The Anaphor Agreement Effect: further evidence against binding-as-agreement
Omer Preminger*
University of Maryland

Handout also available at
https://omer.lingsite.org/talk-in-oslo/
—or just scan this:

1. The Anaphor Agreement Effect (AAE)

1.1. Overview

(1) **Anaphor Agreement Effect** (Rizzi 1990:26)
Anaphors do not occur in syntactic positions construed with agreement.

(2) * 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]

1. Problem:

(3) Vëtja më dhimset. (Albanian)
self.NOM CL.1sg.DAT feel.sorry.for.3sg.pres.nactive

• Notice:
  - ex. (3) involves a DAT-Experiencer verb, much like ex. (2);
  - like in Italian, this means the NOM Theme controls agreement;
  - but unlike in Italian, a verb form with 3sg morphology renders the occurrence of a NOM anaphor well-formed; cf.:

(4) * A voi interessai solo voi stessi.
to you.DAT.pl interest.3sg only yourselves(NOM)
‘You are interested only in yourselves.’

• Thus:

(5) **Anaphor Agreement Effect** (Tucker 2011:8)
Anaphors do not occur in syntactic positions construed with agreement, **unless the agreement does not vary for φ-features**.

  - ex. (3) is exempted from the effects of (1) —
    - because ‘default agreement’ does not vary for φ-features (cf. (5))

• Notice:

  - there is now no meaningful sense in which agreement restricts the possible positions of an anaphor
  - sometimes (e.g. in Italian, or English) they cannot occur in positions that would otherwise trigger agreement (as per Rizzi 1990);
  - and sometimes (e.g. in Nez Perce, or Albanian) they can (as per Tucker 2011, Woolford 1999)

  ▶ The only constant, it seems, is the following:

(6) **Anaphor Agreement Relation**

* H₀ . . . DPANAPH, where R is a nontrivial φ-agreement relation

(7) A φ-agreement relation between α and β is **nontrivial** if there are at least two sets of feature-values, F and F’, such that α takes on one form, α(F), when β bears F; and takes on another form, α(F’), when β bears F’. (And α(F) and α(F’) are both part of the normal φ-agreement paradigm of α.)

---

*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.

1 Woolford (1999) and Tucker (2011) discuss one more pattern related to the AAE, that of Nez Perce. In Nez Perce, verbs with reflexive objects take object-agreement markers from a dedicated anaphoric agreement paradigm, which is distinct from the paradigm of normal object agreement (see also Deal 2010). Nevertheless, the form of these markers still alternates based on the φ-features of the anaphor (or of its binder; there would be no way to tell). I follow Tucker (2011:7–8) in setting this example aside for the purpose of the Anaphoric Agreement Effect. Tucker himself suggests that this agreement morphology may ultimately be agreement with the antecedent, not with the anaphor, and that it therefore spells out a portmanteau of subject agreement with detransitivizing reflexive morphology. If such an analysis were to prove untenable, it would instead be possible to fold in Nez Perce with other genuine AAE-violating languages like Tamil; see §9.1 for further discussion.
for the purposes of this talk, I will treat (6) as the real AAE.

• In this talk, I will argue:
  (i) the AAE, sometimes taken to support a reduction of binding to \( \varphi \)-agreement, actually provides evidence against such a reduction
  (ii) the AAE arises due to \( \varphi \)-encapsulation —
    o the \( \varphi \)-features borne by the anaphor are properly contained in a syntactically opaque structure

\[
\text{THE } \varphi\text{-ENCAPSULATION HYPOTHESIS}
\]

  \( \text{AnaphP} \)

  Anaph^0

  \( \text{PhiP} \)

  \( \text{Phi^0} \)

  o in other words, the structure that is transparently realized in a language like Georgian —

\[
\begin{align*}
(9) & \quad [ \text{t'ems} \quad \text{tav-s} ] \quad \text{vakeb} \\
& \quad 1\text{sg.poss \ head-DAT} \quad \text{praise}(1\text{sgS}, 3\text{sgO}) \\
& \quad \text{‘I praise myself.’} \quad \text{[Harris 1981:27]}
\end{align*}
\]

↑ is at play even in languages where it is less transparently evident.

1.2. Agreement, not case

• Icelandic has a long-distance anaphor, \( \text{sig/sér/sín} \) (\text{ACC/DAT/GEN})\(^2\)
  o I will gloss this element as “sig” henceforth

• In (10), the embedded subjunctive verb agrees with the \text{nom} subject (\( \text{hi} “\text{you}” \)), not with \text{sig}

\[
(10) \quad \text{Jóni heldur [ að hi hatir sig] \quad (Icelandic)}
\]

  John believes that you hate.\text{sbjv.2sg sig.acc}

  ‘John believes that you hate him.’ \quad \text{[Thráinsson 2007:467]}

  \( \Rightarrow \) accordingly, there is no violation in (10)

  ➢ Contrast this with (11):
    o here, the embedded verb is a quirky-subject verb
      – assigning DAT case to its subject (\text{mér “me.dat”})
      – and NOM case to its object
    o finite agreement tracks case, not grammatical function
    \( \Rightarrow \) in such a clause, it is the \text{nom} object, not the DAT subject, that would control agreement

\[
(11) \quad \text{Sigg\(\text{a} \text{telur [ að mér } \text{líki húni/sig]}} \quad \text{Sigga thinks that me.dat likes sbjv.3sg she.nom/sig}
\]

  ‘Sigga thinks that I like her.’ \quad \text{[Maling 1984:217]}

  \( \Rightarrow \) hence sig cannot occur in the object position of the embedded clause in (11).

\text{NB: This is not a ban on sig serving as an argument of a quirky-case verb.}

\[
(12) \quad \text{Húni ságði [ að sig vantaði peninga]. she.nom said that sig.acc lacked.sbjv.3sg money}
\]

  ‘She said that she lacked money.’ \quad \text{[Maling 1984:216]}

  ➢ Because finite agreement in Icelandic tracks \text{nom} case, the data in (10–12) could be characterized in terms of case (e.g. “no \text{nom} anaphors”) —
    o rather than in terms of \( \varphi \)-agreement

  ➢ However, such a characterization fails to generalize (Woolford 1999:262ff.):

\[^2\text{It has been argued, by Thráinsson (1979), Maling (1984), and others, that sig is in fact logophoric in nature (cf. Clements 1975 and subsequent literature), and as such, its distribution is subject to discourse-oriented restrictions. While this is certainly an interesting facet of sig’s distribution, it is ultimately at right angles to our present concerns. That is because it turns out that there is no discourse configuration, however unorthodox, that allows sig to occur in a position violating (6). The latter is clearly not a discourse-related fact, and it is this facet of sig’s distribution that is of interest here. This does illustrate, of course, that (6) should be taken as a necessary-but-not-sufficient condition on the distribution of anaphors—exactly as one would expect. See Maling (1984), Sells (1987), Thráinsson (1991, 1992), among others, for further discussion.}\]
The AAE: against binding-as-agreement

Omer Preminger

University of Maryland

University of Oslo
October 2019

The AAE: against binding-as-agreement

Omer Preminger

University of Maryland

(13) sensei-ni(-wa) zibun-ga wakar-ani-i

teacher-DAT(-TOPIC) self-NOM understand-NEG-PRES

‘The teacher does not understand himself’

Shibatani 1977:800, via Woolford 1999:263

1.3. Syntax, not morphology

The AAE —

• just based on the characterization in §1.1–§1.2, without further analysis

— already has implications for the way we think about the syntax & morphology of ϕ-agreement cross-linguistically

Consider (13) once more, as well as (11), repeated here:

(11) Sigga telur think.DAT that .

Sigga thinks that .

(hún) sig she.

Sigga thinks that I like her.

Maling 1984:217

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)

(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 nom anaphor in (11) from its counterpart in (13)?

• it cannot be the presence of overt agreement with the anaphor in (11), since syntax is modularly encapsulated from morpho-phonology

— and thus the AAE, situated in syntax by hypothesis, could not ascertain whether a particular a morphological term is or isn’t overt;

• and since Japanese is allowed to have a morpho-phonologically null (but syntactically extant) version of the agreement relation in (11) —

— there is no explanation for why the anaphor in (13) is licit, while the one in (11) is not.

⇒ We’ve learned that [(i) ∨ (ii)] must be true:

• Now consider:

Is the idea that (ii) is true—i.e., that the AAE is morphological in nature—actually a tenable position?

⇒ It is not. Here’s why:

• what the AAE prohibits is (nontrivial) agreement between a probe H0 and an anaphor DPANAPH

• crucially, ϕ-agreement is unbounded: there is no upper bound on the amount of structure or linear distance that the relation can span

— for a particularly vivid demonstration of this fact, I refer the reader to Keine (2017)

• who shows that long-distance agreement in Hindi can span across an unbounded number of transitive verb phrases

— and, moreover, that it is subject to exactly the same structural constraints as the licensing of wh-in-situ in Hindi is

⇒ and unbounded dependencies are the domain of syntax, not morphology

3This 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.

4While 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.
to put this another way:
for the AAE to be “morphological” in nature, there would have to be no upper bound on the size of a “morphological domain”
• a position that I take to be self-evidently indefensible

[(i) ∨ (ii)] must be true, and (ii) is false
⇒ (i) must be true.
➢ In other words:
  ○ languages like Japanese 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 in, say, Icelandic

• This is a welcome result:
  ○ it converges with independent arguments (in favor of the same conclusion) from the distribution and nature of Person Case Constraint (PCC) effects
    – see Preminger 2019
  • where I have termed this the no-null-agreement generalization

2. Reductionist theories of binding
• Recent years have seen a proliferation of proposals that subscribe to what I will call a reductionist view of binding:

(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

• It is important to note that the mention of binding index in (14) is a matter of convenience
  ○ and does not reflect any commitment to the ontological reality of indices (in syntax or in semantics)

➢ Concretely, “X and Y share a binding index” should be understood here and throughout as shorthand for “X and Y behave, binding-theoretically, the way that index-based theories of binding predict they would behave if they shared an index.”

• For some proponents of reductionism, it only applies to anaphoric binding —
  ○ e.g. Heinat 2008, Reuland 2011, Rooryck & Vandenhove 2011;
• Whereas for others, it applies to any kind of binding —
  ○ e.g. Kratzer 2009, Wurmbrand 2017; see also the discussion in Pollard & Sag 1994

➢ Now, there is plenty of evidence unrelated to the AAE that reductionism is off track

(i) directionality
  ○ φ-agreement transmits values upward in the structure
    – from a c-commanded goal to a c-commanding probe

  ○ binding, on the other hand, transmits values downward in the structure
    – from a c-commanded goal to a c-commanding probe

⇒ the idea that binding requires φ-agreement is antithetical to what we know from actual φ-agreement

*There have been attempts in the literature to challenge this claim concerning the directionality of φ-agreement. But these are either empirically inadequate (e.g. Zeijlstra 2012, as discussed in Preminger 2013), or require assumptions about φ-agreement that are known to be false. An example of the latter is Bjorkman & Zeijlstra (2019), whose theory depends on the assumption that, alongside valuation, every φ-agreement relation includes a checking component—an assumption known to be incorrect (Preminger 2011a, 2014).

*Not every reductionist approach runs afoul of this directionality consideration. The aforementioned work by Reuland (2011), for example, does not—as it makes use of covert 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).
(ii) the no-null-agreement generalization

- recall that the AAE itself implies that morpho-phonologically covert agreement is banned (§1.3) —
  - a result that enjoys independent support from the PCC (Preminger 2019)
- but it’s a truism that anaphoric binding exists even in languages that lack overt \(\varphi\)-agreement (e.g. Japanese)
  - which, by hypothesis, lack syntactic \(\varphi\)-agreement as well
- Japanese, after all, does have anaphors
  \(\Rightarrow\) the idea that binding requires \(\varphi\)-agreement is again antithetical to what we know from actual \(\varphi\)-agreement\(^7\)

(iii) \(\varphi\)-feature-matching is neither necessary nor sufficient for coreference

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

(16) \(<\text{pointing to different individuals in succession}>\)

You should leave, but you should stay here.

- moreover, it has been reported (by Collins & Postal) that we in (15) even has a bound reading
  \(\Rightarrow\) it seems incongruous that \(\varphi\)-agreement would be inexorably implicated in binding

- In the remainder of this talk, I will present an argument from the AAE against reductionism;
- Followed by a non-reductionist account of the AAE.

3. Step 1: \(\varphi\)-matching \(\not\Rightarrow\) \(\varphi\)-agreement

- One of the arguments frequently proffered in support of reductionist theories of binding:
  - matching in \(\varphi\)-features between binder and bindee

  ➢ In actuality, however, \(\varphi\)-feature-matching provides no support whatsoever for an involvement of \(\varphi\)-agreement in binding
- Specifically, \(\varphi\)-matching is enforced even in scenarios where syntactic \(\varphi\)-agreement could not possibly be involved
  - Donkey Anaphora; cross-utterance anaphora; and even linguistically-unantecedced deixis
- And this is so even when it comes to those \(\varphi\)-features which are not interpreted
  - e.g. grammatical gender on inanimates

- Consider:

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

(18) a. kol \text{exad} \text{fe-ye'j} \text{l-o} \text{maxberet fe-ya-sim}

\text{every one that-exist DAT-3sgM notebook(F) that-3sgM.FUT-put}
\text{ot-a/*ot-o} \text{ba-tik} \text{ACC-3sgF/*ACC-3sgM in.the-case}
\text{‘Everyone who has a notebook(F), put it.F/*it.M in your bag ’}

b. kol \text{exad} \text{fe-ye'j} \text{l-o} \text{maxjevon fe-ya-sim}

\text{every one that-exist DAT-3sgM calculator(M) that-3sgM.FUT-put}
\text{ot-o/*ot-a} \text{ba-tik} \text{ACC-3sgM/*ACC-3sgF in.the-case}
\text{‘Everyone who has a calculator(M), put it.M/*it.F in your bag ’}

- these are cases 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;
- as well as the fact that the antecedent is buried inside a Complex NP Island (of the relative-clause variety), inside a Subject Island

\(^7\)Sandhya Sundaresan (p.c.) points out that it is technically possible to maintain the ban on morpho-phonologically null finite \(\varphi\)-agreement, while assuming that \(\varphi\)-agreement between a DP and another DP (viz. binding) is not subject to the same ban. While I agree that this is technically possible, it seems ad hoc, and in any event, it decreases the amount of ‘reduction’ that reductionist theories can actually claim to have achieved, thus lessening their appeal in the first place.
Clearly, a syntax that can relate two expressions in the absence of c-command, and in a manner that disregards islands, is hopelessly unrestricted.

Cases like (17–18) are a clear indication that even \( \phi \)-features that are not interpreted, such as the grammatical gender features of inanimates, can be transmitted by a non-syntactic mechanism.

Let us refer to this mechanism—whatever it may be—as \( NSM \).

There are candidates in the literature for what this mechanism might be:

- for Elbourne (2013), pronouns are (hidden) definite descriptions
  - and so the matching requirements evidenced in (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 these are the residue of NP ellipsis
  - and so the matching requirements evidenced in (17–18) dissolve into whatever enforces identity of form between ellipses and their antecedents (cf. *John is no longer a bachelor, and Bill did <get married> too*).
  - but it doesn’t really matter for me what \( NSM \) is;
  - its modular affiliation is clearly not syntax, and that is the only point here.

Donkey Anaphora is not the only evidence for the existence of \( NSM \).

Consider cases of cross-utterance anaphora / linguistically-unantecedent deixis:

(19) A: Where are the scissors?  
B: They are right here.
   - these exhibit \( \phi \)-matching across different utterances, across different speakers.\(^8\)

As before, the same holds of grammatical gender on inanimates:

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

\[
\begin{align*}
\text{ir-a-kingu-ye} & \quad \text{ir-a-kingu-ye} \quad \text{(Kinyarwanda)} \\
4\text{SUBJ-PAST-OPEN-PREFV} & \quad 4\text{SUBJ-PAST-OPEN-PREFV} \\
\text{‘They are open.’} & \quad \text{‘They have been opened.’}
\end{align*}
\]

- crucially, the subj.-agreement marker in (20) is \( ir- \) because (plural) “doors” are a member of noun-class 4
  - rather than 2, 6, 8, and so forth.

These data illustrate quite vividly: there is absolutely no argument to be had from overt \( \phi \)-matching to an involvement of syntactic \( \phi \)-agreement.

Before concluding this section, it is worth pausing to entertain a putative swing of the pendulum all the way in the other direction —

- namely, the possibility of \( NSM \) subsuming all of what syntactic \( \phi \)-agreement was posited for.

Dowty & Jacobson (1988), discussing a related set of facts, hint (at the very least) at this possibility.

While one operation is certainly better than two, there is good evidence that \( \phi \)-agreement proper —

- that is, \( \phi \)-feature agreement between a verbal head and the verb’s argument(s)
  - must be syntactically underpinned.

Consider (21) (‘ \( \gg \) ’ represents c-command, and \( H \) is some functional head):

(21) \( H \gg DP_1 \gg DP_2 \)

  - \( H \) cannot enter into \( \phi \)-agreement with \( DP_2 \) across \( DP_1 \) in this case.\(^9\)

\(^8\)It 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 \( \phi \)-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.

\(^9\)For Dowty & Jacobson, \( NSM \) would reside in the semantic-pragmatic component, but that is not crucial for our current discussion.

\(^10\)This phenomenon of intervention in \( \phi \)-agreement is not to be confused with the phenomenon of dative intervention in A-movement, which Bruening (2014) and, more recently, Branen (2018), have challenged the very existence of. The discussion here concerns cases where \( DP_2 \) is not a candidate to
Importantly, such intervention in syntactic $\varphi$-agreement is systematically ameliorated by A-moving the intervener:

- see Anagnostopoulou (2003) and Holmberg & Hróarsdóttir (2003), a.o.

\[
\text{það finnst/*finnast [einverjum stúdent] tölvurnar ljótar.}
\]

\[
\text{EXPL find.sg/*find.pl some student.DAT computers.the.NOM ugly}
\]

‘Some student finds the computers ugly.’

(22) \[\text{það finnst/*finnast [einverjum stúdent] tölvurnar ljótar. (Icelandic)}\]

\[
\text{[Einverjum stúdent] finnast t₁ tölvurnar ljótar}
\]

\[
\text{some student.DAT find.pl computers.the.NOM ugly}
\]

‘Some student finds the computers ugly.’

(23) \[\text{Holmberg & Hróarsdóttir 2003:999–1000}\]

- Binding, on the other hand, behaves differently:\footnote{Thanks to an anonymous reviewer for helping me sharpen this argument.}

(24) a. The children seem to her to have t₁ amused Mary₁.
   \[\text{we know from (24a) that the to-Experiencer of \textit{seem} can bind into the embedded infinitive}
   \]
   \[\text{– at least when locality is not at issue (as is the case with Condition C)}\]
   \[\Rightarrow \text{what, then, is the reason for the failure of reflexive binding in (24b)?}\]
   \[\text{– if it is because of the intervening trace of \textit{the children}, this shows that A-traces count for binding}
   \]
   \[\Rightarrow \text{contra the behavior of \varphi-agreement (22–23)}\]

b. * The children seem to me to have t₁ amused myself.

c. The children₁ seem to me to have t₁ amused each other.

\[\Rightarrow \text{Overall: a Dowty & Jacobson-inspired swing in the other direction —}
\]

\[\text{– does not accord with the facts.}\]

4. Step 2: Evidence for $\varphi$-encapsulation (Middleton 2018)

- Middleton (2018): survey of 86 languages, from 13 language families

\[\Rightarrow \text{looking specifically at the forms that each language uses to express the following 4 meanings (labels are Middleton’s)}\]

\begin{itemize}
  \item “ANAPHOR”
    \[\text{Diana } \lambda x (x \text { thinks that only Charles } \lambda y (y \text { loves } y))\]
  \item “DIAPHOR”
    \[\text{only Diana } \lambda x (x \text { thinks that Charles } \lambda y (y \text { loves } x))\]
  \item “EXOPHOR”
    \[\text{only Diana } \lambda x (x \text { thinks that Charles } \lambda y (y \text { loves } z)), \text{ where } z = \text{Diana}\]
  \item “PRONOUN”
    \[\text{only Diana } \lambda x (x \text { thinks that Charles } \lambda y (y \text { loves } z)), \text{ where } z \neq \text{Diana}\]
\end{itemize}
KEY FINDING: a ban on discontinuous syncretism

(26) English: AAAB

a. Diana thinks that only Charles loves himself.
   ~ Diana \( \lambda x (x \text{ thinks that only } \lambda y (y \text{ loves } y)) \)

b. Only Diana thinks that Charles loves her.
   ~ only Diana \( \lambda x (x \text{ thinks that } \lambda y (y \text{ loves } x)) \)
   ~ only Diana \( \lambda x (x \text{ thinks that } \lambda y (y \text{ loves } z)), \text{ where } z = \text{ Diana} \)
   ~ only Diana \( \lambda x (x \text{ thinks that } \lambda y (y \text{ loves } z)), \text{ where } z \neq \text{ Diana} \)

(27) Icelandic: AABC (also: Malayalam, some dialects of Mandarin)

a. Díana telur að aðeins Karl elski sjálfan sig
   Díana believes that only Charles loves sjálfan sig
   ~ Diana \( \lambda x (x \text{ thinks that only } \lambda y (y \text{ loves } y)) \)

b. Aðeins Díana telur að Karl elski sig
   only Díana believes that Charles loves sig
   ~ only Diana \( \lambda x (x \text{ thinks that } \lambda y (y \text{ loves } x)) \)

c. Aðeins Díana telur að Karl elski hana
   only Díana believes that Charles loves hana
   ~ only Diana \( \lambda x (x \text{ thinks that } \lambda y (y \text{ loves } z)), \text{ where } z = \text{ Diana} \)
   ~ only Diana \( \lambda x (x \text{ thinks that } \lambda y (y \text{ loves } z)), \text{ where } z \neq \text{ Diana} \)

[Jane Middleton, p.c.]


• Caha (2009) (abridged): a similar ban on discontinuous syncretism, in the forms of nom-acc-dat case paradigms (see also Harðarson 2016, Zompí 2016)

(29) || ‘arm’ | ‘land’ | ‘queen’ || (Icelandic)
---|---|---|---|
nom | arm-ur | A | land-Ø | A | drottning-Ø | A |
acc | arm-Ø | B | land-Ø | A | drottning-u | B |
dat | arm-s | C | land-i | B | drottning-u | B |

[Harðarson 2016:1332]
➢ Caha’s (2009) proposal: *structural containment*

(see also Bobaljik 2012, Smith et al. 2016 for containment-based accounts of other bans on discontinuous syncretism)

(30)

\[
\begin{array}{c}
\text{DATATIVE} \\
\text{ACCUSATIVE} \\
\text{NOMINATIVE}
\end{array}
\]

\[
\begin{array}{c}
d \\
a
\end{array}
\]

○ How this works:
  - assume that there can be Vocabulary Insertion rules for non-terminals
  - when more than one of these rules is applicable, the one chosen is the one with the most specified (i.e., structurally-rich) insertion context\(^{12}\)
    (this is the *Superset Principle* of nanosyntax; see Caha 2009:55)

\(^{12}\)As an illustration, consider a hypothetical vocabulary that includes the following items:

(i) a. Num\(^0\)\(\text{P} \quad \Rightarrow \quad \text{mouse} (/mæUs/)

\[
\begin{array}{c}
\text{MOUSE} \\
\end{array}
\]

b. Num\(^0\)\(\text{P} \quad \Rightarrow \quad \text{mice} (/mæJs/)

\[
\begin{array}{c}
\text{MOUSE} \\
\end{array}
\]

With these insertion rules in place, a structure like (ii) could only be spelled out as mice (/mæJs/), and not mouse (/mæUs/) or mouses (/mæUs/+z/), because the insertion rule in (i.b) is both applicable and more specific than the one in (i.a).

(ii) a. Num\(^0\)\(\text{P} \quad \Rightarrow \quad \text{mice} (/mæJs/)

\[
\begin{array}{c}
\text{MOUSE} \\
\end{array}
\]

now consider a language with different forms for the nom and the acc
  - i.e., a pattern that starts with AB,

○ there are two options for how the dat is formed:

(i) the dat is cumulative relative to the acc
  - this would arise if there was no specific insertion rule targeting the dative node
  ⇒ the spellout of the dat would be the spellout of the d node, affixed to the spellout of the acc
  - if the spellout of d is null, the result will be form-identical to the acc, resulting in ABB
  - if, on the other hand, the spellout of d is non-null, then, setting cases of accidental homophony aside, the result will be ABC

(ii) the spellout of the dat is the result of a specific Vocabulary Insertion rule targeting the dative node in (30)
  - this would also yield an ABC pattern (again, setting accidental homophony aside)

➢ crucially, there are only two possible derivations that would lead to the undesired ABA pattern, and both involve accidental homophony:

(i) a Vocabulary Insertion rule that applies to the dative node and inserts phonological material that is accidentally identical to the spellout of the nominative node

(ii) there is no insertion rule that applies to the dative node, but:
  - there is one that applies to the accusative node;
  - the latter is truncative relative to the spellout of the nominative node (i.e., it idiosyncratically inserts a substring of the nom form; cf. the acc of ‘arm’ in (29))
  - and the spellout of the d node just happens to be identical to that truncated portion
    (this would be the case if, e.g., the dat suffix on ‘arm’ in (29) had been -ur, contrary to fact)

○ assuming that such cases of accidental homophony are ruled out —
  - by the learning mechanism if not by the grammar itself
    — the impossibility of ABA patterns (i.e., discontinuous syncretism in the domain of nom-acc-dat case forms) is derived
As noted by Caha (2017):
- a ban on discontinuous syncretism is not an unambiguous indicator of containment (see also Bobaljik & Sauerland 2018)
- It can also arise through partial featural overlap
- Crucially, however, evidence of strict morpho-phonological cumulativity in a given empirical domain militates against such an account, and in favor of a containment-based one
  - see Middleton (2018:12–21) for discussion

And, lo and behold:

(31) \[
\text{ANAPHOR} \quad \text{awake dheen dheewe} \\
\text{DIAPHR/EXOHYP} \quad \text{awake dheen} \\
\text{PRONOUN} \quad \text{dheen}
\] (PJS)

\[\Rightarrow\] Therefore:

(32) \[
\text{ANAPHOR} \\
\text{DIAPHR} \\
\text{EXOHYP} \\
\text{PRONOUN}
\] [Middleton 2018]

\[\Rightarrow\] Two things about (32):
- first, it is necessarily the case that (32) represents a universal, cross-linguistically invariant structure for the expressions in question
  - if we relax this assumption:
    - the explanation for why there is not a single example of discontinuous syncretism in Middleton’s (fairly vast) survey is lost
- second, if all we are interested in is the locus of \(\varphi\)-features relative to the locus of anaphoricity —
  - then (32) reduces to (33):

(33) \[
\text{AnaphP} \\
\text{Anaph}^0 \\
\text{PhiP} \\
\text{Phi}^0 \\
\cdots
\] \[\Rightarrow\] (33) is the partial structure of (32) that’s relevant to the relation between \(\varphi\)-features and anaphoricity

- In other words, what we have seen in this section is that \(\varphi\)-features are properly contained within the structural layer that turns an expression into an anaphor
  - I refer to this as the \(\varphi\)-encapsulation hypothesis.
5. The challenge to reductionism from the AAE

- Recall the reductionist position, repeated below:

\[(34)\]
\[\alpha\text{ and }\beta\text{ can share a binding index only if }\alpha\text{ and }\beta\text{ have entered into syntactic agreement in }\varphi\text{-features}\]

- where, for the purposes of this definition, entering into \(\varphi\)-agreement is subject to transitive closure
  - i.e., if \(\alpha\) agrees with \(\gamma\), and \(\gamma\) agrees with \(\beta\), then \(\alpha\) and \(\beta\) count as having entered into agreement with one another for the purposes of this definition

\[\Rightarrow\]

Taken at face value, (34) predicts the systematic absence of anything like the AAE

- That is because:
  - anaphors are, by definition, bound;
  - and by (34), binding entails \(\varphi\)-agreement;
  - and an anaphor is, by definition, a target with valued \(\varphi\)-features\(^\text{13}\)

\[\Rightarrow\]

successful, nontrivial agreement with anaphors should be the natural state of affairs —
  - in direct opposition to (35):

\[(35)\] ANAPHOR AGREEMENT EFFECT

\[\ast H^0 \ldots \text{DP_ANAPH}, \text{where } R \text{ is a nontrivial } \varphi\text{-agreement relation } \Rightarrow (6)\]

\[\Rightarrow\]

Two approaches suggest themselves in response to this apparent paradox, and I will discuss them each in turn

- an approach based on derivational timing (§6)
- an approach based on structural encapsulation (§7)

\[\text{\textbf{6. A timing-based approach to the AAE}}\]

- One way around the apparent paradox identified in §5:
  - assume that the anaphor has not yet acquired valued \(\varphi\)-features at the derivational stage at which it is targeted for \(\varphi\)-agreement
  - On this view, anaphors begin their derivational life in a \(\varphi\)-deficient state —
  - and they acquire \(\varphi\)-feature values via the very \(\varphi\)-agreement relation implicated in the reductionist position (34)
  - Suppose this is so; let us see what this entails for anaphoric binding, \(\varphi\)-agreement, and reductionism . . .

\[\text{\textbf{6.1. Reflexives in Basque}}\]

\[\text{\textbf{6.1.1. The basics}}\]

\[(36)\]
\[a. <\text{pro.2pl.ERG}> [ \text{zeuen buru-a } ] \text{ saldu d-\varphi-u-zue} \quad 2\text{pl(strong).poss head-ART}_{sg} (\text{ABS}) \text{ sold } 3\text{A-sgA-}\sqrt{2}\text{-2plE}\]
\[\text{‘Y’all have given yourselves away’ (lit. ‘Y’all have sold y’all’s head.’)}\]
\[\text{\textit{[Artiagoitia 2003:620]}\]

\[b. <\text{pro.2pl.ERG}> [ \text{zuen buru-a } ] \text{ saldu d-\varphi-u-zue} \quad 2\text{pl(weak).poss head-ART}_{sg} (\text{ABS}) \text{ sold } 3\text{A-sgA-}\sqrt{2}\text{-2plE}\]
\[\text{‘Y’all have given yourselves away’ (lit. ‘Y’all have sold y’all’s head.’)}\]

- The structure of Basque reflexives:

\[(37)\] \text{[PRON.GEN N D]}

where:

- \text{PRON.GEN can be a strong or weak possessive pronoun;}
- and \text{N} is a designated noun \text{buru} (“head”)

- As the examples in (36a–b) demonstrate, Basque obeys the AAE:
  - the reflexives (bracketed above) trigger invariant 3rd person agreement
  - Morphologically speaking, the head noun \text{buru} is indeed 3rd person
  - while the possessor bears \(\varphi\)-features matching those of the antecedent

\[\Rightarrow\]

For the purposes of this section, I’m assuming for the sake of argument that once these feature-values are present on the possessor —
  - they are also visible from outside the entire anaphor

\[\text{\textbf{6.1.2. Reflexives with possessive pronouns}}\]

\[\ast H^0 \ldots \text{DP_ANAPH}, \text{where } R \text{ is a nontrivial } \varphi\text{-agreement relation } \Rightarrow (6)\]

\[\Rightarrow\]

Two approaches suggest themselves in response to this apparent paradox, and I will discuss them each in turn

- an approach based on derivational timing (§6)
- an approach based on structural encapsulation (§7)

\[\text{\textbf{6.1.3. Reflexives with proper nouns}}\]

\[\ast H^0 \ldots \text{DP_ANAPH}, \text{where } R \text{ is a nontrivial } \varphi\text{-agreement relation } \Rightarrow (6)\]

\[\Rightarrow\]

Two approaches suggest themselves in response to this apparent paradox, and I will discuss them each in turn

- an approach based on derivational timing (§6)
- an approach based on structural encapsulation (§7)

\[\text{\textbf{6.1.4. Reflexives with proper nouns}}\]

\[\ast H^0 \ldots \text{DP_ANAPH}, \text{where } R \text{ is a nontrivial } \varphi\text{-agreement relation } \Rightarrow (6)\]

\[\Rightarrow\]

Two approaches suggest themselves in response to this apparent paradox, and I will discuss them each in turn

- an approach based on derivational timing (§6)
- an approach based on structural encapsulation (§7)

\[\text{\textbf{6.1.5. Reflexives with proper nouns}}\]

\[\ast H^0 \ldots \text{DP_ANAPH}, \text{where } R \text{ is a nontrivial } \varphi\text{-agreement relation } \Rightarrow (6)\]

\[\Rightarrow\]

Two approaches suggest themselves in response to this apparent paradox, and I will discuss them each in turn

- an approach based on derivational timing (§6)
- an approach based on structural encapsulation (§7)

\[\text{\textbf{6.1.6. Reflexives with proper nouns}}\]

\[\ast H^0 \ldots \text{DP_ANAPH}, \text{where } R \text{ is a nontrivial } \varphi\text{-agreement relation } \Rightarrow (6)\]

\[\Rightarrow\]

Two approaches suggest themselves in response to this apparent paradox, and I will discuss them each in turn

- an approach based on derivational timing (§6)
- an approach based on structural encapsulation (§7)
This could be because of:

- feature-percolation (from the possessor to the entire possessum DP); or
- the head noun *bura* behaving in a manner analogous to pseudo-partitives
  - i.e., projecting an extended nominal projection that is transparent to
    \( \varphi \)-agreement

(The alternative, whereby valued \( \varphi \)-features on the possessor of *bura* are
not visible from outside the reflexive anaphor, is what I will actually be
arguing for; see §7 onwards.)

### 6.1.2. A note about number in Basque reflexives

- In cases where the binder (and, consequently, the possessive pronoun) is plural,
  the head-noun *bura* can optionally be plural as well
  - e.g. in the ABS: *bura-a* (“head-\( \text{ART}_\text{sg}(\text{ABS}) \)” or *bura-ak* (“head-\( \text{ART}_\text{pl}(\text{ABS}) \)”)
- But notice that, given that *bura* can still be singular even when antecedent &
  possessive pronoun are plural —
  - it is quite clearly *not* the number features associated with the binding index
    that are being expressed here
- Rather, this is probably a ‘dependent plural’ effect (cf. 
  *The plastic surgeons gave each other a new nose / new noses*)

⇒ I leave this issue aside here\(^{14}\)

### 6.1.3. Are we “digging in the right place”??

- At this juncture, one may wonder if reflexives in Basque are really what we
  should be focusing on —
  - given that they have this particular possessed-body-part structure
    (much like Georgian reflexives, discussed in §1.1)
  - But recall the results of §4:

(38) THE \( \varphi \)-ENCAPSULATION HYPOTHESIS

\[ \begin{array}{c}
\text{AnaphP} \\
\text{Anaph}^0 \\
\Phi P \\
\Phi^0 \\
\end{array} \]

- what Basque (and Georgian) shows overtly —
  - a pronominal, \( \varphi \)-bearing layer, properly contained in a separate
    structural layer (in this case, the layer headed by the body-part noun)
    that renders the expression anaphoric

— is precisely what’s going on, albeit in a way that is morphologically less
transparent, even in a language like, e.g., Albanian:

(39) *Vetja më dhimset.*

\[ \text{self.NOM Cl.1sg.DAT feel.sorry.for.3sg.pres.nonactive} \]

‘I feel sorry for myself.’

\( \Rightarrow \) This has to be so, otherwise we lose the explanation for the universality of the
no-discontinuous-syncretism result (see §4)

- Let me be perfectly clear about this: given Middleton’s results —
  - an expression like the Albanian *vetja* necessarily includes a pronominal
    layer, properly contained in an anaphoric layer

    – just like reflexives in Basque/(Georgian) do

\( \Rightarrow \) Reflexives like those in Basque are exactly what we should be looking at —
  - since they show transparently that which is going on in all reflexives!

### 6.1.4. Anaphoric binding in Basque transitives

- As argued in detail by Arregi & Nevins (2012), the head responsible for
  \( \varphi \)-agreement with ABS DPs in Basque is \( T^0 \)

\( \Rightarrow \) rather than, say, \( t^0 \)

\(^{14}\)Though I note (following a suggestion by a reviewer) that this may be taken to support a variation
of the current proposal in which the AAE is defined in terms of \textit{person} features only (see Abramovitz
2019, who develops precisely such an alternative).
⇒ Consider, now, what this entails for a timing-based approach to the AAE:

(Basque is head-final, of course; diagram is head-initial for purely graphical reasons)

(40) [Diagram of TP structure]

‘Y’all have given yourselves away’
(lit.: ‘Y’all have sold y’all’s head.’)  [= (36)]

- By hypothesis, given the timing-based approach, the ϕ-bearing portion of the Basque reflexive —
  - the possessive pronoun, zeuen/zuen
  (‘2pl(strong/weak).poss’(=‘y’all’s’))
  — will receive its ϕ-features once it is bound by its antecedent (in this case, pro.2pl.ERG)
- And note: the trace in [Spec,vP] is an A-trace —
  - as illustrated in (41a–c) (repeated from earlier), A-traces can & do bind

(41) a. The children\(_1\) seem to her\(_3\) to have \(t_1\) amused Mary\(_1\).  [= (24a–c)]
  b. * The children\(_1\) seem to me to have \(t_1\) amused myself.
  c. The children\(_1\) seem to me to have \(t_1\) amused each other.

⇒ The ϕ-bearing subpart of the reflexive will acquire its ϕ-features before \(T^0\) probes for the ϕ-features of the ABS anaphor

➢ This, in turn, means that a timing-based account of the AAE won’t work in Basque
  - or for any other language where \(T^0\) agrees with the internal argument in a transitive clause\(^{15}\)

6.2. Further Issues

(i) PHASES:
Can an approach that makes use of phases (and/or cyclic spellout) salvage the timing-based account of the AAE in Basque?
- The answer, I argue, is “no”:
  - if ϕ-agreement between α and β is a necessary condition for binding of β by α (=the reductionist position) —
    - then there can be no relevant phase boundary in between the subject and object in (40)
      - since the subject is able to bind the object (and, by hypothesis, agree with it)

(ii) FEATURE-SHARING:
There is arguably another problem, unrelated to anything discussed in this subsection so far, with the timing-based approach.
- The problem is that syntactic ϕ-agreement does not “check features” (see Preminger 2011a, 2014);
- Nor does it “copy” feature-values from one place to another;
- Instead, it creates feature-sharing structures

\(^{15}\)A significant subset of ergative languages fit this description. These are the “high-abs” language in Coon et al.’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 et al. 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).
• The result of $\varphi$-agreement between a direct-object anaphor and whichever $\varphi$-probe agreed with it would be a feature-sharing structure like (42)

- Once binding did occur —
  - and, by the reductionist hypothesis, the features of the direct-object anaphor were valued
  - the resulting valuation would affect the anaphor and the probe equally

⇒ Meaning we would see full, nontrivial agreement with the antecedent
  - contrary to the AAE facts.

• And this would be so even if we were looking at a language where the relevant agreement probe was $\nu^0$, rather than $T^0$
  - e.g. “ABS=DEF”/“LOW-ABS” languages (Coon et al. 2014, Legate 2008)

• That is because, due to the nature of feature-sharing, the relative order of $\varphi$-probing and binding-cum-$\varphi$-agreement does not matter.

➢ Crucially, however: even if we were to reject feature-sharing as a syntactic mechanism —
  - it would not affect the argument against timing-based approaches from languages where the probe that targets the internal argument is $T^0$
  (see §6.1.4)

6.3. Summary

• §6.1–§6.2:
  A timing-based approach to reconciling the AAE with reductionism fails.
7. An encapsulation-based approach to the AAE

- **Upshot of §4:**

\[ \text{(43) THE } \varphi\text{-ENCAPSULATION HYPOTHESIS} \]

\[
\begin{array}{c}
\text{AnaphP} \\
\text{Anaph}^0 \\
\text{Phi}^0 \\
\text{...}
\end{array}
\]

- **Upshot of §6.1:**

\[ \text{(44) TP} \]

\[
\begin{array}{c}
\text{pro.2pl.ERG}_1 \\
\text{T'} \\
\text{t}_1 \\
\text{sP} \\
\text{have} \\
\text{v'} \\
\text{y'} \\
\text{V'} \\
\text{sold} \\
\text{DP}_{\text{ABS}=(\text{AnaphP})} \\
\end{array}
\]

\[
\begin{array}{c}
\text{DP}_{\text{POSS}}=\text{PhiP} \\
\text{D'} \\
\text{NP} \\
\text{N}^0 \\
\text{head}
\end{array}
\]

\[ \text{‘Y’all have given yourselves away’} \]
\[ \text{(lit.: ‘Y’all have sold y’all’s head.’)} \]

- Suppose the AAE arises because \( \varphi \)-agreement comes upon a structural layer —
  - AnaphP in (43) (instantiated by DP\(_{\text{ABS(anaph.)}}\) in (44)) — that prevents access to the feature-values hosted on the \( \varphi \)-bearing portion of the reflexive
  - PhiP in (43) (instantiated by DP\(_{\text{POSS}}\) in (44))

- That such a layer exists does not need to be stipulated; it follows from Middleton’s results, surveyed in §4

- All we’re adding here is the assumption that this universally-present layer has the relevant properties to halt \( \varphi \)-probing
  - be it because it is phasal;
  - or because it bears its own set of valued \( \varphi \)-features, distinct from the ones that covary with the antecedent;
  - or both.

➢ This clearly suffices to derive the AAE:
  - \( \varphi \)-matching between the PhiP layer (DP\(_{\text{POSS}}\) in (44)) and the antecedent is enforced by NSM (see §3)
  - but a \( \varphi \)-probe located properly outside the AnaphP layer (DP\(_{\text{ABS(anaph.)}}\) in (44)) will run into this opaque layer
    - eliminating the possibility of nontrivial \( \varphi \)-agreement
    - and yielding the AAE (repeated below)

\[ \text{(45) ANAPHOR AGREEMENT EFFECT} \]

\* \( \text{H}^0 \ldots \text{DP}_{\text{ANAPH}}, \text{where } \mathcal{R} \text{ is a nontrivial } \varphi \text{-agreement relation} \]

\[ \text{[{=} (6, 35)]} \]
8. Encapsulation and reductionism

8.1. A false prediction

- The reductionist position, one more time:

(46) **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

⇒ That means that in a structure like (47) —

(47) AnaphP

  Anaph0

  PhiP

  Phi0

  · · ·

  [=(8, 33, 38, 43)]

— it is necessarily PhiP that is bound by the antecedent (!)

- This yields a testable prediction

- Because AnaphP and PhiP are distinct structural layers (§4), and PhiP is the one bound by the antecedent:16

(48) **REDUCTIONIST PREDICTION:**

the outermost layer of a reflexive anaphor should behave as if it is not the bearer of the relevant binding index

- In light of this, consider (49a–b):

(49) a. Johni expects Mary to outdo himi/k.

  [Norvin Richards, p.c.]

  - notice first that (49a) does not give rise to a disjoint-reference effect between John and the pronoun him

    - meaning John is too far away from the pronoun, structurally speaking, for the two to enter into a local binding relation

  ⇒ crucially, this means that the cause of the disjoint-reference effect observed in (49b) must be the anaphor, himself

  ➢ but this could only be the case if the binding index resided on the outermost projection of the anaphor

    - in contradiction to the predictions of the reductionist hypothesis.

8.2. Cross-linguistic variation?

- Iatridou (1988):

  - reflexive anaphors in Greek have a possessed-body-part structure (like that seen earlier for Georgian and Basque);

  - and they can be clitic-doubled (just like other DPs in Greek can);

  - clitics that double full DPs (in Greek, and more generally) typically behave as though they carried the same referential index as their doubled DP does;

  - finally, clitics in Greek can independently be shown to obey Condition B.

  ➢ now, suppose that the binding index on Greek reflexives resided on their outermost layer:

    - clitic-doubling an anaphor that is bound by the local subject (in accordance with Condition A) would give rise to a clitic, which —

      - being itself bound by the local subject

    — would trigger a Condition B effect

  ⇒ crucially, no such Condition B effect arises in this case

  ⇒ suggesting that the binding index does not, in fact, reside on the outermost layer of the Greek reflexive.17

---

16 It 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 §6). 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. (46)), and nothing more would need to be said in this section.

17 I thank an anonymous reviewer for bringing this work to my attention.
• **Is this a problem for the argument in §8.1...?**
  
  o first, note that when juxtaposed with the results in §8.1 (in particular, exx. (49a–b) and the surrounding discussion) —
    
    - Iatridou’s results suggest a genuine point of cross-linguistic variation
      
      - since English anaphors do behave as though the binding index resides on their outermost layer

  o but are the Greek results truly problematic for \( \varphi \)-encapsulation?
    
    - recall that, as discussed in §8.1, reductionism entails that the binding index must reside on PhiP, the bearer of valued \( \varphi \)-features
      
      - an entailment that was shown to be false, at least in English

  ➢ but, crucially, rejecting reductionism does not entail that binding indices cannot reside on PhiP ever (i.e., in any language)

  ➢ nor does rejecting reductionism entail that if the binding index does not reside on PhiP in a given language, it must reside specifically on AnaphP —

    
    - it could reside, in Greek, on some other layer in the extended projection of the anaphor, that is neither of these two

⇒ That there is a language that behaves the way Iatridou shows that Greek behaves does not change the status of the argument in §8.1 one bit.

### 8.3. Summary

- **§8.1–§8.2:**
  Reductionism, when juxtaposed with the facts of \( \varphi \)-encapsulation, yields false predictions about the structural locus of binding indices (e.g. in English).

- **Note:** This result has a different status than, e.g., the result in §6 showing that timing-based approaches cannot reconcile reductionism with the AAE.

  o to that result one could respond, correctly, that it does not prove the requisite negative
    
    - *viz.* that there is no reconciling reductionism with the AAE

  o it merely eliminates one contender for doing so (timing-based accounts)

  ➢ but the results from this section are more broad —

    
    - showing that if \( \varphi \)-encapsulation holds (and there’s every reason to believe it does), then reductionism just won’t work.

### 9. Remaining issues

#### 9.1. Non-AAE-obeying languages

- The account presented in the previous sections derives the AAE as a universal property of all anaphors in all languages
  
  o its universality deriving from the universality of \( \varphi \)-encapsulation (§4)

  ➢ Murugesan (2018, 2019), however, shows that some languages genuinely violate the AAE

    
    - i.e., exhibit nontrivial agreement with anaphors

#### 9.1.1. Tamil: the basics

- Tamil (Dravidian): NOM-ACC language where only NOM DPs control verbal agreement

  (50) Meena Kohli.y.ai paar-t-aal Meena.F(OM) Kohli.M-ACC see-PAST-3sGf

  ‘Meena.F saw Kohli.M.’ [Murugesan 2018:(13)]

- Tamil also has DAT-subject verbs, whose object then surfaces as NOM

  - and, unsurprisingly, it is then this NOM object that controls agreement

  (51) Kohli-ukku Meena kidai-t-aal Kohli.M-DAT Meena.F(OM) get-PAST-3sGf

  ‘Kohli.M got Meena.F.’ [Murugesan 2018:(14)]

- No accessible NOM DP \( \rightarrow \) only 3sGn agreement

  (52) a. Kohli.ukku pasi-t-atu/*aan Kohli.M-DAT hungry-PAST-3sGn/*3sGm

  ‘Kohli.M was hungry.’


Crucially, there appears to be nontrivial agreement with the anaphor \textit{taan}:\footnote{Jeffrey Lidz (p.c.) informs me that a similar pattern obtains in Kannada (Dravidian).}

\begin{enumerate}
\item[a.] Kohli-ukku \textit{taan} tirumba kidai-t-\textit{aan} \\
\quad Kohli\textit{M} -DAT \textit{REFL}(NOM) again \quad \text{get-PAST-3sgM} \\
\quad ‘Kohli\textit{M} got himself back again.’
\item[b.] Meena-ukku \textit{taan} tirumba kidai-t-\textit{aal} \\
\quad Meena\textit{F} -DAT \textit{REFL}(NOM) again \quad \text{get-PAST-3sgF} \\
\quad ‘Meena\textit{F} got herself back again.’ \quad [Murugesan 2018:(15a–b)]
\end{enumerate}

\textit{and, in light of (51–52), we can be quite sure that the source of non-3sgN agreement in (53a–b) is not the \textit{DAT} antecedent.}

\section*{9.1.2. Murugesan’s (2018, 2019) analysis}

\begin{itemize}
\item Murugesan’s (2018, 2019) analysis of these facts appeals to the relative structural height of the $\varphi$-probe relative to the antecedent:
\begin{itemize}
\item he proposes that the AAE arises in those languages where the $\varphi$-probe is located below the antecedent, and therefore enters the structure earlier
\item – at a point when the anaphor has not yet been bound
\item whereas in those languages where the $\varphi$-probe is located above the antecedent, the anaphor has already been bound
\item – and thus the AAE does not arise
\end{itemize}
\item Tamil, on Murugesan’s account, is a language of the latter type
\begin{itemize}
\item assuming that the $\varphi$-probe is located on T$^0$ and the subject is base-generated lower, e.g. in [Spec,\textit{vP}]
\end{itemize}
\item Notice that this account is a timing-based account
\item And furthermore, it requires reductionism
\begin{itemize}
\item because the difference between a not-yet-bound anaphor and a bound one consists, on this account, in the difference between not having valued $\varphi$-features and having valued $\varphi$-features
\end{itemize}
\end{itemize}

\begin{itemize}
\item Crucially, we have already seen that neither assumption is tenable:
\begin{itemize}
\item timing-based approaches do not work (§6)
\item nor does reductionism, even independent of derivational timing (§8)
\end{itemize}
\item Nevertheless, much of the empirical burden in arguing against these approaches was carried by languages like Georgian, Basque, etc.
\begin{itemize}
\item where anaphors are transparently analytic, involving a possessor-possessum structure in which the possessum is a body-part noun
\item \Rightarrow Could we retreat from the conclusions drawn from such data, and reinstate a timing-based account of the kind Murugesan envisions?
\item The answer, I think, is “no.”
\item Recall that our explanandum here is the AAE, \textit{viz.} the ban on nontrivial agreement with anaphors
\item If we endorse the retreat in question, languages like Georgian and Basque would then stand as a twofold coincidence:
\begin{itemize}
\item first, we’d have to say that the reason languages like Basque and Georgian exhibit the AAE is unrelated to (and, in fact, disjoint from) the reason why other languages exhibit the AAE
\begin{itemize}
\item in Basque transitive clauses with a reflexive object, the \textit{ABS} $\varphi$-probe is located on T$^0$ (Arregi & Nevins 2012)
\item above both the anaphor and the base position of the antecedent
\item \Rightarrow Basque is precisely the kind of language predicted \textit{not} to exhibit the AAE, on Murugesan’s account
\item on this view, the reason Basque (as well as Georgian) exhibits the AAE is the possessed-body-part structure of its reflexives
\item which is entirely unrelated to the reason other languages exhibit the AAE (the antecedent being higher than the $\varphi$-probe)
\end{itemize}
\item second, this purportedly idiosyncratic structure of reflexives, which causes the AAE in languages like Basque and Georgian, is in fact what Middleton (2018) has shown holds \textit{universally}
\begin{itemize}
\item anaphors involve additional structural layers encapsulating the $\varphi$-feature-bearing layer of structure
\item whether it happens to be transparently detectable in a particular language or not
\end{itemize}
\end{itemize}
\item Nevertheless, much of the empirical burden in arguing against these approaches was carried by languages like Georgian, Basque, etc.
\begin{itemize}
\item where anaphors are transparently analytic, involving a possessor-possessum structure in which the possessum is a body-part noun
\item \Rightarrow Could we retreat from the conclusions drawn from such data, and reinstate a timing-based account of the kind Murugesan envisions?
\item The answer, I think, is “no.”
\item Recall that our explanandum here is the AAE, \textit{viz.} the ban on nontrivial agreement with anaphors
\item If we endorse the retreat in question, languages like Georgian and Basque would then stand as a twofold coincidence:
\begin{itemize}
\item first, we’d have to say that the reason languages like Basque and Georgian exhibit the AAE is unrelated to (and, in fact, disjoint from) the reason why other languages exhibit the AAE
\begin{itemize}
\item in Basque transitive clauses with a reflexive object, the \textit{ABS} $\varphi$-probe is located on T$^0$ (Arregi & Nevins 2012)
\item above both the anaphor and the base position of the antecedent
\item \Rightarrow Basque is precisely the kind of language predicted \textit{not} to exhibit the AAE, on Murugesan’s account
\item on this view, the reason Basque (as well as Georgian) exhibits the AAE is the possessed-body-part structure of its reflexives
\item which is entirely unrelated to the reason other languages exhibit the AAE (the antecedent being higher than the $\varphi$-probe)
\end{itemize}
\item second, this purportedly idiosyncratic structure of reflexives, which causes the AAE in languages like Basque and Georgian, is in fact what Middleton (2018) has shown holds \textit{universally}
\begin{itemize}
\item anaphors involve additional structural layers encapsulating the $\varphi$-feature-bearing layer of structure
\item whether it happens to be transparently detectable in a particular language or not
\end{itemize}
\end{itemize}
\end{itemize}
– 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 §4 for details)
⇒ far from being an exception, these languages are transparent exemplars
  of the universal structure of anaphors
– yet their behavior, on Murugesan’s account, would have to be cast as
  some sort of outlier

9.1.3. 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, taan simply cannot be bound by (grammatically) 1st/2nd person
  antecedents
• For these cases, Tamil instead uses forms that are indistinguishable from the
  1st/2nd person pronominals in the language
  ◦ behavior that is reminiscent of the reflexive in Romance

(54) En-akk\textsubscript{u}\ naan\textsubscript{u} tirumba kidai-tt-een
  1sg-DAT 1sg-NOM(PRON) again get-PAST-1sg
  ‘I got myself\textsubscript{u} back again.’ [Sandhya Sundaresan, p.c.]
⇒ While agreement with anaphors in Tamil does qualify as nontrivial —
  · given, e.g., the bona fide variance in gender features seen in (53a–b)
  — it would also be imprecise to characterize it as full-fledged
  · given that the only true anaphor is restricted to 3rd person \(\varphi\)-features
• One could therefore imagine a further attenuation of how we define the AAE
  such that nontrivial agreement in person features is the crucial element
  ◦ see Abramovitz 2019 for an approach along these lines
➤ But since number agreement with anaphors is not generally permitted
  (see, e.g., Amiridze 2003 on the AAE in Georgian)
  — it seems to me that adopting a person-only AAE misses an important part of
  the picture.

• Second, there are cases that appear to instantiate agreement with a (nom)
  subject anaphor
  ◦ as opposed to the (nom) object anaphors we saw earlier

(55) a. Mani\textsubscript{t} [taan\textsubscript{t} satath-ai sapi-t\textsuperscript{a}aan-nnu]\son-n-aan
  \text{Mani}\textsubscript{t} self rice-ACC eat-PAST-3sgM-comp say-PAST-3sgM
  ‘Mani\textsubscript{t} said that self\textsubscript{t} ate the rice.’
b. Banu\textsubscript{F} [taan\textsubscript{F} satath-ai sapi-t\textsuperscript{a}aal-nnu] son-n-aal
  \text{Banu}\textsubscript{F} self rice-ACC eat-PAST-3sgF-comp say-PAST-3sgF
  ‘Banu\textsubscript{F} said that self\textsubscript{F} ate the rice.’ [Murugesan 2018:(46a–b)]
  ◦ but as Murugesan points out:
    – Sundaresan (2016) has shown that this is actually a case of agreement
      with a perspective-holder
      · and, as such, need not target one of the core arguments at all

(56) Banu\textsubscript{F} [taan\textsubscript{F}/\*ja sataat-ai sapi-t\textsuperscript{a}een-nnu] so-n-aal
  \text{Banu}\textsubscript{F} self rice-ACC eat-PAST-1sg-comp say-PAST-3sgF
  ‘Banu\textsubscript{F} said that self\textsubscript{F}/\*ja ate the rice.’ [Murugesan 2018:(48)]

9.1.4. Weighing the options

• Suppose we did not pursue an attenuation of the AAE to person features only
  ➤ Consider what each competing account (Murugesan’s, and the one proposed
    here) would then have to do to accommodate the data that remains recalcitrant
    for that account:

• We have already discussed the pitfalls of excluding languages like Basque —
  · languages in which the \(\varphi\)-probe is situated higher than the typical
    antecedent
  — from the purview of an AAE account
  · see §9.1.2 for details
⇒ There doesn’t 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 \( \varphi \)-encapsulation account pursued here, languages like Tamil could simply be a case where AnaphP is exceptionally not syntactically opaque

\[
\text{(57)}
\]

\[
\begin{array}{c}
\text{AnaphP} \\
\text{Anaph}^0 \\
\Phi^0 \\
\Phi P \\
\cdots
\end{array}
\]

\[
[= (8, 33, 38, 43, 47)]
\]

- either because the normally-phasal AnaphP is exceptionally non-phasal in these languages;
- or because the normally \( \varphi \)-bearing AnaphP instead behaves like a pseudo-partitive
  - in contrast to our conclusions about Basque (see §6.1)

• The former, phase-based approach has parallels in Abels (2003) analysis of preposition stranding\(^{19}\)
  - 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)

• This case would be similar:
  - AnaphP would be phasal in the vast majority of languages —
    - but can be deemed non-phasal by the learner, in light of positive evidence (viz. nontrivial agreement with anaphors)\(^{20}\)

• It is also suggestive, in this regard, that AAE-violating languages like Tamil are roughly as common as preposition-stranding languages like English
  - that is, very rare (modulo the usual caveats on the pitfalls of counting languages)

---

\(^{19}\)Abels (2012) later pursues an alternative, wherein PP is always phasal, and the difference between P-stranding and non-P-stranding languages consists in the presence of a (possibly unpronounced) morpheme between P and its supposed complement.

\(^{20}\)See also Preminger (2019), for an argument from an unrelated empirical domain that children are sensitive to overt morpho-phonological covariance (i.e., nontrivial agreement) in setting the parameters of their language.

9.2. The fate of attempted agreement with anaphors

• One thing we haven’t really come back to is:
  - what happens when a derivation occurs in which agreement with an anaphor is attempted

• We know that it won’t be successful (nontrivial) agreement, because:

\[
\text{(58) ANAPHOR AGREEMENT EFFECT}
\]

\[
\begin{array}{c}
\text{ANAPH} \\
\text{R} \\
\text{DPANAPH}
\end{array}
\]

\[
[= (6, 35, 45)]
\]

- But this still leaves us with two possible outcomes:
  - (i) a grammatical sentence with a default / nonvarying agreement form
  - (ii) ungrammaticality

• Rizzi (1990) was operating on the assumption that only (ii) was possible;

• But we’ve seen that that’s clearly false
  - instead, we find variation;
  - in Icelandic, the result is outright ungrammaticality (ii):

\[
\text{(59) Sigga telur [ að mér líki hún/\*sig\_ ]}
\]

\[
\text{Sigga thinks \_ that me.DAT likes.SBJV.3sg she.NOM/\*SIG}
\]

\[
\text{‘Sigga thinks that I like her.’} \quad [= (11)]
\]

- in Albanian, the result is a (grammatical) default (i):

\[
\text{(60) Vetja më dhimset.}
\]

\[
\text{self.NOM cl.1sg.DAT feel.sorry.for.3sg.PRES.NONACTIVE}
\]

\[
\text{‘I feel sorry for myself.’} \quad [= (3, 39)]
\]

• Furthermore, as pointed out by Murugesan (2018):
  - the Icelandic pattern, where sig in agreeing positions results in outright ungrammaticality, is surprising in and of itself
  - that is because, as shown in (61a–b):
    - actual agreement with nominative objects in Icelandic exhibits optionality (see also Hornstein 2018)
(61) a. Henni leiddist þeir
    she.DAT was.bored.by.3sg they.NOM
    ‘She was bored with them.’

b. Henni leiddust þeir
    she.DAT was.bored.by.3pl they.NOM
    ‘She was bored with them.’

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 w.r.t. the AAE.

There are simplex anaphors that:
◦ can occur in the relevant positions with invariant 3sg agreement
  (Albanian)\(^{21}\)
◦ cannot occur in the relevant positions at all (Icelandic)

There are complex anaphors that:
◦ can occur in the relevant positions with invariant 3sg agreement (Basque)
◦ cannot occur in the relevant positions at all (Italian)

There are ϕ-varying anaphors that:
◦ can occur in the relevant positions with invariant 3sg agreement (Basque)
◦ cannot occur in the relevant positions at all (Italian)

There are ϕ-invariant anaphors that:
◦ can occur in the relevant positions with invariant 3sg agreement (Inuktitut)\(^{22}\)
◦ cannot occur in the relevant positions at all (Icelandic)

Similarly, the properties of the language don’t seem to be predictive of these behaviors, either:
◦ Basque is an erg-abs language and its reflexive is complex
  — and its reflexive occurs with invariant 3sg agreement
◦ Albanian is a nom-acc language and its reflexive is simplex
  — and yet it behaves exactly like Basque
◦ Icelandic is like Albanian in being nom-acc and having a simplex reflexive
  — but behaves the opposite way with respect to the fate of this reflexive in
  positions that would otherwise trigger agreement

This is important 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 for anaphors
— is not only problematic from the perspective of the facts surveyed in §4;
➢ It also provides no apparent help in understanding the phenomena at hand.

Finally:
◦ while the encapsulation hypothesis provides no particular insight into
whether you’ll get (i) (a grammatical sentence with invariant ϕ-agreement)
or (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 one assumed that such agreement gives rise to default, 3sg agreement —
  ⇒ then Icelandic, English, and languages like them would remain
  unexplained

\(^{21}\)On the simplex nature of Albanian vetja, see Franks (2013).
\(^{22}\)On Inuktitut reflexive anaphors, see Yuan (2018).
As it concerns English and Icelandic, one could appeal to a morphological gap in the nominative cell of the reflexives paradigms; 

But since Albanian does not have such a gap, this seems like a restatement of the explananda, rather than an explanation.

Moreover, Rizzi (1990:34) provides good reason to think that the Italian pattern (which is essentially English/Icelandic-like) could not possibly arise because of a paradigm gap

- since nom-acc syncretisms in the 1st/2nd person in Italian mean that the form of the allegedly "missing" reflexives is in fact fully predictable

Overall:

- there doesn't seem to be a theory of the outcomes of AAE violations (default vs. ungrammaticality) currently in the offing
- and this doesn't distinguish the approaches under considerations here (timing and/or reductionism, vs. encapsulation)

10. Conclusion

- The AAE is about a ban on (nontrivial) agreement with an anaphor
- not about restrictions on the positions that anaphors can occur in
- ϕ-feature-matching between binder and bindee is in no way an argument for any involvement of syntactic ϕ-agreement
- Anaphors involve structural encapsulation (following Middleton 2018)
- A timing-based reductionist approach to the AAE doesn't work
- In fact, any reductionist approach doesn't work, in light of the AAE
- Encapsulation, without reductionism (which is untenable anyway), suffices to explain the AAE
  - or at least those aspects of the AAE that anyone currently purports to have an explanation for

References


23Thanks to an anonymous reviewer for help here.


