

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

Omer Preminger*

July 2019

Abstract

This paper is about the Anaphor Agreement Effect (AAE): the ban on φ -agreement with an anaphor. While the AAE has sometimes been taken to support a reduction of anaphoric binding to syntactic φ -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 φ -agreement, but because the φ -features of the anaphor are (always) encapsulated within a larger anaphoric structure. This 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 1990, Woolford 1999, *a.o.*) seems to suggest a rather tight interaction between syntactic φ -agreement on the one hand, and binding on the other. This has led some (e.g. Reuland 2011) to take the AAE as support for theories where syntactic φ -agreement is a necessary condition for binding. I refer to such theories as *reductionist*. I will show that upon closer inspection, the AAE in particular and anaphoric binding more generally provide fairly strong evidence against reductionist theories. That is because such theories require, somewhat paradoxically, assumptions about φ -agreement and about the structure of reflexive anaphors which break their compatibility with an agreement-based theory of binding, in the first place.

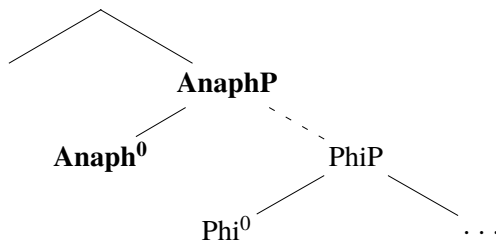
I start by re-examining how the AAE should be characterized. Building on observations by Woolford (1999) and Tucker (2011), I argue that Rizzi's (1990) original characterization—in terms of a restriction on the possible positions in which anaphors can occur—has little cross-linguistic merit. While it is true that certain languages ban the occurrence of anaphors in positions that would otherwise control agreement, many other languages (perhaps the vast majority) exhibit no such ban. Nevertheless, there is a constant underlying both patterns: true, nontrivial φ -agreement with anaphors is banned. Limiting the distribution of anaphors is but one way to achieve this; another is 'default' (i.e., non-alternating) agreement. This formulation, in terms of nontrivial φ -agreement with anaphors (and not in terms of their occurrence) is the only cross-linguistically viable statement of the AAE.

I then turn to purported interactions of binding with φ -agreement. First, I address the question of whether φ -feature matching between binder and bindee constitutes an argument for syntactic φ -agreement between the two. I argue that it does not. The argument is based on cases in which

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

φ -feature matching between the binder and bindee is enforced even in the absence of any syntactic relation between the two. These include cases of Donkey Anaphora, cases of cross-utterance (and cross-speaker) anaphora, and cases of linguistically-unantecedented deixis. These all provide evidence for φ -feature matching, even in uninterpreted φ -features like grammatical gender on inanimates, in the absence of any syntactic relation between the matched phrases. Once we come to terms with these facts, it becomes clear that whatever mechanism underpins these non-syntactic cases is also sufficient to ensure φ -matching under anaphoric binding, without implicating syntactic φ -agreement in any way. Next, I review Middleton’s (2018) morphologically-based argument that anaphors have a nested containment structure, the inner layer of which is pronominal. I argue that because this pronominal core is the locus of valued φ -features, anaphors necessarily involve φ -encapsulation: the φ -bearing portion of an anaphor is properly contained—universally and without exception—in an additional layer of structure that is responsible for the anaphoric nature of the expression:

(1) THE φ -ENCAPSULATION HYPOTHESIS



Given Middleton’s results, (1) must be the universal structure of anaphors, regardless of the particular morphology of the language in question. Thus, languages like Georgian, where anaphors are pronominally-possessed body-part nouns (as shown in (2)), are merely a more transparent reflection of the containment structure that all anaphors have, even in those languages where it is less clearly reflected in the morphology.

- (2) [[tʃems] tav-s] vakeb (Georgian)
 1sg.POSS head-DAT praise<1sgS,3sgO>
 ‘I praise myself.’ [Harris 1981:27]

I discuss the prima facie challenge that the AAE poses for reductionist theories of binding. I examine one solution to this challenge, in terms of the relative timing of φ -agreement and binding, and show that it does not work. I then show that similar facts actually militate against reductionism entirely (i.e., whether it is paired with a timing-based approach, or not).

I then propose an alternative account of the AAE (as construed above) in terms of φ -encapsulation alone, with no appeal to derivational timing or reductionism.

Finally, I turn to several outstanding issues. The first is the proper treatment of the handful of languages that disobey the AAE even when formulated in terms of agreement (cf. Murugesan 2018, 2019). The second is the aforementioned cross-linguistic variation in the fate of attempted agreement with anaphors: while in some languages (e.g. English, Icelandic, Italian) this results in ungrammaticality, in others (Albanian, Georgian) it does not, and the result is instead a grammatical utterance with a non-alternating agreement form. The third is divergences between finite, verbal φ -agreement on the one hand, and nominal and adjectival concord, on the other, when it comes to the AAE.

Section 12 concludes.

2. The Anaphor Agreement Effect

2.1. Is the AAE a distributional constraint, or a constraint on agreement?

Rizzi (1990) defines the AAE as follows:

(3) ANAPHOR AGREEMENT EFFECT (Rizzi 1990:26)

Anaphors do not occur in syntactic positions construed with agreement.

He exemplifies this condition with examples like (4). This example involves a dative-Experiencer verb, whose nominative Theme argument would normally control finite agreement. As Rizzi notes, it is precisely in this case that the nominative argument cannot be an anaphor in Italian. That is despite the fact that binding-theoretically, this should be a well-formed configuration. See Belletti & Rizzi 1988:335 for evidence, from long-distance anaphora, that the dative Experiencer is structurally higher than the nominative Theme.¹

- (4) * A voi interessate solo voi stessi. (Italian)
to you(DAT.PL) interest(2PL) only yourselves(NOM)
'You are interested only in yourselves.' [Rizzi 1990:34]

As noted by Woolford (1999) and Tucker (2011), however, Rizzi's definition of the AAE fails to generalize. There are languages where anaphors can surface in positions that would otherwise control agreement without causing ungrammaticality. Instead, the verb surfaces with invariant 3rd person singular morphology. To illustrate this, Woolford provides examples like (5), from Albanian. Much like (4), this example involves a dative-Experiencer verb whose nominative Theme typically controls finite agreement morphology on the verb. However, in contrast to (4), the occurrence of an anaphor in this position is not ruled out. Rather, the verb simply surfaces with 3rd person singular agreement morphology, regardless of the features of the antecedent (and, presumably, of the reflexive anaphor).

- (5) 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]

Anticipating some of the discussion in later sections, I will note that *vetja* in Albanian is a simplex anaphor (Franks 2013), whereas *voi stessi* in Italian (or its 3rd person singular counterpart, *se stesso*) is a complex one. This is noteworthy because the behaviors exemplified by (4) and (5) (ungrammaticality vs. a non-alternating, fixed agreement form) are crucially not predictable from the structure of the reflexive anaphor in question. The simplex anaphor in Icelandic behaves in this respect like the complex anaphor of Italian, whereas the complex anaphor of Georgian behaves like the simplex Albanian anaphor. (This will be discussed in greater detail in section 11.2.)

¹Glossing conventions used in this paper: 1: first person; 2: second person; 3: third person; ABS: absolutive; ACC: accusative; ANAPH: anaphor(ic); ART: article; ASSOC: associative; AUX: auxiliary; BEN: benefactive; CL: clitic; COMP: complementizer; COP: copula; DAT: dative; DET: determiner; ERG: ergative; F: feminine; FUT: future; HAB: habitual; IMPF: imperfective; INF: infinitive; INSTR: instrumental; LOC: locative; M: masculine; NEG: negation; NMZ: nominalization; NOM: nominative; PL/pl: plural; PAST: past; PROG: progressive; PRT: participle; REFL: reflexive; REL: relational; SBJV: subjunctive; SG/sg: singular; ϕ : a morphological slot which is phonologically empty; $\sqrt{\quad}$: lexical root.

(8) ANAPHOR AGREEMENT EFFECT

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

- (9) A φ -agreement relation between α and β is nontrivial if there are at least two sets of φ -feature values, F and F' , such that α takes on one form, $\alpha(F)$, when β bears F ; and takes on another, distinct form, $\alpha(F')$, when β bears F' .

If a language bans anaphors in positions that would otherwise control agreement (as Italian does), then (8) holds a fortiori. If a language allows anaphors to surface in the relevant positions, but the result is a non-alternating, fixed agreement form (as in Albanian or Chichewa), (8) still holds.

For the remainder of this paper, I will concentrate on (8) as a target of explanation (though I will briefly return to the issue of Italian-like vs. Albanian/Chichewa-like behavior in section 11.2).

2.2. Agreement, not case

In this subsection, I review Rizzi's (1990) argument that the AAE should indeed be described in terms of φ -agreement, and not in terms of case. I do this both for the sake of completeness, and because it constitutes one of the premises for the argument in the next subsection, that the AAE is a properly syntactic phenomenon, rather than a morphological one.

The argument is most clearly made by juxtaposing a language like Icelandic with a language like Japanese. 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).⁴ The occurrence of SIG in (10) clearly adheres to the AAE, as the embedded subjunctive verb agrees with the NOM subject (*þú* “you”), not with SIG:

- (10) Jón_i heldur [að þú hatir sig_i] (Icelandic)
John believes that you hate.SBJV.2sg SIG.ACC
'John_i believes that you hate him_i.' [Thráinsson 2007:467]

Contrast this with (11): here, the embedded verb is a quirky-subject verb, whose subject is the dative *mér* (“me.DAT”), and whose object is nominative. As in many other languages, finite agreement in Icelandic tracks case, not grammatical function. Thus, it is the nominative object, not the dative subject, that controls finite agreement in this instance. Crucially, SIG is now banned from occurring in the embedded object position, even though (11) is binding-theoretically equivalent to (10).

- (11) Sigga_i telur [að mér líki hún_i/*sig_i]
Sigga thinks that me.DAT likes.SBJV.3sg she.NOM/*SIG
'Sigga_i thinks that I like her_i.' [Maling 1984:217]

⁴It 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 (8). 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 (8) 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.

As one might expect, SIG can serve as an argument of a quirky-case verb, just as long as it is not the nominative argument and is therefore not a putative controller of finite agreement. An example is given in (12).

- (12) Hún_i sagði [að sig_i vantaði peninga].
 she.NOM said that SIG.ACC lacked.SBJV.3sg money
 ‘She_i said that she_i lacked money.’ [Maling 1984:216]

Because finite agreement in Icelandic tracks nominative case, data like (10–12) *could* be characterized in terms of case (*viz.* “no nominative anaphors”), rather than in terms of agreement. As Woolford (1999:262ff.) points out, however, such a characterization would fail to generalize. Consider an example from Japanese:

- (13) sensei_i-ni(-wa) zibun_i-ga wakar-ani-i (Japanese)
 teacher-DAT(-TOPIC) self-NOM understand-NEG-PRES
 ‘The teacher_i does not understand herself_i.’ [Shibatani 1977:800, via Woolford 1999:263]

As (13) shows, Japanese allows anaphors bearing nominative case (assuming that their occurrence is binding-theoretically well formed). A case-based characterization of the AAE, while it would account for the Icelandic data above, would make the wrong predictions for data like (13). Importantly, the latter pattern is quite general: languages that lack overt φ -agreement morphology seem to be systematically exempt from the AAE.

It is for these reasons that researchers (going back to Rizzi 1990) have chosen to characterize the AAE in terms of agreement, not case.

2.3. Syntax, not morphology

The AAE as characterized in (8), even without further analysis, already has implications for the way we think about the syntax and morphology of φ -agreement, cross-linguistically.

Consider once more the contrast between the Icelandic (11), where a nominative anaphor is banned in a position that would otherwise control overt finite agreement, and the Japanese (13), where a nominative anaphor is licit in a very similar configuration but in a language that lacks overt φ -agreement morphology. To maintain an account of (11) without incorrectly ruling out (13), it is necessary for at least one of the following conditions to hold: (i) Japanese, Korean, and all other languages like them do not have an abstract, morpho-phonologically unexpressed version of the agreement relation in (11); or (ii) the AAE is a morphological constraint, rather than a syntactic one.

To see why, suppose that both (i) and (ii) were false. What would distinguish the nominative anaphor in (11) from its counterpart in (13)? It cannot be the presence of overt agreement with the anaphor (in (11) but not in (13)). That is because 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 (11), there would be no explanation for why the anaphor in (13) is licit while the one in (11) 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) agreement between a φ -probe, H^0 , and an anaphor, DP_{ANAPH} (see (8–9), above, and the surrounding discussion). No module of grammar could be the locus of AAE effects, then, unless it was able

to hold both H^0 and DP_{ANAPH} within a single domain of computation, to examine whether the offending agreement relation had indeed been established. Crucially, however, φ -agreement is *unbounded*: there is no upper bound on the amount of structure or linear distance that the relation can span.⁵ For a particularly vivid demonstration of this fact, I refer the reader to Keine (2017), who shows that long-distance agreement (LDA) in Hindi can span across an unbounded number of transitive verb phrases, and moreover, that it is subject to exactly the same structural constraints as the licensing of *wh*-in-situ in Hindi is. If ‘morphology’ differs from ‘syntax’ in any contentful way, surely unbounded dependencies are the purview of the latter, not the former.⁶ 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.

If at least one of (i)–(ii) must be true, and (ii) (the claim that the AAE is morphological in nature) is false, it follows that (i) is necessarily true: languages like Japanese must lack any finite φ -agreement relation whatsoever. In particular, they do not have a morpho-phonologically unexpressed counterpart of the φ -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 very same conclusion (Preminger 2019).

3. Reductionist theories of binding

Recent years have seen a proliferation of proposals that subscribe to what I will call a *reductionist* view of binding, as formalized in (14):

(14) THE REDUCTIONIST POSITION

α and β can share a binding index **only if** α and β have entered into syntactic agreement in φ -features

- where, for the purposes of this definition, entering into φ -agreement is subject to transitive closure
 - i.e., if α agrees with γ , and γ agrees with β , then α and β count as having entered into agreement with one another for the purposes of this definition

Some of these proposals apply reductionism only to anaphoric binding (Heinat 2008, Reuland 2011, Rooryck & Vanden Wyngaerd 2011); while others apply it to any kind of binding relation (Kratzer 2009, Wurmbrand 2017; see also the discussion of binding in Pollard & Sag 1994).

⁵This is not to be confused with the (obviously false) claim that there is no structure that would block φ -agreement. Much like *wh*-movement is unbounded but is stopped by certain structures (syntactic islands), so too is φ -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.

⁶While there have been recent proposals that allow morphology to traffic in objects like ‘chains’, ‘copies’, ‘traces’, etc. (cf. Bobaljik 2008, Marantz 1991), modular separation should entail *some* difference in the sets of primitives available to each module. That is not to say that the two sets should be disjoint: there must be some overlap in the primitives of syntax and morphology, otherwise the output of one would be wholly illegible to the other. Heads and their features seem like good candidates to fill this role of “shared vocabulary” between syntax and morphology. But if there was ever a candidate for a primitive that is syntactic but not morphological, it would be the ‘chains’ / ‘copies’ / etc. formed by syntactic movement. So either there is literally no difference between syntax and morphology (in which case (ii) is definitionally false), or else there is, and (ii) is false for the reasons outlined in the text.

Now, there is plenty of evidence unrelated to the AAE indicating that reductionist approaches are off track. While it is not the purpose of the present paper to rehash these arguments, I will briefly mention three of them here. One source of evidence concerns *directionality*. It is well established at this point that φ -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 to appear, Preminger 2013, Preminger & Polinsky 2015, *a.o.*).⁷ Anaphoric binding, in contrast, transmits values in the other direction, from a c-commanding antecedent to a c-commanded bindee. Thus, the idea that anaphoric binding is underpinned by φ -agreement is antithetical to the apparent structural properties of the two relations.⁸

Another source of evidence against reductionist theories has already been mentioned here, and concerns the ban on syntactic φ -agreement that is morpho-phonologically null (Preminger 2019). Recall that the AAE itself already implies that morpho-phonologically covert agreement is banned (section 2.3), a result that enjoys independent support from the distribution and nature of PCC effects (Preminger 2019). It is a truism, however, that anaphoric binding exists even in languages that lack overt φ -agreement—Japanese—which, by hypothesis, lack syntactic φ -agreement as well.⁹ Japanese, after all, does have anaphors. Thus, the idea that anaphoric binding is underpinned by φ -agreement is antithetical to the ban on null agreement, which is required for the AAE to even be statable with any generality.¹⁰

Finally, it is worth noting a peculiar (but often overlooked) aspect of the reductionist position: in reality, φ -feature matching is neither a necessary nor sufficient condition for coreference, as demonstrated in (15) and (16), respectively.

(15) Only the present authors think we know how to do syntax. [Collins & Postal 2012:253n1]

(16) <pointing to different individuals in succession>

You should leave, but you should stay here.

It is logically possible that φ -identity, while not required for coreference, is nevertheless required for binding. But a strict connection between φ -features and binding would need to be argued for quite rigorously, given that no such strict connection holds between φ -matching and coreference.

⁷See also Bjorkman & Zeijlstra (2019), who, by making explicit the vast and problematic array of assumptions that would be required to support a theory in which φ -feature values were transmitted downward in the structure, inadvertently provide a fairly strong argument against such a theory. For example, Bjorkman & Zeijlstra's theory must assume that all φ -agreement includes a *checking* component alongside *valuation*—an assumption that is known to be false (Preminger 2011, 2014).

⁸Not *every* reductionist approach runs afoul of this directionality consideration. For example, the aforementioned work by Kratzer (2009) does, but the aforementioned work by Reuland (2011) and by Rooryck & Vanden Wyngaerd (2011) does not—as it makes use of syntactic movement to ensure that all agreement relations that are well-behaved from a directionality perspective (*i.e.*, relations where the recipient of values c-commands the supplier of values).

⁹A reviewer points out that there are languages that lack morpho-phonologically overt φ -agreement and yet manifest AAE or AAE-like effects (*e.g.* Mainland Scandinavian; Woolford 1999:283n31). While I leave the ultimate resolution of this matter for future research, I will note here that a full retreat from the ban on null agreement lands us back in the conundrum identified in section 2.3 concerning why the AAE applies in, *e.g.*, Icelandic but not in, *e.g.*, Japanese.

¹⁰Sandhya Sundaresan (*p.c.*) points out that it is technically possible to maintain the ban on morpho-phonologically null finite φ -agreement, while assuming that φ -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.

Moreover, Collins & Postal report that a bound-variable (i.e., covarying) reading is available for *we* in (15), suggesting that it is not only coreference that is dissociable from φ -matching, but binding as well.¹¹

All that being said, the focus of the remainder of this paper is different. I will present a separate argument, from the AAE, against reductionism, followed by a non-reductionist accounts of the AAE.

4. Step I: φ -matching $\not\Rightarrow$ φ -agreement

One of the arguments frequently proffered in support of reductionist theories of binding is the rather general pattern of matching in φ -features between binder and bindee (though see (15–16) and the surrounding discussion). In actuality, however, φ -feature-matching provides no support whatsoever for an involvement of φ -agreement in binding. That is because, as we will see shortly, φ -matching is enforced even in scenarios where syntactic φ -agreement could not possibly be at play. And this is so even when it comes to those φ -features that are not interpreted (e.g. grammatical gender on inanimates). Whatever mechanism is responsible for such cases would suffice to enforce φ -matching even in syntactically local cases, dissolving φ -matching as a source of evidence for the involvement of φ -agreement in binding.

The scenarios in question include cases of Donkey Anaphora, cases of cross-utterance (and cross-speaker) anaphora, and even linguistically-unantecedented deixis. Consider (17–18):

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

(Hebrew)

(18) a. kol exad [e-ye]f l-o maxberet [e-ya-sim] ot-a/*ot-o
 every one that-EXIST DAT-3sgM notebook<F> that-3sgM.FUT-put ACC-3sgF/*ACC-3sgM
 ba-tik
 in.the-case

‘Everyone who has a notebook<F>, put it.F/*it.M in your bag ’

b. kol exad [e-ye]f l-o max[evon] [e-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 ’

These English and Hebrew sentences are examples of Donkey Anaphora: on the intended reading, the underlined expressions covary. This, despite the absence of *c*-command (in either direction) between the covarying expressions, and despite the fact that the antecedents are located within a Complex NP Island (of the relative-clause variety), which is itself buried inside a Subject Island.

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. Cases like (17–18), then, are a clear indication that even φ -features that are not interpreted, such as the grammatical gender features of inanimates, can be transmitted by some non-syntactic mechanism. Let us refer to this mechanism—whatever it may be—as *NSM*.

¹¹Thanks to Itai Bassi and Jason Merchant for helpful discussion. See also Podobryaev (2014).

There are some candidates in the literature for what *NSM* might be. For Elbourne (2013), pronouns are (hidden) definite descriptions. On this view, the matching requirements evidenced in (17–18) dissolve into whichever 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 the ones in (17–18) are the residue of NP ellipsis. On this view, the attested matching requirements dissolve into whatever it is that enforces identity of form between ellipses and their antecedents (cf. **John is no longer a bachelor, and Bill did <get married> too*).

For our current purposes, however, it is not terribly important what *NSM* is; its modular affiliation is clearly not syntax, and that is the only relevant point here.

Donkey Anaphora is far from the only evidence for the existence of *NSM*. Consider cases of cross-utterance anaphora, like (19). These exhibit φ -matching across different utterances, and even across different speakers. It would be logically incoherent to speak of a syntactic relation (be it Agree or something else) holding between *they* in Speaker B’s utterance and *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.¹²

(19) A: Where are the scissors?

B: They are right here.

As before, the same holds of grammatical gender on inanimates. This is demonstrated in (20) using a case of linguistically-unantecedent deixis:

(20) <pointing to a pair of saloon-style doors>

ir-a-kingu-ye

(Kinyarwanda)

4SUBJ-PAST-OPEN-PRFV

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

The subject-agreement marker in (20) is *ir-* because (plural) “doors” are a member of noun-class 4 (rather than 2, 6, 8, and so forth). Again, there seems to be no coherent way that facts like these could be attributed to syntactic agreement in φ -features.

Overall, the data surveyed in this section illustrate quite vividly that there is absolutely no argument to be had from overt φ -feature matching to the involvement of syntactic φ -feature agreement. That is because these data demonstrate the need for a mechanism that can enforce φ -feature matching in the absence of c-command; across island boundaries; across different speakers; and even in the absence of an explicit linguistic antecedent. Whatever mechanism underlies these cases, it is clearly not a syntactic one. Consequently, the same mechanism is sufficient to ensure φ -matching between an anaphor and its binder, as well, with no involvement of syntactic φ -feature agreement.

¹²It is possible, of course, that the actual relation of interest holds not between *they* in Speaker B’s utterance and *the scissors* in Speaker A’s utterance, but between *they* and some mental representation of (parts of) speaker A’s utterance that speaker B has constructed. But then we must ask what mechanism ensures that the mental representation that speaker B builds maintains the same φ -features used in speaker A’s original utterance (as opposed to, say, speaker B building a representation that replaces *scissors* with *object*). The answer to that question cannot be ‘syntax’, and that establishes once again the need for a non-syntactic mechanism (*NSM*) of one kind or another that is involved in the relevant mediation.

This does not constitute an argument *against* the involvement of syntactic φ -feature agreement in anaphoric binding, of course. But it does remove one very common argument in favor of such involvement, namely the purported argument from φ -matching.

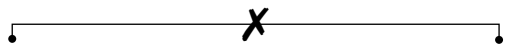
5. Interlude: Could NSM replace syntactic φ -agreement entirely? (No.)

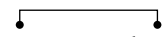
The previous section established the necessity of a non-syntactic mechanism, which I labeled *NSM*, capable of enforcing φ -feature matching between binder and bindee. In light of these results, one might entertain a putative swing of the pendulum all the way in the other direction, towards the possibility that *NSM* could subsume all of what syntactic φ -agreement was posited for in the first place. Dowty & Jacobson (1988), discussing a related set of facts, hint (at the very least) at this very possibility.¹³

In a vacuum, one operation is always better than two. But there is good evidence that φ -agreement proper—that is, φ -feature agreement between a verbal head and the verb’s argument(s)—must be syntactically underpinned. Consider, first, the configuration in (21), where ‘ \gg ’ represents c-command, and H is some functional head:

(21) H \gg DP₁ \gg DP₂

As discussed extensively by Abels (2012), Béjar & Rezac (2009), Boeckx & Jeong (2004), Nevins (2007), Preminger (2014), Rizzi (2001) and Starke (2001), H cannot enter into φ -agreement with DP₂ across DP₁ in this case.¹⁴ Importantly, such intervention in syntactic φ -agreement is systematical ameliorated by A-moving DP₁ out of its intervening position (Anagnostopoulou 2003, Holmberg & Hróarsdóttir 2003, *a.o.*). This is demonstrated in the Icelandic (22–23): intervention by *einhverjum stúdent* in (22) is ameliorated by A-movement of this DP to a position not c-commanded by the finite verb, as in (23).

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

(23)  [Holmberg & Hróarsdóttir 2003:999–1000]
 [Einhverjum stúdent]₁ finnst t₁ tölvurnar ljótar
 some student.DAT find.PL computers.the.NOM ugly
 ‘Some student finds the computers ugly.’

Binding, on the other hand, exhibits a markedly different pattern of behavior. Consider the following sentences:

¹³For Dowty & Jacobson, *NSM* would reside in the semantic-pragmatic component, but that is not crucial for our current discussion.

¹⁴This phenomenon of intervention in φ -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 DP₂ is not a candidate to A-move, regardless of whether DP₁ is there or not, and only *agreement* between H and DP₂ is at stake (e.g. in Basque or Icelandic). Bruening and Branan are mum on such cases.

- (24) a. The children₁ seem to her_{j/*i} to have *t*₁ amused Mary_i.
 b. * The children₁ seem to me to have *t*₁ amused myself.
 c. The children₁ seem to me to have *t*₁ amused each other.

We know from (24a) that the *to*-Experiencer of *seem* can bind into the embedded infinitive, at least when locality is not at issue (as is the case with Condition C). What, then, is the reason for the failure of reflexive binding in (24c)? If it is because of the intervening A-trace of *the children*, this shows that A-traces count for binding.

If, on the other hand, it is because the embedded infinitive in (24) constitutes its own binding domain, then in (24b), there must be a binder within that embedded binding domain to satisfy Condition C with respect to *each other*. The only viable candidate to serve as that binder is the A-trace of *the children*, showing once again that A-traces must count for binding.

Either way, the conclusion is the same: A-traces count for binding, whereas we've seen above that they do not count for φ -agreement.¹⁵

Furthermore, φ -agreement is sensitive to the case of its operands. It is very common for a DP to be inaccessible to finite agreement unless it is a nominative or absolutive DP (see Bobaljik 2008, and references therein). Binding, on the other hand, does not care about case. For example: 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 (Zaenen, Maling & Thráinsson 1985).

Perhaps most importantly, however: syntactic φ -agreement can never do anything like (17–18) (the Donkey Anaphora cases presented earlier). Verbs—or, more accurately, the functional heads that verbs typically occur with—can agree with DPs that are not their arguments (e.g. in cases of raising-to-subject in English; or in cases of raising-to-ergative in Basque, as shown by Artiagoitia 2001, Rezac, Albizu & Etxepare 2014). They can agree with DPs in other (lower) clauses (see Bhatt & Keine 2017, Polinsky 2003). But they cannot blatantly ignore the contours of syntactic structure (c-command, islands, etc.), in the manner shown in (17–18).

Overall, then, a Dowty & Jacobson (1988)-inspired swing in the other direction, whereby even φ -agreement would be handled by *NSM*, would not accord with the facts.

6. Step II: Evidence for φ -encapsulation

In this section, I review a recent argument by Middleton (2018) (henceforth, M18), showing that anaphoric expressions have a cross-linguistically universal containment structure. As we will see, from this universal structure one can then deduce that the outermost layer of an anaphor is not the one associated with valued φ -features.

M18 surveys 86 languages, from 13 different languages families, looking specifically at the forms that each language uses to express the four meanings in (25a–d) (the given labels are M18's).

¹⁵I thank a reviewer for helping me sharpen this argument.

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

The key finding is a *ban on discontinuous syncretism* (cf. Bobaljik 2012, Bobaljik & Sauerland 2018, Caha 2017). To see what this means, we can adopt, as M18 does, the notational convention used by Bobaljik (2012). For each language, let us represent the forms used to express the meanings in (25a–d) by going from the bottom of the paradigm upwards, and assigning a new capital letter every time a new form enters the paradigm.¹⁶ English, then, could be said to exhibit an “AAAB” pattern, e.g. *them* (“A”), *them* (“A”), *them* (“A”), *themselves* (“B”). This is illustrated in greater detail in (26):

(26) English: **ABBB**

- a. Diana thinks that only Charles loves HIMSELF.
 ~ Diana $\lambda x(x \text{ thinks that only Charles } \lambda y(y \text{ loves } \boxed{y}))$
- b. Diana thinks that only Charles loves HER.
 ~ only Diana $\lambda x(x \text{ thinks that Charles } \lambda y(y \text{ loves } \boxed{x}))$
 ~ only Diana $\lambda x(x \text{ thinks that Charles } \lambda y(y \text{ loves } \boxed{z}))$, where $z = \text{Diana}$
 ~ only Diana $\lambda x(x \text{ thinks that Charles } \lambda y(y \text{ loves } \boxed{z}))$, where $z \neq \text{Diana}$

Icelandic, in contrast, exhibits an “ABCC” pattern (as does Malayalam, as well as certain varieties of Mandarin). This is illustrated in (27):

¹⁶At this juncture, there is no substantive reason for proceeding in this direction, rather than going from the top downwards. As we will see, however, M18 ultimately argues for a containment structure in which every category in (25a–d) properly contains the structure associated with the category below it. Consequently, starting from the bottom and proceeding upwards maximizes the parallelism with the way Bobaljik (2012) discusses attributive-comparative-superlative adjectival morphology.

(27) Icelandic: **ABCC**

[Jane Middleton, p.c.]

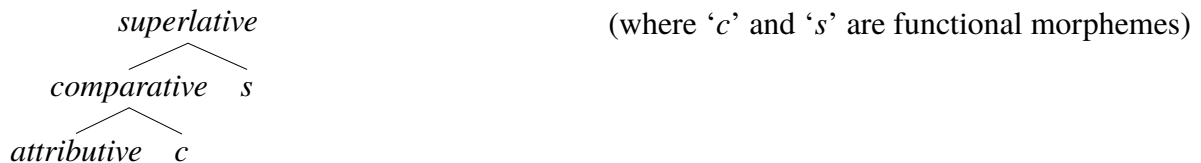
- a. Díaná telur að aðeins Karl elski sjálfan sig
Diana believes that only Charles loves SJALFAN SIG
~ Diana $\lambda x(x$ thinks that only Charles $\lambda y(y$ loves $\underline{y}))$
- b. Aðeins Díaná telur að Karl elski sig
only Diana believes that Charles loves SIG
~ only Diana $\lambda x(x$ thinks that Charles $\lambda y(y$ loves $\underline{x}))$
- c. Aðeins Díaná telur að Karl elski hana
only Diana believes that Charles loves HANA
~ only Diana $\lambda x(x$ thinks that Charles $\lambda y(y$ loves $\underline{z}))$, where $z =$ Diana
~ only Diana $\lambda x(x$ thinks that Charles $\lambda y(y$ loves $\underline{z}))$, where $z \neq$ Diana

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:

(28) *unattested*: ***ABBA**, ***AABA**, ***ABCA**, ***ABCB**, ***ABAC**, etc.

Recent work on similar bans has revealed two ways in which such a pattern could arise. The first is a containment structure (Bobaljik 2012, Caha 2009, 2013). As an example, Bobaljik (2012) discusses an ***ABA** ban, found in the domain of attributive-comparative-superlative adjectival morphology, and derives it as in (29):

(29) ADJECTIVAL CONTAINMENT HYPOTHESIS



If the comparative form is created by affixing the functional morpheme *c* to the attributive, and the superlative properly contains the comparative, then it follows that *c* will be present in the superlative, as well. If the comparative is suppletive relative to the attributive, then *c* must be a trigger for contextual allomorphy of the attributive stem. And since *c* is also present in the superlative, the same allomorphy will also be triggered, there. Alternatively, *s* could itself be an allomorphy trigger, in which case it would give rise to a third, distinct form for the superlative; in other words: an **ABC** pattern.

On these assumptions, the only way to create an **ABA** pattern is as a pathological subcase of **ABC**, in which **C** is accidentally homophonous with **A** (while both remain distinct from **B**). Bobaljik stipulates that there is a weak type of anti-homophony at play. This type of anti-homophony does not ban accidental homophony outright (that would obviously be far too strong, given, e.g., the existence of open-class homophones like *bank*). Instead, it bans context-sensitive allomorphs of a given morpheme from being accidentally homophonous with the 'elsewhere' allomorph of the same morpheme:

(30) ANTI-HOMOPHONY: Bobaljik 2012 version

A context-sensitive allomorph of μ cannot be homophonous with the default exponent of μ .

[Bobaljik 2012:35]

In the pathological derivation of **ABA** as a subcase of **ABC**, the shape of the ‘elsewhere’ allomorph of the *attributive* in (29) would have to be homophonous with the *s*-conditioned allomorph of that same morpheme, in violation of (30).

As noted earlier, however, a containment structure is not the only way in which a ban on discontinuous syncretism could arise. It can also arise through partial featural overlap (Caha 2017). Focusing again on ***ABA** as an example, consider (31):

(31)

EXPONENT	FEATURAL MAKE-UP
α	[<i>f</i>]
β	[<i>f</i> <i>g</i>]
γ	[<i>g</i>]

If β is different from both α and γ , the only way for α and γ to be identical is to have accidentally-homophonous spellouts for *f* and for *g* in the absence of the other feature. If we adopt a slightly different version of anti-homophony, given in (32), then the featural configuration in (31) yields a ***ABA** ban.

(32) ANTI-HOMOPHONY: featural-overlap version

A single morpheme μ cannot have two form-identical allomorphs that are triggered by disjoint featural contexts.

While having two competing routes to the same surface pattern is generally taken as an indication of a problem in the theory, this may not be the case here. That is because there are specific empirical domains where one or the other structure is transparently realized, lending support to the existence of both of these routes.

As noted earlier, Bobaljik (2012) shows that attributive-comparative-superlative morphology across languages exhibits a ban on discontinuous morphology. Crucially, certain languages, such as Persian, show a transparent containment structure in this empirical domain:

(33) Persian: TRANSPARENT CONTAINMENT

<i>bozorg</i>	“big”
<i>bozorg-tar</i>	“bigger”
<i>bozorg-tar-in</i>	“biggest”

Caha (2017) shows that the morphology associated with dative, allative, and locative cases also shows a cross-linguistic ban on discontinuous syncretism (i.e., ***ABA**). In contrast to attributive-comparative-superlative morphology, however, here we find transparent realization of the partial featural overlap structure, as exemplified by Tigrinya:

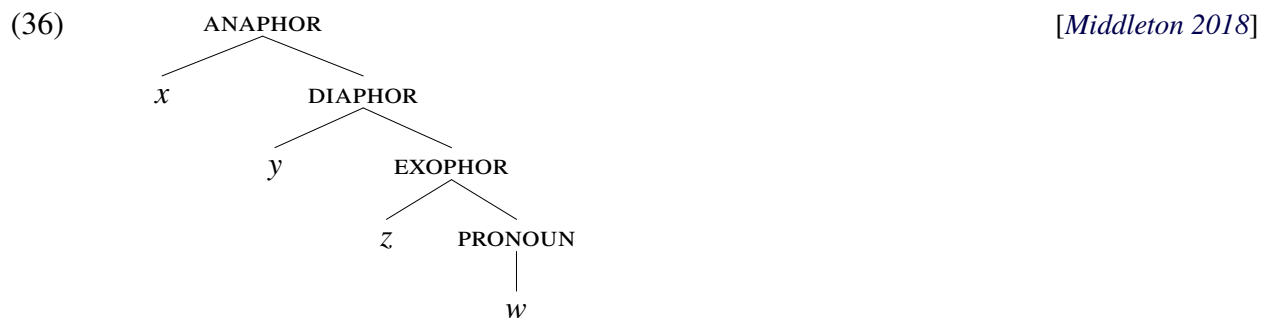
(34) Tigrinya: TRANSPARENT FEATURAL OVERLAP

CATEGORY	EXPONENT
DATIVE	ne
ALLATIVE	ne + ab
LOCATIVE	ab

Against this backdrop, M18 shows that one can find (near-)transparent reflection of a containment structure in the pronoun-exophor-diaphor-anaphor data. The case in point comes from Peranakan Javanese of Semarang (henceforth, PJS), and is based on the work of Cole et al. (2007):

- (35) ANAPHOR *awake dheen dhewe* (PJS)
 DIAPHOR/EXOPHOR *awake dheen*
 PRONOUN *dheen* [Cole et al. 2007]

As M18 discusses in greater detail, this pattern is not only suggestive of a containment structure; it in fact rules out an account of the ban on discontinuous syncretism in this empirical domain in terms of partial featural overlap (see M18:12–21). This disambiguates among the competing analyses of the cross-linguistic pattern at hand, in favor of the containment structure given in (36):



Let us note two important aspects of (36). First, it is necessarily the case that (36) represents a *universal*, cross-linguistically invariant structure for the expressions in question. If we relax the assumption, the explanation for why there is not a single example of discontinuous syncretism in M18’s (fairly vast) survey is lost. Second, suppose all we were interested in was the locus of valued φ -features relative to the structure responsible for anaphoricity. In that case, (36) reduces to (37):



The reasons (36) reduces to (37) are as follows. First, the φ -bearing layer of (36) can be no higher than PRONOUN, given that pronouns can bear the full range of available φ -features. Second, the categories that are subject to the AAE are what M18 classifies as DIAPHOR and ANAPHOR (cf. Rizzi

1990, Woolford 1999). Thus, (37) is the partial structure of (36) relevant to the relation between φ -features and anaphoricity.

What we have seen in this section, then, is that valued φ -features are *properly contained* within the structural layer that turns an expression into an anaphor. I refer to this as the **φ -encapsulation hypothesis**.

7. The challenge to reductionism from the AAE

Recall the reductionist position, repeated here:

- (38) THE REDUCTIONIST POSITION [=(14)]
 α and β can share a binding index **only if** α and β have entered into syntactic agreement in φ -features
- where, for the purposes of this definition, entering into φ -agreement is subject to transitive closure
 - i.e., if α agrees with γ , and γ agrees with β , then α and β count as having entered into agreement with one another for the purposes of this definition

Taken at face value, (38) predicts the systematic absence of anything like the AAE. The reason is as follows: anaphors are, by definition, bound; according to (38), binding entails φ -agreement; and an antecedent is, by definition, a target with valued φ -features.¹⁷ Successful, nontrivial agreement with anaphors should therefore be the natural state of affairs, in direct opposition to (39):

- (39) ANAPHOR AGREEMENT EFFECT [=(8)]
 $* H^0 \dots DP_{\text{ANAPH}}$, where \mathcal{R} is a nontrivial φ -agreement relation
-

Two approaches suggest themselves in response to this apparent paradox, and the next two sections will address these approaches in turn: an approach based on derivational timing (section 8), and an approach based on structural encapsulation (section 10).

8. A timing-based approach to the AAE

One way around the apparent contradiction, identified in section 7, between reductionism and the AAE, is to assume that the anaphor has not yet acquired valued φ -features *at the derivational stage at which it is targeted for φ -agreement*. On this view, anaphors begin their derivational life in a “ φ -deficient” state; and they acquire φ -feature values via the very φ -agreement relation implicated in the reductionist position (38).

¹⁷A possible exception involves cases where the binder of an anaphor is another anaphor, as in (i):

(i) Mary_i expected herself_i to outdo herself_i.

Nevertheless, even in such cases, this daisy-chain of bound-and-binding anaphors must ultimately resolve in a non-anaphoric antecedent if it is to be well-formed. And since (38) is closed under transitivity, all the anaphors in this scenario will ultimately have entered into φ -agreement with the non-anaphoric antecedent, by definition (see Sundaresan 2018).

2010 and Heck 2004 for arguments against this mechanism). Alternatively, it could arise if the head noun *buru* behaved in a manner analogous to pseudo-partitives, where the outermost extended nominal projection is transparent for φ -agreement (cf. *There are_{pl} [a bunch_{sg} of [children_{pl}]] here*). The choice between these options is not crucial for the discussion that follows.

8.1.2. A note on NUMBER in Basque reflexives

In cases where the binder (and consequently, the possessive pronoun within the reflexive anaphor) is plural, the head noun can *optionally* be plural, as well. (E.g. in the absolutive: *buru-ak* “head-ART_{pl}(ABS)” alongside *buru-a* “head-ART_{sg}(ABS)”); see *Artiagoitia 2003:621* for discussion.) In this case, the anaphor will indeed govern plural agreement where applicable—though, crucially, it will still be 3rd person agreement, regardless of the person features of the antecedent and of *buru*’s possessor. This can be seen, for example, in the finite auxiliary *ditugu* in (42b), which carries 1st person plural ergative agreement but 3rd person plural absolutive agreement (*d-it-u-gu* “3.ABS-pl.ABS-√-1pl.ERG”).

- (42) a. Orduan, etxetiarrek *beren buruak* enganaturik, beren baithan erraiten dute:
 then tenants.ERG *their heads.ABS* deceive.PRT their inside.LOC say.IMPX AUX
 Aurthen hemen naiz, . . .
 this.year here AUX
 ‘Then, the tenants, deceiving themselves, say in their mind: I am here this year, . . .’
- b. Geure egiazko irudia atzendu edo zaigu, eta amets irudipenezko argitan
 our true.INSTR.REL picture forget sort AUX and dream image.INSTR.REL light.LOC
 ikusten ditugu behialako *geure buruak*
 see.IMPX AUX long.REL *our heads.ABS*
 ‘We have kind of forgotten our true picture, we see our old selves in the light of dreamy images.’
[*Artiagoitia 2003:621*]

Notice, however, that *buru* can still be singular even when the antecedent and possessive pronoun are plural (cf. (43) below, as well as (40a–b), earlier).

- (43) Kirolari hauek *beren burua* erakustera etorri dira
 athlete these.ERG *their head(sg).ABS* show.NMZ.ALL come AUX
 ‘These athletes came just to show off.’
 (*lit.* ‘These athletes came to exhibit themselves.’)
[*Artiagoitia 2003:621*]

The number features expressed on *buru*, then, are quite clearly not the same entity as the number features associated with the binding index (as that would make the wrong prediction for (43)). Rather, what is going on in examples like (42a–b) probably is, or at least closely resembles, the phenomenon of “dependent plurality”—as demonstrated in the English (44), for example. I therefore leave this issue aside for the present purposes.¹⁸

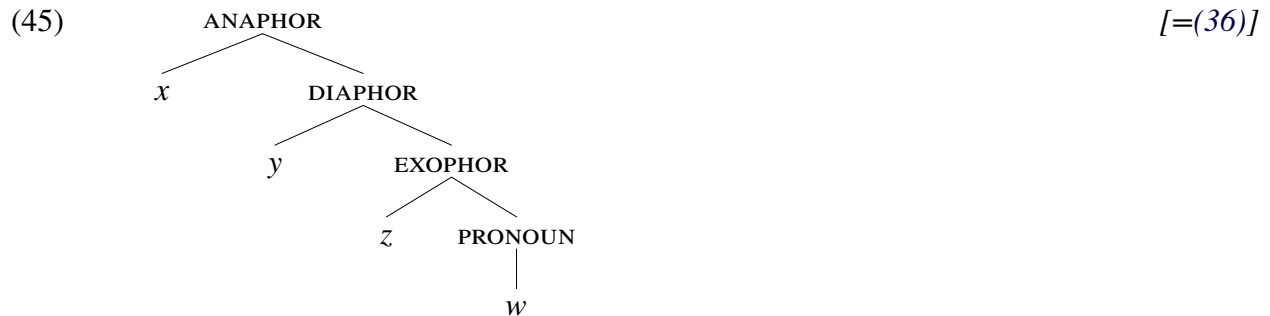
- (44) The plastic surgeons gave each other a new nose / new noses.

¹⁸A reviewer suggests an alternative view of this fact, one where it is taken to support a definition of the AAE in terms of PERSON features alone (see *Abramovitz 2019*, who develops a variant of the current proposal along those very lines). Such an approach, however, would be on the hook to explain why plural number surfaces only *optionally* in these cases.

8.1.3. Are we “digging in the right place”?

At this juncture, one might wonder whether Basque reflexives are truly what we should be focusing our investigation on, given that they have the very particular possessed-body-part structure outlined in section 8.1.1 (similar to Georgian reflexives, discussed in section 2.1).

Recall, however, the discussion in section 6. As Middleton (2018) has shown, the cross-linguistic landscape of syncretism in pronominal and anaphoric expressions can only be explained if reflexives universally conform to the structure in (45)—whether the language-specific morphology reveals it transparently (e.g. in the Peranakan Javanese of Semarang), or not. This has to be so, otherwise we lose the explanation for the universality of the no-discontinuous-syncretism result.



The structure in (45) implies that all reflexive anaphors must involve a pronominal (and hence φ -bearing) layer, which I have labeled PhiP, properly contained in a separate structural layer that is responsible for rendering the expression anaphoric, which I have labeled AnaphP (see section 6 for details). It just happens to be the case that in Basque (and in Georgian), these two layers are realized in a way that is morpho-syntactically very conspicuous. (The possessor constituting the φ -bearing layer, and the designated reflexive head-noun *buru* projecting the outer, anaphoric layer).

Thus, the structure of reflexives in Basque (and in Georgian) is merely a very transparent reflection of this universal containment structure. To put this another way: given Middleton’s results, an expression like the Albanian *vetja* in (46) necessarily includes a pronominal layer properly contained in an anaphoric layer, just like reflexives in Basque(/Georgian) do. It just happens to be the case that in Albanian, this containment structure is less clearly reflected in the morphology of the reflexive anaphor.

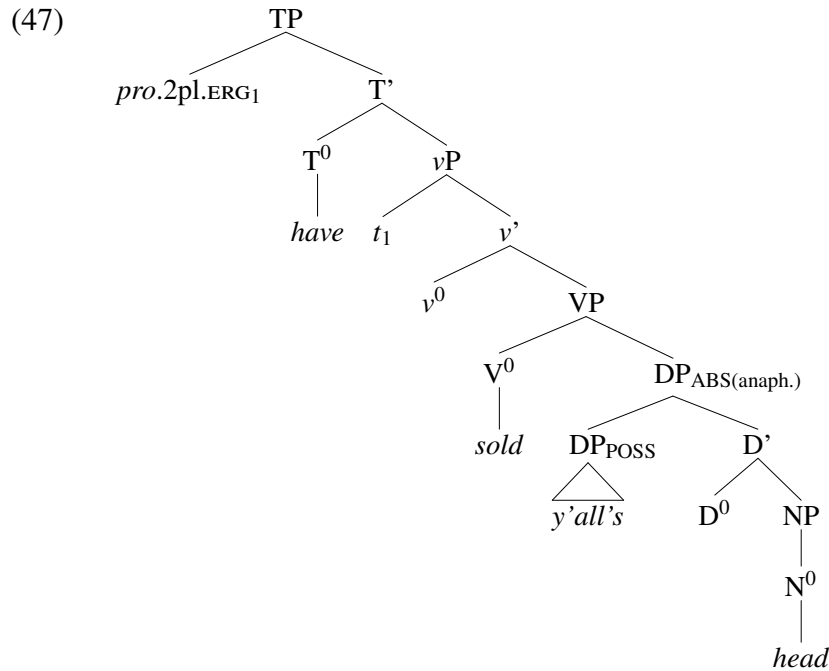
(46) Vetja më dhimset. (Albanian)
 self.NOM CL.1sg.DAT feel.sorry.for.3sg.PRES.NONACTIVE
 ‘I feel sorry for myself.’ [=(5)]

In light of this, reflexives of the Basque sort are arguably exactly what we should be looking at—given that they are undisguised realizations of what is going on (sometimes less transparently) in all reflexives cross-linguistically.

8.1.4. Anaphoric binding in Basque transitives

We are now in a position to examine a fairly straightforward scenario of anaphoric binding in Basque, from the perspective of timing-based accounts of the AAE. The scenario in question is binding of a transitive object by its clausemate subject.

As argued in detail by Arregi & Nevins (2012), the head responsible for φ -agreement with absolutive DPs in Basque is in the inflectional domain (T^0), and not in the verb-phrase domain (e.g. v^0 , as claimed for example in Preminger 2009). Let us now consider what this entails for timing-based accounts (even though Basque is head-final, diagrams have been drawn head-initially for ease of reading):



‘Y’ all have given yourselves away’
 (lit.: ‘Y’ all have sold y’ all’s head.)

[=(40)]

On a timing-based approach, the anaphor will receive its φ -feature values as soon as it is bound by the antecedent (in this case *pro.2pl.ERG*). By hypothesis, these φ -feature values will be visible on the entire anaphor, $DP_{ABS(anaph.)}$ (see section 8.1.1).

Importantly, even if there is indeed A-movement of the external argument to [Spec,TP], as indicated in (47), the A-trace in [Spec,vP] counts as a local binder for the reflexive. We have already established, independent of the data under discussion, that A-traces are perfectly capable binders—see the discussion in section 5 of data like (24a–c), repeated here:

- (48) a. The children_i seem to her_j/*_i to have t_1 amused Mary_i. [=(24a–c)]
 b. * The children_i seem to me to have t_1 amused myself.
 c. The children_i seem to me to have t_1 amused each other.

The net result is that, derivationally speaking, the reflexive anaphor will acquire its φ -feature values *before* it is probed by T^0 . And given the reductionist position, whereby φ -agreement is a necessary condition for binding, it follows that the values of those φ -features will be identical to those of the antecedent. T^0 is thus predicted to come upon φ -features matching those of the antecedent. This is precisely the wrong prediction, as Basque in fact obeys the AAE.

Thus, a timing-based account of the AAE will not work in Basque. Nor, for the same reasons, will it work in any other language in which T^0 agrees with the internal argument in a transitive

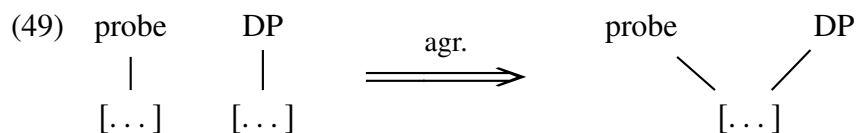
clause.¹⁹ Since it cannot generalize to such languages, a timing-based approach cannot serve as a general account of the AAE. And since the structure of reflexives in Basque is a fairly transparent reflection of the structure of reflexives cross-linguistically (see section 8.1.3), this result strikes at the root of the viability of timing-based accounts.

8.2. Further issues

The challenge identified in the previous subsection to timing-based approaches to the AAE stands on its own right; however, this is not all there is to be said about the relationship between timing-based approaches and what we know about the AAE. In this subsection, I will briefly address some further issues that arise in relation to these approaches.

The first matter concerns phases. One might wonder whether an appeal to phases (and/or cyclic spellout) might salvage the timing-based account of the AAE in, e.g., Basque. The answer, I contend, is “no.” The reason is as follows: if φ -agreement is a necessary condition for binding of β by α (as per the reductionist position), then there cannot be a relevant locality boundary separating the subject and the object in, e.g., (47). That is because it is manifestly the case that the subject is able to bind the object (and thus, by the reductionist hypothesis, agree with it). Moreover, in finite transitive clauses with *non*-anaphoric objects in Basque, T^0 is able to freely access the features of the object. It would be quite unexpected if clauses with reflexive objects contained *more* phasal boundaries than their non-reflexive counterparts.

The second matter concerns feature-sharing, and the additional problem (unrelated to the one identified in section 8.1) that it poses for timing-based accounts. As much work on agreement has shown, agreement does not consist in the “checking” of features (see Preminger 2011, 2014); nor does it “copy” feature-values from one place to another. Instead, it creates feature-sharing structures (see Andrews 1971, Frampton & Gutmann 2000, 2006, Gazdar et al. 1985, Pesetsky & Torrego 2007, Pollard & Sag 1994, Preminger 2017, *a.o.*):



Consequently, the result of φ -agreement between a direct-object anaphor and whichever φ -probe entered into agreement with it would be a feature-sharing structure like (49). Once binding did occur (and, by the reductionist hypothesis, the φ -features of the direct object were valued), the resulting valuation would affect the anaphor and the aforementioned φ -probe equally. That is because they are both linked to what is literally one and the same feature-structure. This predicts that *by the end of the derivation*, we would see full, nontrivial agreement with the anaphor—contrary to the AAE facts. And this would be so even if we were looking at a language where the relevant φ -probe were low (e.g. v^0) rather than high (e.g. T^0)—as in, for example, “ABS=DEF”/“LOW-ABS” languages

¹⁹A significant subset of ergative languages fit this description. These are the “HIGH-ABS” language in Coon, Mateo Pedro & Preminger’s (2014) parlance, or the “ABS=NOM” ones in Legate’s (2008)—languages where T^0 is the functional head that enters into a relationship with the absolutive argument. Interestingly, Basque itself does not fit Legate’s (2008) case-theoretic diagnostics (which Coon, Mateo Pedro & Preminger 2014 adopt) for being an “ABS=NOM”/“HIGH-ABS” language. Nevertheless, it is quite clear that the locus of absolutive agreement in Basque is T^0 (see Arregi & Nevins 2012 for extensive evidence of this).

(Coon, Mateo Pedro & Preminger 2014, Legate 2008). Because of the nature of feature-sharing, the relative order of φ -probing and binding-cum- φ -agreement would not matter.

Therefore, the evidence that agreement gives rise to a feature-sharing structure (see references above) constitutes an additional reason, unrelated to the one discussed in section 8.1, why timing-based approaches to the AAE are unsuccessful.

9. Encapsulation and reductionism

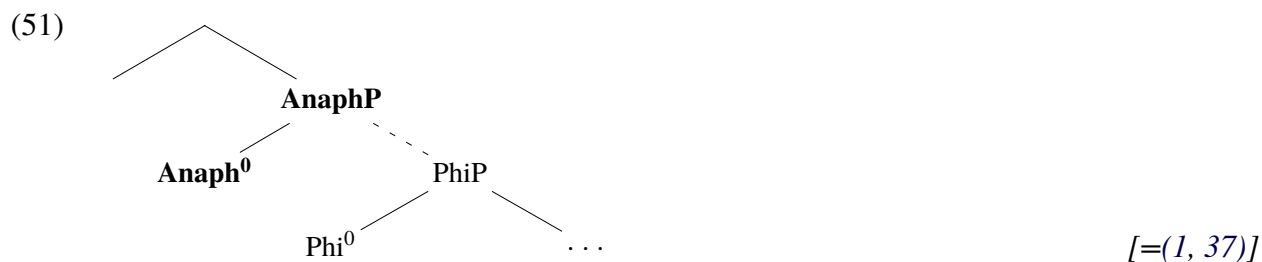
In this section, I will consider the fate of the reductionist position—repeated again in (50)—in light of the evidence for φ -encapsulation adduced in section 6.

(50) THE REDUCTIONIST POSITION [=(14, 38)]
 α and β can share a binding index **only if** α and β have entered into syntactic agreement in φ -features

- where, for the purposes of this definition, entering into φ -agreement is subject to transitive closure
 - i.e., if α agrees with γ , and γ agrees with β , then α and β count as having entered into agreement with one another for the purposes of this definition

9.1. A false prediction: English

Taking the reductionist position seriously entails that in a structure like (51), it is necessarily PhiP—and crucially, not AnaphP in its entirety—that enters into a binding relation with the antecedent. That is because it is PhiP, and not AnaphP, that bears valued φ -features corresponding to those of the antecedent (section 6).



This yields a concrete prediction, given in (52):

(52) REDUCTIONIST PREDICTION
 The outermost layer of a reflexive anaphor should behave as if it is *not* the bearer of the relevant binding index.²⁰

In light of this prediction, consider (53a–b):

²⁰It is logically possible to assume that the binding index percolates from PhiP to AnaphP. But note that in order to capture the AAE, it is necessary to assume that the valued φ -features on PhiP do not similarly percolate (see section 8). If we assume that φ -features and binding indices diverge in this manner, we have in fact already abandoned the reductionist position (which holds that the two travel in concert; cf. (50)), and nothing more would need to be said in this section.

(53) a. John_i expects Mary to outdo him_{i/k}.

b. John_i expects himself_i to outdo him_{k/*i}.

[Norvin Richards, p.c.]

Notice first that (53a) does not give rise to a disjoint-reference effect between *John* and the pronoun *him*. This means that *John* is too far a way from the pronoun, structurally speaking, for the two to enter into a local binding relation. Crucially, this means that the cause of the disjoint-reference effect observed in (53b) cannot be *John*, and must instead be the anaphor, *himself*. But this could only be the case of the binding index resided on the outermost projection on the anaphor, contrary to the reductionist prediction in (52).

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

(54) The children seem [_{PP} *to* [_{DP} *her*_{j/*i}]] to have amused Mary_i.

[=(24a)]

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 (54) demonstrates). However, PPs in English do not typically behave as bearers of φ -features; it stands to reason that the bearer of φ -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* (Gazdar et al. 1985, Webelhuth 1992, a.o.; but see Cable 2007, 2010, Heck 2004 for criticism). On this view, the φ -features of *her* in (54) percolate to the PP layer, together with the binding index, explaining the binding-theoretic behavior observed in (54).

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 (40a–b), above. If φ -feature percolation (from the unmistakably φ -feature-bearing possessor, to the entire anaphoric expression) were what underpinned these cases, we would see these percolated φ -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 φ -features have percolated in the manner one might hypothesize for (54), the verdict is that they have not. (See Amiridze 2003 for converging evidence from Georgian.) This renders the percolation explanation of (54) 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 φ -features. In other words: to reject reductionism.

9.2. Potential cross-linguistic variation in the locus of binding indices

Iatridou (1988) discusses reflexive anaphors in Greek, which have a possessed-noun structure (as seen earlier for Georgian and Basque):²¹

(55) Costas thavmazi [ton eafton tu]
Costas admires DET.ACC.Msg self 3sg.GEN.Msg
'Costas admires himself.'

[Iatridou 1988:699]

²¹I thank a reviewer for bringing this work to my attention.

that align with the discrete overt morphemes in *ton eaf ton tu*. One important lesson of Middleton’s (2018) findings is that this is simply wrong: anaphoric expressions contain at least two structural layers (and per Middleton, at least three or four, depending on whether the expression is a diaphor or an anaphor proper), regardless of which of these are morphologized, and if so, how they are morphologized. All that examples like (58) show us, then, is that the binding on *ton eaf ton tu* is not on its very outermost structural layer (on the assumption that the preverbal clitic *ton* would then behave as if it bore the same index). That hardly suffices to show that the index therefore resides exactly on that XP which is the locus of valued φ -features (e.g. the possessor *tu*).

More important, however, is that none of this bears at all on the argument against reductionism presented in section 9.1. Given universal φ -encapsulation (section 6), the reductionist position entails that a language like English—where anaphors behave as if their binding index were on the outermost layer of the anaphoric expression—could not exist. But abandoning the reductionist position does not entail that in *no* language would it (appear to) be the case that the binding index does cohabit with valued φ -features on the same XP.

9.3. Interim summary

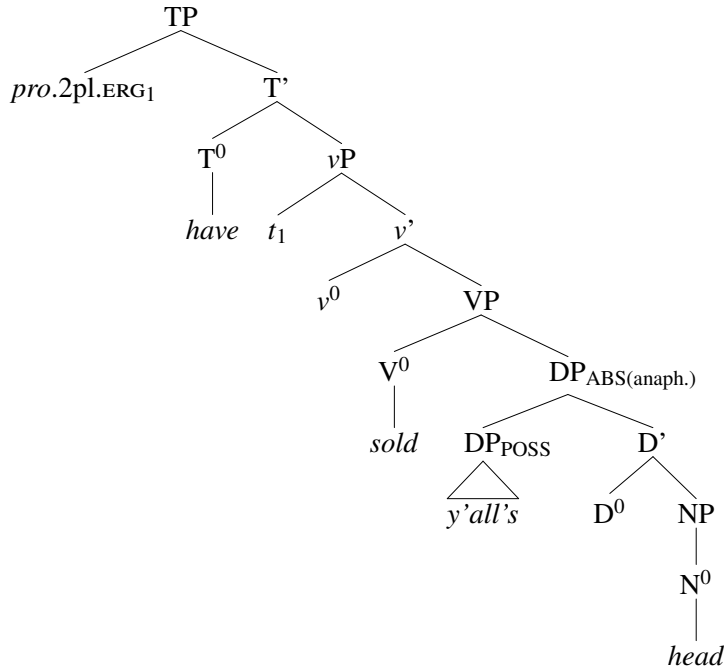
Reductionism, when juxtaposed with the evidence for φ -encapsulation, yields false predictions about the structural locus of binding indices (e.g. in English).

Importantly, this result has a different status than, e.g., the result in section 8 showing that timing-based approaches cannot reconcile reductionism with the AAE. To that result, one could respond—correctly—that it does not prove the requisite negative, *viz.*, that there is *no* way of reconciling reductionism with the AAE; it merely eliminates one contender for doing so (timing-based accounts). But the results from this section are more broad, showing that if φ -encapsulation holds (and there is every reason to believe it does), then reductionism just will not work.

10. An encapsulation-based approach to the AAE

Given that timing-based approaches cannot account for the AAE (section 8), and that reductionism is quite generally unsuccessful in light of φ -encapsulation (section 9), what we are still in need of is an explanation of the AAE. Suppose, then, that the AAE arises because φ -agreement comes upon a structural layer (AnaphP in (51), instantiated by $DP_{ABS(anaph.)}$ in the Basque (59), for example) that prevents access to the feature-values hosted on the φ -bearing portion of the reflexive (PhiP in (51), instantiated by DP_{POSS} in (59)).

(59)



‘Y’all have given yourselves away’
 (lit.: ‘Y’all have sold y’all’s head.’)

[=(47)]

That a structural layer of this sort exists need not be stipulated; it follows from Middleton’s (2018) results. All that is being added here is the assumption that this universally-present layer (AnaphP) has properties that cause it to halt φ -probing. This could be because it is phasal; it could be because it bears its own set of valued φ -features, distinct from the ones that covary with the antecedent; or both things could be true. In any case, the assumption that this layer halts φ -probing clearly suffices to derive the AAE: a φ -probe located properly outside the AnaphP layer (or $DP_{ABS(anaph.)}$ in (59)) will run into this opaque layer, eliminating the possibility of nontrivial φ -agreement between the φ -probe and the anaphor.

(60) ANAPHOR AGREEMENT EFFECT

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

[=(8, 39)]

Finally, as discussed in detail in section 4, φ -matching between the PhiP layer (DP_{POSS} in (59)) and the antecedent in no way implicates the existence of syntactic agreement between the two. It can just as well arise through what I have labeled *NSM*: whatever non-syntactic mechanism is responsible for φ -matching in those cases of binding where syntax cannot possibly be involved (e.g. Donkey Anaphora).

11. Remaining issues

11.1. Non-AAE-obeying languages

The account presented in the previous sections derives the AAE as a universal property of all anaphors in all languages. Its universality on this account derives from the universality of φ -encapsulation (section 6). Murugesan (2018, 2019), however, shows that some languages genuinely violate the AAE, exhibiting nontrivial agreement with anaphors.

11.1.1. Tamil: the basics

Tamil (Dravidian) is a nominative-accusative language where only nominative DPs control verbal agreement:

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

Tamil also has verbs whose subject is dative. The object of such verbs surfaces as nominative. Unsurprisingly, in light of these case facts, it is the object that controls agreement in this case:

- (62) Kohli-ukku Meena kidai-t-aal
Kohli-DAT Meena(NOM) see-PAST-3sgF
'Kohli got Meena.' [Murugesan 2018:ex. (14)]

In clauses where there is *no* accessible nominative DP, the verb surfaces with invariant 3rd person singular neuter agreement morphology:

- (63) 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:exx. (16–17)]

Crucially, there appears to be nontrivial agreement with the anaphor *taan*:²³

- (64) 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:exx. (15a–b)]

In light of (61–63), we can be quite sure that the source of non-3sgN agreement in (64a–b) is *not* the dative antecedent. These facts therefore stand as a genuine exception to the AAE.

²³Jeffrey Lidz (p.c.) informs me that a similar pattern obtains in Kannada (Dravidian).

11.1.2. Murugesan’s (2018, 2019) analysis

Murugesan (2018, 2019) puts forth a timing-based, reductionist analysis of these Tamil facts. Notwithstanding the shortcomings, discussed in previous sections, of timing-based accounts, and of reductionism in general—let us hold these factors in abeyance and outline the proposal itself. Appealing to the relative structural height of the φ -probe and the antecedent, Murugesan proposes that the AAE arises when the φ -probe is located below the antecedent, and therefore enters the structure earlier, at a point when the anaphor has not yet been bound. If one were to assume reductionism, this absence of binding would entail an absence of φ -feature values, yielding the AAE. In contrast, in those languages where the φ -probe is located above the antecedent, the anaphor will have already been bound by the time φ -probing occurs, and thus the AAE would not arise. Tamil, on Murugesan’s account, is a language of the latter type, as he assumes that the φ -probe is located on T^0 and the subject is base-generated lower (e.g. in [Spec, ν P]).

But as noted, we have already seen why such an analysis will not generalize beyond Tamil. Timing-based accounts cannot be maintained as general accounts of the AAE (section 8). Nor can reductionism—and this is so whether derivational timing plays a role in the account or not (section 9). Nevertheless, much of the empirical burden in arguing for these conclusions was borne by languages like Georgian, Basque, 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 on the basis of such data, and reinstate a timing-based account of the kind Murugesan envisions?

The answer, I think, is “no.” 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 Georgian and Basque would stand as a twofold coincidence. First, we would have to say that the reason these languages exhibit the AAE is unrelated to (and, in fact, disjoint from) the reason why other languages exhibit the AAE. In Basque, the absolutive φ -probe is located on T^0 (Arregi & Nevins 2012), above both the anaphor and the base position of the antecedent. Basque is therefore precisely the kind of language predicted *not* to exhibit the AAE, on Murugesan’s account. The reason Basque (as well as Georgian) exhibits the AAE would have to have something to do with the possessed-body-part structure of its reflexives—which, on this view, would be entirely unrelated to the reason other languages exhibit the AAE (the relative structural height of the antecedent and the φ -probe).

Second, this purportedly special structure of reflexives, which would have to be taken as the cause of the AAE in languages like Basque and Georgian, is in fact what Middleton (2018) has shown holds *universally*. Anaphors involve additional structural layers encapsulating the φ -feature-bearing layer of structure, whether this happens to be transparently detectable in a particular language or not. If it were possible to deviate from this on a language-specific basis, her results, concerning the universal ban on discontinuous syncretism in anaphoric expressions, would be left unexplained (see section 6 for details). Far from being an exception, then, languages like Basque and Georgian 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.

11.1.3. More Tamil facts

Before sketching an alternative, non-reductionist, non-timing-based analysis of Tamil (and languages like it), let me mention two additional facts about agreement with anaphors in Tamil.

First, there are cases that appear to instantiate agreement with a (nominative) anaphor in subject position (as opposed to the nominative object anaphors we saw in section 11.1.2).

- (65) a. Mani_i [taan_i sathat-ai sapi-t-**aan**-nnu] son-n-aan
 Mani_M self rice-ACC eat-PAST-**3sgM**-COMP say-PAST-3sgM
 ‘Mani_i said that self_i ate the rice.’
 b. Banu_i [taan_i sathat-ai sapi-t-**aal**-nnu] son-n-aal
 Banu_F self rice-ACC eat-PAST-**3sgF**-COMP say-PAST-3sgF
 ‘Banu_i said that self_i ate the rice.’ [Murugesan 2018:exx. (46a–b)]

But as Murugesan points out, this is almost certainly a case of agreement with a syntactically realized (but phonologically-null) perspective-holder (Sundaesan 2016). As such, it need not target one of the core arguments at all, as (66) demonstrates:

- (66) Banu_i [taan_{i/*j} saatat-ai sapi-t-**een**-nnu] so-n-aal
 Banu_F self rice-ACC eat-PAST-**1sg**-COMP] say-PAST-3sgF
 ‘Banu_i said that self_{i/*j} ate the rice.’ [Murugesan 2018:ex. (48)]

Second, *taan* itself simply cannot be bound by grammatically 1st/2nd person antecedents. In such cases, Tamil employs forms that are morpho-phonologically indistinguishable from 1st/2nd person pronominals. (A similar phenomenon is found in Romance.)

- (67) En-akku_i naan_i tirumba kidai-tt-**een**
 1sg-DAT 1sg.NOM(PRON) again get-PAST-1sg
 ‘I_i got myself_i back again.’ [Sandhya Sundaesan, p.c.]

This means that, while agreement with anaphors in Tamil does qualify as *nontrivial* (given, e.g., the bona fide covariance in GENDER features seen in (64a–b)), it would also be imprecise to characterize it as full-fledged agreement. The only true, unambiguous anaphor (*taan*) is restricted to 3rd person. One could therefore imagine a further attenuation of how we define the AAE, such that nontrivial agreement *in PERSON features* is the crucial element (see Abramovitz 2019 for an approach along these lines). If we were to pursue such an approach, a language like Tamil would no longer stand as a counterexample to the AAE. However, NUMBER agreement with anaphors is not generally permitted (see, e.g., Amiridze 2003 on Georgian), it seems to me that adopting a PERSON-only AAE misses an important part of the picture.

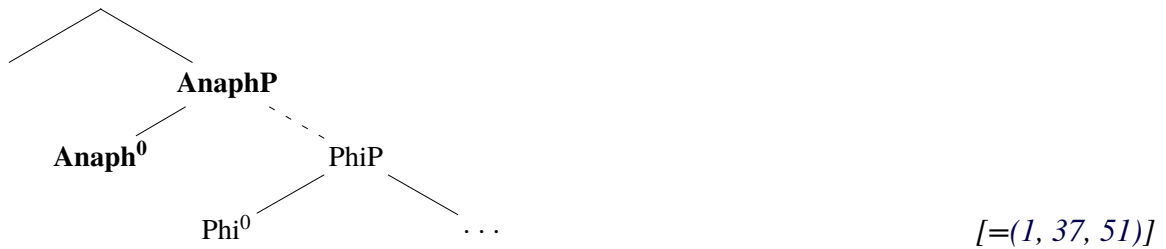
11.1.4. Weighing the options

Consider now what each competing account (Murugesan’s, and the encapsulation-based account proposed here) would have to do to accommodate the data that remains recalcitrant for that account. We have already discussed the pitfalls of excluding languages like Basque (languages with complex reflexives in which the φ -probe is nevertheless situated higher than the typical antecedent) from the purview of an AAE account (section 11.1.2). There does not seem to be a good way for a timing-based account to deal with the variation between, e.g., Tamil on the one hand, and Basque on the other.

On the φ -encapsulation account pursued here, languages like Tamil could simply be a case where AnaphP is exceptionally *not* syntactically opaque—either because the normally-phasal AnaphP is

exceptionally non-phasal in these languages; or because the normally φ -bearing AnaphP instead behaves like a pseudo-partitive (cf. the earlier discussion of Basque; see section 8.1):

(68)



The former, phase-based approach has parallels in Abels (2003) analysis of preposition stranding, wherein PP is a phase in most languages, but can be deemed non-phasal by the learner in light of positive evidence (*viz.* P-stranding).²⁴ This case would be similar: AnaphP would be phasal in the vast majority of languages, but could be deemed non-phasal by the learner, in light of positive evidence (*viz.* nontrivial agreement with anaphors).²⁵ It is also suggestive, in this regard, that AAE-violating languages like Tamil are roughly as common as preposition-stranding languages like English—very rare, that is (modulo the usual caveats on the pitfalls of counting languages).

11.2. The fate of attempted agreement with anaphors

One issue which was touched on in section 2.1, but which I have not returned to so far, concerns what happens when a derivation occurs in which agreement with an anaphor is *attempted*. We know the result will not be successful (nontrivial) agreement, given (69):

(69) ANAPHOR AGREEMENT EFFECT

$$* \text{H}^0 \dots \text{DP}_{\text{ANAPH}}, \text{ where } \mathcal{R} \text{ is a nontrivial } \varphi\text{-agreement relation} \quad [=(8, 39, 60)]$$

$\underbrace{\hspace{1.5cm}}_{\mathcal{R}}$

But this still leaves us with two possible outcomes:

- (70) **i.** a grammatical utterance with a default / nonvarying agreement form
- ii.** ungrammaticality

Rizzi (1990) was operating on the assumption that (70.ii) was the only possible outcome; but we have seen that this is false. Instead, we find variation. In Icelandic, the result is outright ungrammaticality:

(71) Sigga_i telur [að mér líki hún_i/*sig_i]
 Sigga thinks that me.DAT likes.SBJV.3sg she.NOM/*SIG
 ‘Sigga_i thinks that I like her_i.’ [=(11)]

In Albanian, the result is a (grammatical) default form:

²⁴Abels (2012) later pursues an alternative, wherein PP is always phasal, and the difference between P-stranding languages and non-P-stranding ones consists in the presence of a (possibly unpronounced) morpheme between P and its supposed complement. Be that as it may, it does not bear directly on whether this is the right analysis of the linguistic variation under discussion here.

²⁵See also Preminger (2019), for an independent argument that children are sensitive to overt morpho-phonological covariance (i.e., nontrivial agreement) in setting the parameters of their language.

(72) *Vetja mē dhimset.*
 self.NOM CL.1sg.DAT feel.sorry.for.3sg.PRES.NONACTIVE
 ‘I feel sorry for myself.’

[=(5, 46)]

Furthermore, as pointed out by Murugesan (2018), the Icelandic pattern, in which *sig* is unable to occur as a nominative object, is surprising in and of itself. That is because, as shown in (73), actual agreement with nominative objects in Icelandic exhibits optionality (see also Hornstein 2018).

(73) *Henni leiddist/leiddust þeir*
 she.DAT was.bored.by.3sg/3pl they.NOM
 ‘She was bored with them.’

[Taraldsen 1995:307]

Importantly, the surface structure of an anaphor—whether it appears morphologically simplex or complex, whether it alternates overtly for φ -features or not—is not predictive of the behavior of the anaphor with respect to the AAE. When it comes to simplex anaphors: some can occur in the relevant positions with invariant 3sg agreement (Albanian);²⁶ and some cannot occur in the relevant positions at all (Icelandic). The state of affairs with respect to complex anaphors is similar: some can occur in the relevant positions with invariant 3sg agreement (Basque); and some cannot occur in the relevant positions at all (Italian).

Likewise, when it comes to φ -varying anaphors: some can occur in the relevant positions with invariant 3sg agreement (Basque); and some cannot occur in the relevant positions at all (Italian). The state of affairs with respect to φ -invariant anaphors is again similar: some can occur in the relevant positions with invariant 3sg agreement (Inuktitut);²⁷ and some cannot occur in the relevant positions at all (Icelandic).

Nor are other properties of the languages in question predictive of these behaviors. Basque is an ergative-absolutive language, its reflexive is complex, and it occurs with invariant 3sg agreement. Albanian is a nominative-accusative language, its reflexive is simplex, and yet it behaves exactly like Basque. Icelandic is like Albanian in being nominative-accusative and having a simplex reflexive, but behaves the opposite way with respect to the fate of this reflexive when it occurs in positions that would otherwise control agreement.

All of this is noteworthy because it shows that rejecting the universality of φ -encapsulation (i.e., breaking the class of anaphors into separate subclasses, each with a different syntactic structure) is not only problematic from the perspective of the facts surveyed in section 6. It also provides no apparent help in understanding the phenomena at hand.

Finally, while the encapsulation hypothesis provides no particular insight into whether a given language/derivation will yield (70.i) (a grammatical sentence with invariant φ -agreement) or (70.ii) (ungrammaticality), neither do approaches based on reductionism and/or derivational timing. Consider: if one assumed that agreement with an anaphor that has not yet been bound, or whose φ -features have not yet been valued, gives rise to ungrammaticality, then Albanian, Basque, and languages like them would remain unexplained. If, on the other hand, one assumed that such agreement gives rise to default, 3sg agreement, then Icelandic, English, and languages like them would remain unexplained.

²⁶On the simplex nature of Albanian *vetja*, see Franks (2013).

²⁷On Inuktitut reflexive anaphors, see Yuan (2018).

11.4. Reciprocals

A reviewer points out that in contrast to the state of affairs with reflexives, there seems to be a much weaker ban (if any) on agreement with reciprocals (see Tucker 2011:11–12 for a discussion of some of the variation found with respect to such examples):

(76) % Linguists and philosophers seldom know what each other are talking about.

Given that both reflexives and reciprocals involve binding, it is not at all clear how reductionist approaches could possibly deal with this contrast between reflexives and reciprocals vis-à-vis φ -agreement.

On the φ -encapsulation proposal, things may not be so bleak. As the reviewer notes, Middleton's (2018) results surveyed in section 6 concern reflexives and not reciprocals, and it is likely that they do not extend to reciprocals. If this is the case, then the encapsulation-based account of the AAE (section 10) would not extend to reciprocals, either, suggesting a way that the present proposal might have a handle on this issue.

12. Conclusion

This paper has been concerned with the Anaphor Agreement Effect (AAE). I began by arguing that the AAE is a ban on (nontrivial) agreement with an anaphor—and not, as Rizzi (1990) claimed, a restriction on the positions in which an anaphor can occur. I then proceeded to establish two central points about how the theories of φ -agreement and binding relate to one another. First, I showed that overt φ -feature matching between a binder and a bindee is in no way an argument for syntactic φ -agreement between the two (contra, e.g., Kratzer 2009, Reuland 2011, Rooryck & Vanden Wyngaerd 2011). Second, I reviewed Middleton's (2018) argument for universal *structural encapsulation* in the structure of anaphors, and showed how the (more coarse-grained) *φ -encapsulation hypothesis* can be derived as a particular consequence of her results.

I then turned to timing-based approaches to the AAE, showing that the finer details of anaphoric binding (especially in languages where T^0 agrees with the transitive object) render such accounts untenable. Next, I showed that φ -encapsulation alone (i.e., even if one abstracts away from derivational timing) militates against reductionist theory of binding—that is, against any theory in which (possibly mediated) syntactic φ -agreement between α and β is a necessary condition for a binding relation between the two.

Finally, I argued that φ -encapsulation (sans reductionism) is sufficient to account for the AAE as construed above.

One facet of the picture that remained unexplained was the fate of attempted AAE violations (default morphology vs. ungrammaticality); but it was shown that none of the current theories of the AAE offer any purchase on this issue.

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.

- Abramovitz, Rafael. 2019. *What can agreeing anaphors tell us about the Anaphor Agreement Effect?*, Paper presented at the 42nd Generative Linguistics in the Old World (GLOW) Colloquium, Oslo.
- 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 Zagana, 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.
- Baker, Mark C. 2008. *The syntax of agreement and concord*. Cambridge Studies in Linguistics 115, Cambridge: Cambridge University Press.
- Béjar, Susana & Milan Rezac. 2009. Cyclic Agree. *Linguistic Inquiry* 40:35–73, doi: <10.1162/ling.2009.40.1.35>.
- Belletti, Adriana & Luigi Rizzi. 1988. Psych-verbs and Theta-Theory. *Natural Language & Linguistic Theory* 6:291–352, doi: <10.1007/BF00133902>.
- Bhatt, Rajesh & Stefan Keine. 2017. Long-distance agreement. In *The Wiley companion to syntax*, eds. Martin Everaert & Henk van Riemsdijk, 2nd Edition, Oxford: Wiley.
- Bjorkman, Bronwyn & Hedde Zeijlstra. 2019. Checking up on (φ -)agree. *Linguistic Inquiry* 50:527–569, 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](https://doi.org/10.5334/gjgl.348)>.
- 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.
- Collins, Chris & Paul M. Postal. 2012. *Imposters: a study of pronominal agreement*. Cambridge, MA: MIT Press.
- 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](https://doi.org/10.1075/lv.14.2.01coo)>.
- 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.
- Embick, David. 2010. *Localism versus globalism in morphology and phonology*. Linguistic Inquiry Monographs 60, Cambridge, MA: MIT 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 & Sam Gutmann. 2000. *Agreement is feature sharing*. Ms., Boston, MA: Northeastern University. URL: <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.
- Harris, Alice C. 1981. *Georgian syntax: a study in Relational Grammar*. Philadelphia, PA: 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](https://doi.org/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.
- Iatridou, Sabine. 1988. Clitics, anaphors, and a problem of coindexation. *Linguistic Inquiry* 19:698–703.
- 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>.
- 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. 2019. *Predicting the AAE and its violations*. Doctoral dissertation, Leipzig: Institut für Linguistik, Universität Leipzig.
- Nevins, Andrew Ira. 2007. The representation of third person and its consequences for Person-Case effects. *Natural Language & Linguistic Theory* 25:273–313, DOI: <10.1007/s11049-006-9017-2>.
- Norris, Mark. 2012. Towards an analysis of concord (in Icelandic). In *Proceedings of the 29th West Coast Conference on Formal Linguistics (WCCFL 29)*, eds. Jaehoon Choi, Alan Houge, Jessamyn Schertz, Jeff Punske, Deniz Tat & Alex Trueman, Somerville, MA: Cascadilla Press, 205–213.
- Norris, Mark. 2014. *A theory of nominal concord*. Doctoral dissertation, Santa Cruz, CA: University of California.
- 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. Emonds*, eds. Simin Karimi, Vida Samiian & Wendy Wilkins, 262–294. Amsterdam: John Benjamins.
- Podobryaev, Alexander. 2014. *Persons, imposters, and monsters*. Doctoral dissertation, Cambridge, MA: MIT.
- 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. to appear. *The Agreement Theta Generalization*. *Glossa*, 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. 2009. Breaking agreements: distinguishing agreement and clitic doubling by their failures. *Linguistic Inquiry* 40:619–666, DOI: <10.1162/ling.2009.40.4.619>.
- Preminger, Omer. 2011. *Agreement as a fallible operation*. Doctoral dissertation, Cambridge, MA: MIT.
- 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* 4(1):13, 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>.
- Rizzi, Luigi. 1990. On the anaphor-agreement effect. *Rivista di Linguistica* 2:27–42.
- 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>.
- 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. 2018. Perspective is syntactic: evidence from anaphora. *Glossa* 3:1–40, doi: <10.5334/gjgl.81>.
- 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.
- Tucker, Matthew A. 2011. *On the derivation of the Anaphor Agreement Effect*. Ms., Santa Cruz, CA: University of California. URL: <<https://matthew-tucker.github.io/files/papers/anaphor-agreement-effect.pdf>>.
- 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>.
- Wurmbrand, Susi. 2017. Feature sharing or how I value my son. In *A Pesky Set: papers for David Pesetsky*, eds. Claire Halpert, Hadas Kotek & Coppe van Urk, 173–182. Cambridge, MA: MITWPL.
- Yuan, Michelle. 2018. *Dimensions of ergativity in Inuit: theory and microvariation*. Doctoral dissertation, Cambridge, MA: MIT.

Zaenen, Annie, Joan Maling & Höskuldur Thráinsson. 1985. Case and grammatical functions: the Icelandic passive. *Natural Language & Linguistic Theory* 3:441–483.

svn revision code: 10652