare phrase structure*. MIT Occasional Papers in Linguistics 5, Cambridge, MA: MITWPL.
- Chomsky, Noam. 2000. Minimalist inquiries: the framework. In *Step by step: essays on minimalist syntax in honor of Howard Lasnik*, eds. Roger Martin, David Michaels & Juan Uriagereka, 89–155. Cambridge, MA: MIT Press.
- Chomsky, Noam. 2001. Derivation by phase. In *Ken Hale: a life in language*, ed. Michael Kenstowicz, 1–52. Cambridge, MA: MIT Press.
- Clements, George N. 1975. The logophoric pronoun in Ewe: its role in discourse. *Journal of West African Languages* 10:141–177.
- Cole, Peter, Gabriella Hermon, Yassir Tjung, Chang-Yong Sim & Chonghyuck Kim. 2007. *Anaphoric expressions in the peranakan javanese of semarang*. LINCOM Studies in Asian Linguistics 72, München: LINCOM.
- Coon, Jessica, Pedro Mateo Pedro & Omer Preminger. 2014. The role of case in A-bar extraction asymmetries: evidence from Mayan. *Linguistic Variation* 14:179–242, doi: <10.1075/lv.14.2.01coo>.

- Cowper, Elizabeth. 1987. Pied piping, feature percolation and the structure of the noun-phrase. *Canadian Journal of Linguistics* 32:321–338.
- Dowty, David R. & Pauline Jacobson. 1988. Agreement as a semantic phenomenon. In *Proceedings of the 5th Eastern States Conference on Linguistics (ESCOL 1)*, eds. Joyce Powers & Kenneth De Jong, Columbus, OH: Ohio State University, 95–108.
- Elbourne, Paul D. 2013. *Definite descriptions*. Oxford: Oxford University Press.
- Everaert, Martin. 1991. Nominative anaphors in Icelandic: morphology or syntax. In *Issues in Germanic syntax*, eds. Werner Abraham, Wim Kosmeijer & Eric Reuland, 277–305. Berlin: Mouton de Gruyter.
- Frampton, John. 1991. Relativized minimality: a review. *The Linguistic Review* 8:1–46.
- Frampton, John & Sam Gutmann. 2000. *Agreement is feature sharing*. Ms., Boston, MA: Northeastern University. URL: <[mathserver.neu.edu/~ling/pdf/agrisfs.pdf](http://mathserver.neu.edu/~ling/pdf/agrisfs.pdf)>.
- Frampton, John & Sam Gutmann. 2006. How sentences grow in the mind: agreement and selection in efficient minimalist syntax. In *Agreement systems*, ed. Cedric Boeckx, 121–157. Amsterdam: John Benjamins.
- Franks, Steven L. 2013. Bulgarian *nego si* is a Balkan anaphor. *Linguistique Balkanique* 52:3–28.
- Gazdar, Gerald, Ewan Klein, Geoffrey K. Pullum & Ivan A. Sag. 1985. *Generalized Phrase Structure Grammar*. Cambridge, MA: Harvard University Press.
- Grimshaw, Jane. 2000. Locality and extended projections. In *Lexical specification and insertion*, eds. Peter Coopmans, Martin Everaert & Jane Grimshaw, 115–134. Amsterdam: John Benjamins.
- Heck, Fabian. 2004. *A theory of pied-piping*. Doctoral dissertation, Tübingen: Universität Tübingen.
- Heinat, Fredrik. 2008. *Probes, pronouns and binding in the Minimalist Program*. Saarbrücken: VDM Verlag.
- Holmberg, Anders & Þorbjörg Hróarsdóttir. 2003. Agreement and movement in Icelandic raising constructions. *Lingua* 113:997–1019, DOI: <10.1016/S0024-3841(02)00162-6>.
- Hornstein, Norbert. 2018. *Case and agreement in a corner of Icelandic*. Ms. URL: <<https://ling.auf.net/lingbuzz/004273>>.
- Hubbard, Philip. 1985. *The syntax of the Albanian verb complex*. New York, NY: Garland Publishing.
- Kayne, Richard S. 1983. Connectedness. *Linguistic Inquiry* 14:223–249.
- Keine, Stefan. 2017. Agreement and vP phases. In *A schrift to fest Kyle Johnson*, eds. Nicholas LaCara, Keir Moulton & Anne-Michelle Tessier, 177–185. Amherst, MA: University of Massachusetts Linguistics Open Access Publications, URL: <[https://scholarworks.umass.edu/cgi/viewcontent.cgi?filename=19&article=1000&context=linguist\\_oapubs&type=additional](https://scholarworks.umass.edu/cgi/viewcontent.cgi?filename=19&article=1000&context=linguist_oapubs&type=additional)>.
- Keine, Stefan & Bhamati Dash. 2018. *The ups and downs of agreement*. Ms., Los Angeles, CA: University of Southern California.
- Kratzer, Angelika. 2009. Making a pronoun: fake indexicals as windows into the properties of pronouns. *Linguistic Inquiry* 40:187–237, DOI: <10.1162/ling.2009.40.2.187>.
- Legate, Julie Anne. 2008. Morphological and abstract case. *Linguistic Inquiry* 39:55–101, DOI: <10.1162/ling.2008.39.1.55>.
- Maling, Joan. 1984. Non-clause-bounded reflexives in Modern Icelandic. *Linguistics & Philosophy* 7:211–241, DOI: <10.1007/BF00627706>.
- Marantz, Alec. 1991. Case and licensing. In *Proceedings of the 8th Eastern States Conference on Linguistics (ESCOL 8)*, eds. German Westphal, Benjamin Ao & Hee-Rahk Chae, Ithaca, NY: CLC Publications, 234–253.
- Merchant, Jason. 2014. Gender mismatches under nominal ellipsis. *Lingua* 151:9–32, DOI: <10.1016/j.lingua.2014.01.008>.

- Middleton, Jane. 2018. *Everyone left the room, except the logophor: \*ABA patterns in pronominal morphology*. Paper presented at LinG1: Agreement and Anaphoricity, Göttingen.
- Murugesan, Gurujegan. 2018. *Merge matters for AAE*. Paper presented at LinG1: Agreement and Anaphoricity, Göttingen.
- Murugesan, Gurujegan. in prep. *Predicting the AAE and its violations*. Doctoral dissertation, Leipzig: Institut für Linguistik, Universität Leipzig.
- Pesetsky, David & Esther Torrego. 2007. The syntax of valuation and the interpretability of features. In *Phrasal and clausal architecture: syntactic derivation and interpretation, In honor of Joseph E.monds*, eds. Simin Karimi, Vida Samiian & Wendy Wilkins, 262–294. Amsterdam: John Benjamins.
- Polinsky, Maria. 2003. Non-canonical agreement is canonical. *Transactions of the Philological Society* 101:279–312, DOI: <10.1111/1467-968X.00120>.
- Polinsky, Maria & Omer Preminger. 2017. *The Agreement Theta Generalization*. Ms. URL: <<https://ling.auf.net/lingbuzz/003521>>.
- Pollard, Carl & Ivan A. Sag. 1994. *Head-driven Phrase Structure Grammar*. Chicago, IL: University of Chicago Press.
- Preminger, Omer. 2011a. *Agreement as a fallible operation*. Doctoral dissertation, Cambridge, MA: MIT.
- Preminger, Omer. 2011b. Asymmetries between person and number in syntax: a commentary on Baker's SCOPA. *Natural Language & Linguistic Theory* 29:917–937, DOI: <10.1007/s11049-011-9155-z>.
- Preminger, Omer. 2013. That's not how you agree: a reply to Zeijlstra. *The Linguistic Review* 30:491–500, DOI: <10.1515/tlr-2013-0015>.
- Preminger, Omer. 2014. *Agreement and its failures*. Linguistic Inquiry Monographs 68, Cambridge, MA: MIT Press, DOI: <10.7551/mitpress/9780262027403.001.0001>.
- Preminger, Omer. 2017. How can feature-sharing be asymmetric? Valuation as UNION over geometric feature structures. In *A Pesky Set: papers for David Pesetsky*, eds. Claire Halpert, Hadas Kotek & Coppe van Urk, 493–502. Cambridge, MA: MITWPL.
- Preminger, Omer. 2019. What the PCC tells us about 'abstract' agreement, head movement, and locality. *Glossa* 3:1–42, DOI: <10.5334/gjgl.315>.
- Preminger, Omer & Maria Polinsky. 2015. *Agreement and semantic concord: a spurious unification*. Ms. URL: <<https://ling.auf.net/lingbuzz/002363>>.
- Reuland, Eric. 2011. *Anaphora and language design*. Cambridge, MA: MIT Press.
- Rezac, Milan, Pablo Albizu & Ricardo Etxepare. 2014. The structural ergative of Basque and the theory of case. *Natural Language & Linguistic Theory* 32:1273–1330, DOI: <10.1007/s11049-014-9239-7>.
- Richards, Norvin. 2001. *Movement in language: interactions and architecture*. Oxford: Oxford University Press.
- Ritter, Elizabeth. 1991. Two functional categories in noun phrases: evidence from Modern Hebrew. In *Perspectives on phrase structure: heads and licensing*, ed. Susan D. Rothstein, 37–62. San Diego, CA: Academic Press.
- Ritter, Elizabeth. 1992. Cross-linguistic evidence for number phrase. *Canadian Journal of Linguistics* 37:197–218.
- Rizzi, Luigi. 1990a. On the anaphor-agreement effect. *Rivista di Linguistica* 2:27–42.
- Rizzi, Luigi. 1990b. *Relativized minimality*. Cambridge, MA: MIT Press.
- Rizzi, Luigi. 2001. Relativized minimality effects. In *The handbook of contemporary syntactic theory*, eds. Mark R. Baltin & Chris Collins, 89–110. Oxford: Blackwell Publishers, DOI: <10.1002/9780470756416.ch4>.
- Rooryck, Johan & Guido Vanden Wyngaerd. 2011. *Dissolving binding theory*. Oxford: Oxford University Press.
- Sells, Peter. 1987. Aspects of logophoricity. *Linguistic Inquiry* 18:445–481.

- Shibatani, Masayoshi. 1977. Grammatical relations and surface cases. *Language* 53:789–807, DOI: <10.2307/412912>.
- Sigurðsson, Halldór Ármann. 1993. Agreement as head visible feature government. *Studia Linguistica* 47:32–56.
- Starke, Michal. 2001. *Move dissolves into Merge: a theory of locality*. Doctoral dissertation, Geneva: University of Geneva. URL: <<https://ling.auf.net/lingbuzz/000002>>.
- Sundaresan, Sandhya. 2016. Anaphor vs agreement: a new kind of Anaphor Agreement Effect in Tamil. In *The impact of pronominal form on interpretation*, eds. Patrick Grosz & Pritty Patel-Grosz, 77–106. Berlin: Mouton de Gruyter.
- Sundaresan, Sandhya. to appear. Distinct featural classes of anaphor in an enriched PERSON system. In *Agreement in the minimalist program*, Open Generative Syntax, Language Science Press.
- Taraldsen, Knut Tarald. 1995. On agreement and nominative objects in Icelandic. In *Studies in comparative Germanic syntax*, eds. Hubert Haider, Susan Olsen & Sten Vikner, 307–327. Dordrecht: Kluwer Academic Publishers.
- Thráinsson, Höskuldur. 1976. Reflexives and subjunctives in Icelandic. In *Proceedings of the 6th meeting of the North East Linguistic Society (NELS 6)*, eds. Alan Ford, John Reighard & Rajendra Singh, Montreal, QC: McGill University and University of Montreal, 225–239.
- Thráinsson, Höskuldur. 1979. *On complementation in Icelandic*. Outstanding Dissertations in Linguistics, New York, NY: Garland Publishing.
- Thráinsson, Höskuldur. 1991. Long distance reflexives and the typology of NPs. In *Long distance anaphora*, eds. Jan Koster & Eric Reuland, 49–75. Cambridge: Cambridge University Press.
- Thráinsson, Höskuldur. 1992. A note on underspecification and binding. In, Harvard Working Papers in Linguistics 1, 73–90. Cambridge, MA: Harvard University.
- Thráinsson, Höskuldur. 2007. *The syntax of Icelandic*. Cambridge: Cambridge University Press.
- Webelhuth, Gert. 1992. *Principles and parameters of syntactic saturation*. Oxford: Oxford University Press.
- Woolford, Ellen. 1999. More on the Anaphor Agreement Effect. *Linguistic Inquiry* 30:257–287, DOI: <10.1162/002438999554057>.
- Yip, Moira, Joan Maling & Ray Jackendoff. 1987. Case in tiers. *Language* 63:217–250.
- Zaenen, Annie, Joan Maling & Höskuldur Thráinsson. 1985. Case and grammatical functions: the Icelandic passive. *Natural Language & Linguistic Theory* 3:441–483.

svn revision code: 10080