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

Omer Preminger*
University of Maryland

1. A very brief overview of the AAE

- The Anaphor Agreement Effect (henceforth, AAE):
  - an avoidance of ϕ-feature agreement with anaphors
    - or ~
  - an avoidance of anaphors in positions targeted by ϕ-feature agreement

- Example:

(1) a. Húnk sagðisaid [að sigk vantaði peninga].
   Shek said that shek lacked money.
   (Icelandic)

b. Sigga telur [að méér líki húnk/*sigk]
   Sigga thinks that me.dat likes sbjv she.nom/*refl(nom)
   ‘Sigga thinks that I like her.’
   [Maling 1984:216–217]

- the embedded verbs in (1a–b) assign quirky case (acc and dat, respectively) to their subjects;
- in quirky-subject clauses, it is the object, rather than the subject, that controls agreement

⇒ the refl. sig is grammatical in the subject position of such clauses (as shown in (1a))
⇒ and ungrammatical in the object position of such clauses (as shown in (1b))

2. What this talk is & isn’t about

- The AAE seems to suggest that anaphora and ϕ-agreement interact with each other quite closely
  - But as pointed out by Woolford (1999), inter alia, this fails to generalize cross-linguistically:
    - it is common for languages with case-marking but no agreement to allow nominative anaphors

(2) sensek-ni(-wa) zibunk-ga wakar-ani-i
   teacher-dat(-topic) self-nom understand-NEG-PRES
   ‘The teacherk does not understand themselvesk.’
   [Shibatani 1977:800, via Woolford 1999:263]

- An important caveat:
  - this cross-linguistic story requires a ban on null agreement
    - as argued, on independent grounds, in Preminger (2019)
  - if Japanese is allowed to have phonologically-null agreement with the nominative, we lose our explanation for why an anaphor is bad in (1b) but good in (2)
    - they’d both involve agreement with the nominative in the syntax, with Japanese just happening not to expone this at PF

⇒ the explanation for the contrast in the acceptability of a nominative reflexive would be lost
  - Keep this in mind...

- Internal to Icelandic, it looks like one could also capture the effect in terms of case — as a ban on nominative anaphors

- Now, there is plenty of evidence unrelated to the AAE showing that anaphoric binding does not reduce to ϕ-agreement
  - seec, e.g., Reuland (2011:261–262)

- Also relevant here is that in Icelandic dative-nominative ECM constructions, pronominal disjoint-reference effects (i.e., Condition B effects) seem to covary with whether or not the nominative infinitival subject has been agreed with (see Taraldsen 1995, Murugesan et al. 2017). This is equally unexpected if the AAE is a case-theoretic ban on nominative anaphors.
One such source of evidence concerns directionality: anaphoric binding transmits values downward in the structure
- from a c-commanding binder to a c-commanded bindee

Whereas ϕ-agreement only ever transmits values upward in the structure
- from a c-commanded goal to a c-commanding probe
  - but see also Bjorkman & Zeijlstra (2019), who inadvertently provide a fairly strong argument against downward-valuation in ϕ-agreement —
    - by making explicit the vast and problematic set of assumptions that would be required to support such a theory²

⇒ The idea that anaphora is underpinned by ϕ-agreement flies in the face of what we know about the structural properties of the two relations.

Another source of evidence that anaphoric binding does not reduce to ϕ-agreement concerns the ban on null agreement
- I’ve argued in Preminger (2019) that there is generally no such thing as null ϕ-agreement
  - i.e., syntactic agreement in ϕ-features that is not exponed—either as “pure” agreement or as clitic doubling—simply does not exist
- and, as noted above, such an assumption is independently necessary in order for the AAE to even be stateable, cross-linguistically

But it is a truism that anaphora exists where no overt exponence of ϕ-agreement is found
- e.g. there are many languages where there is no overt ϕ-agreement
- but those languages still have anaphoric binding

⇒ The idea that anaphora is underpinned by ϕ-agreement requires assumptions (viz. the existence of null agreement) that make the AAE unstateable in the first place.

Today, I’d like to present a different kind of argument against a reduction of anaphora to ϕ-agreement
- also from the AAE
- I will do this by taking a careful look at what a reductionist theory entails for accounts of the AAE.

3. Reductionist theories in the harsh light of the AAE

Some utility definitions:

(3) Let $F_\varphi$ be the formal process or relation that values the ϕ-features on a functional head using the ϕ-feature values found on one or more DPs.
  - e.g. Chomsky’s 2001 Agree — or, for those who prefer a theory of ϕ-agreement that actually works, Preminger’s 2014 $F_{\text{find}}$

(4) Let DP[A] be the anaphor, and let $\varphi(DP[A])$ be the ϕ-features borne by DP[A] at a given stage of the derivation.

(5) Let DP[B] be a potential binder, and let $\varphi(DP[B])$ be the ϕ-features borne by DP[B] at a given stage of the derivation.

(6) Let $H^0$ be the ϕ-probe that putatively agrees with DP[A] (i.e., the ϕ-probe that, if DP[A] were replaced with a non-anaphoric nominal $\alpha$, would enter into $F_\varphi$ with $\alpha$).

3.1. ϕ-(in)completeness

- Foundational assumptions for this approach:
  - anaphors are referentially dependent
  - referential dependence is syntactically represented by ϕ-feature deficiency

- Immediately, we need to recognize two logical possibilities:

²To name one example, Bjorkman & Zeijlstra’s (2019) proposal requires that all ϕ-agreement include a checking component alongside valuation—an assumption that we know is false (see Preminger 2011, 2014).
(i) $\varphi$-deficiency is a matter of ‘unvaluedness’
   - that is, anaphors can, structurally speaking, carry the full range
     of $\varphi$-feature values that other DPs carry
     - they just happen not to, at the derivational stage at which
       they are targeted for $\varphi$-agreement
   - i.e., $\varphi(DP[A]) \neq \varphi(DP[B])$ when $F_\varphi(H^0, A)$ applies
     - and changes yielding $\varphi(DP[A]) = \varphi(DP[B])$ occur only later
       in the derivation

(ii) $\varphi$-deficiency is a matter of inability to carry $\varphi$-feature values at all
   - i.e., $\varphi(DP[A]) \neq \varphi(DP[B])$ throughout the course of the
     derivation

- While logically possible, it seems clear that we can dismiss (ii) outright
  - since it is the case that anaphors in many languages show the full
    range of $\varphi$-feature distinctions available to other DPs in the language

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

- So this brings us back to (i): anaphors lack $\varphi$-features at the derivational
  stage at which they are targeted for $\varphi$-agreement
  - which leads us to . . .

INTERLUDE: What if ‘anaphor’ is not a natural class?
- The inference just drawn, as well as others I will be drawing in
  this talk, only goes through if ‘anaphor’ is a natural class
  - otherwise, it is unsound to carry over conclusions from one
    member of the class to the others
- And there is a growing body of work suggesting that there are in fact
  different types of anaphors (see, e.g., Sundaresan to appear)
- That said, different types do not yet imply that ‘anaphor’ is not a
  natural class (cf. ‘plosive’)
- In fact, if it is not, then the AAE is (definitionally) an epiphenomenon
  - and couldn’t possibly provide support for a reduction of anaphora
    to $\varphi$-agreement (or for anything else, for that matter)
- And so, my work here would, in some sense, be done!

⇒ I will continue to assume there is a natural class ‘anaphor’
  - WORKING DEFINITION: those nominal elements whose reference is
    necessarily dependent on another nominal in the sentence
~ Please enjoy this gratuitous doodling space! ~
(Handout continues on next page.)
3.2. \( (\text{timing-based approach to AAE}) \land (\text{binding-as-agreement}) = \bot \)

- As already noted, a timing-based approach to the AAE requires the following to hold:

\[ (7) \quad \text{NECESSARY CONDITION FOR TIMING-BASED ACCOUNTS OF THE AAE} \]

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

- I will now demonstrate that (7) is incompatible with a reduction of anaphora to \( \varphi \)-agreement

- If the reduction of anaphora to \( \varphi \)-agreement holds, then, by hypothesis, \( F_{\varphi} \) (e.g. \( \text{Agree}_\varphi \)) underpins anaphoric binding, as well

- In order to follow along with the next part, it may be useful to envision a rather straightforward anaphoric binding scenario —
  - binding of a reflexive direct object by its clausemate subject, in a language with \( \varphi \)-agreement between \( v \) and the direct object

\[ (8) \quad \text{Diagram of an anaphoric binding scenario} \]

- On the reductionist view, the following must be true of \( F_{\varphi} \):
  - (i) \( F_{\varphi} \) is able to apply countercyclically
    - changing (valuing) the features on a syntactic object, e.g. \( \text{DP}[A] \), long after that object is no longer near the root of the tree
    - at most, one could envision a demand whereby the \textit{supplier} of the relevant values, e.g. \( \text{DP}[B] \), must be near the root of the tree
  - (ii) \( F_{\varphi} \) does not treat the verb phrase (VP/\( v \)/\( \text{VoiceP} \)/\(<\text{whatever}>\)) as a relevant locality boundary
    - whether this is because the verb phrase is never a phase (Keine 2017; see section 5.1); or because \( F_{\varphi} \) is actually composite, and mediated by intervening functional heads
    - (it makes no difference for our current concerns)

- Now here’s where the rubber meets the road:
  - if \( F_{\varphi} \) is countercyclic, \( v^0 \) should be able to probe \( \text{DP}[A] \) after the latter’s features have been valued by \( \text{DP}[B] \)
    \[ \Rightarrow \text{premise (7) is false} \]
  - even if we add the assumption that at least one of the operands of \( F_{\varphi} \) must be near the root of the tree —
    - then at the point that \( \text{DP}[B] \) is merged, if \( v^0 \) still lacks \( \varphi \)-features, \( \text{DP}[B] \) should value it
    - in fact, because \( v^0 \) and \( v \text{P} \) are one and the same object (\textit{Bare Phrase Structure}; Chomsky 1994):
      - \( \text{DP}[B] \) would have to value \( v^0/v\text{P} \) and not \( \text{DP}[A] \), at least initially; anything else would violate minimality
    \[ \Rightarrow \text{again, showing that premise (7) is false} \]
  - at best, this hypothesis predicts that anaphors would never show \( \varphi \)-feature-matching with their antecedents, but the proximate \( \varphi \)-probes would behave as though they did (which is obviously incorrect) \( \Rightarrow \bot \)

- \textbf{QUESTION:} Could the problem be \textit{phases}...?
  - perhaps all of the above does happen, but \( v^0 \) has already been spelled out by the time \( \text{DP}[B] \) tries to value its \( \varphi \)-features
  - and that’s why we never see valuation of \( v^0/v\text{P} \) by \( \text{DP}[B] \)

- \textbf{ANSWER:} To the extent that this is true of \( v^0 \), it is also true of \( \text{DP}[A] \)
  \[ \Rightarrow \text{falsely predicting that anaphors could never exhibit } \varphi \text{-feature-matching with their binders} \Rightarrow \bot \]
  (To put this another way: a reduction of anaphora to \( \varphi \)-agreement requires assumption (ii), above, and so phases are not a possible explanation for what’s going on here.)

- Finally, I’ve been assuming that the relevant copy of \( \text{DP}[B] \) is the one in [Spec,TP]
  - to give the timing story its best chance at success

- If the relevant copy is the one in [Spec,\( v \text{P} \)], things are \textit{at least} as bad (and possibly worse) for the timing story
  - since there is even less time between the merger of \( v^0 \) and \( \text{DP}[A] \) and the merger of \( \text{DP}[B] \).
But wait! Things actually get even worse for the reductionist view —
  o if $F_\varphi$ is supposed to underpin $\varphi$-agreement, then there’s a few other things we know about it:
    - $F_\varphi$ does not “check features”;
    - nor does $F_\varphi$ “copy” features values from one head to another;
  ➤ rather, $F_\varphi$ creates a feature-sharing structure
    - wherein multiple syntactic nodes become linked to a single feature structure

$$v^0 \text{ DP}[A] \quad \Rightarrow \quad v^0 \text{ DP}[A]$$


⇒ meaning any subsequent $\varphi$-values associated with DP[A] will, automatically and unavoidably, show up on $v^0$ as well

All in all, it’s clear that there’s overwhelming reason to reject a timing-based account of the AAE, when coupled with a reduction of anaphoric binding to $\varphi$-agreement

### 3.3. Timing without reductionism

- A super-duper important caveat: the contradictions noted above only obtain if we insist on reducing anaphoric binding to $\varphi$-agreement

⇒ If we accept that anaphoric binding is a formally separate process —
  o which is counter-cyclic and downward-valuing
    — and unrelated to $\varphi$-agreement —
      o which is cyclic and upward-valuing
    — then we can maintain premise (7) (repeated here)
      o and derive the AAE from it

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

By the time we are done, I will have argued that this is not quite the right approach to the AAE, either

- nevertheless, it is a logical possibility;
- but only if we accept that $\varphi$-agreement and anaphoric binding are formally distinct, as detailed above.

### 3.4. Encapsulation

- Given that we cannot maintain a timing-based account of the AAE while reducing anaphoric binding to $\varphi$-agreement —

  o let us consider an alternative, which I’ll refer to as encapsulation

- Suppose that the anaphor does enter into $F_\varphi$ with DP[B] —

  o but that the constituent whose $\varphi$-features are consequently valued is only a proper subpart of DP[A]

```
(10) \begin{array}{c}
\text{DP}[A] \\
\text{D}^0[A] \\
\text{InnerP} \\
\text{Inner}^0 \\

\Rightarrow \text{When probed from the outside, DP}[A] \text{ will appear to lack the relevant } \varphi\text{-feature values transmitted from DP}[B]  \\
\Rightarrow \text{since those will reside on InnerP (and/or Inner}^0\text{), and be inaccessible from the outside}
\end{array}
```

- But now notice:
  If anaphoric binding reduces to $F_\varphi$ (=the reductionist hypothesis) —

  ⇒ it follows that InnerP, and not DP[A], is what bears the binding index
    ⇒ and, consequently, it follows that the binding index on anaphors does not c-command out of DP[A]

- This is a testable prediction:

  (11) a. John expects Mary to outdo him$_{k}$.
  b. John expects himself$_{i}$ to outdo him$_{k}$. [Norvin Richards, p.c.]

  o ex. (11a) does not induce a disjoint-reference effect on the pronoun
    meaning John is too far away from the pronoun to enter into a binding relation

  ⇒ the cause of the disjoint-reference effect in (11b) must be the reflexive anaphor, himself
but this could only be the case if the binding index on \textit{himself} resided 
on the outermost projection of the anaphor
  
  – in contradiction to (10)
⇒ The encapsulation hypothesis might be right (in fact, I will argue shortly 
that it is) —
  
  ◦ but only if the \(\varphi\)-features of the anaphor and its binding index are not 
linked to one another
  
  ◦ i.e., only if the reduction of anaphoric binding to \(\varphi\)-agreement is 
incorrect.

Here’s another argument to the same effect —

• Consider reflexives in Basque
• These have the structure in (12):

(12) \[\text{[PRON.GEN N D]}\]
  
  ◦ where PRON.GEN is a (strong or weak) possessive pronoun, and N is 
the noun \textit{buru} (“head”)
  
  (in line with a common crosslinguistic strategy of forming reflexives from 
possessed body-part nouns)

(13) a. \(<\text{pro}_{2\text{pl}.\text{erg}}\>\text{ [zeuen buru-a ] saldu d-}\varphi\text{-u-zue} \)
    \text{2pl(\text{strong}).GEN head-\text{ART}_{\text{sg}}(\text{ABS}) sold} \text{ 3A-\text{sgA}-}\varphi\text{-2plE}
    \text{‘Y’all have given yourselves away’}
    \text{(lit.: ‘Y’all have sold y’all’s head.’)}

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

\(\text{[Artiagoitia 2003:620]}\)

\textbf{NB:} both examples are also possible with a plural head noun (\textit{buru-ak} 
“head-\text{ART}_{\text{pl}}(\text{ABS})”; Artiagoitia 2003:621)
  
  – cf.: \textit{The plastic surgeons gave each other a new nose / new noses.}

• As the \text{ABS} agreement morphology in exx. (13a–b) makes clear, Basque 
obeyes the AAE:
  
  ◦ Basque reflexives don’t exhibit “true” \(\varphi\)-agreement;
  
  ◦ they are agreed with as \(3\text{sg}\) expressions regardless of the \(\varphi\)-features of 
their referent or antecedent
    
    – the relevant sentences are in fact all ambiguous with a (usually 
nonsensical) reading where they are literally about someone’s 
body-part, \textit{head}
  
  • Under a view that assumes that \(F_\varphi\) underpins anaphoric binding:
    
    ◦ these facts entail that the binding index associated with this reflexive 
binding relation cannot reside at the level of the outermost DP —
      
      – since the \(\varphi\)-features of the antecedent \textit{are} tracked by the reflexive, 
but not at the level of the outermost DP
    
    ◦ instead, the binding index must reside on the \(\varphi\)-bearing subpart of the 
reflexive, i.e., on PRON.GEN

• But this is incongruous:
  
  ◦ the possessor in these Basque examples is pronominal, not anaphoric;
  
  ◦ it is the kind of expression that is subject to Condition B of the 
binding theory, \textit{not} Condition A

• Consider:

(14) Mirande\(\_\text{erg}\) \[\text{[bere-ak buru-a ]} \text{hil z-}\varphi\text{-u-en} \]
    \text{Mirande.\text{\text{ART}}_{\text{sg}}(\text{ABS}) killed 3A.\text{sgE}-\text{sgA}-}\varphi\text{-past}
    \text{‘Mirande killed himself.’}
    \text{(lit.: ‘Mirande killed his head.’)}
    \text{[\text{Artiagoitia 2003:621]}\]

(15) Peio\(_\text{er}\) \text{erransay d-}\varphi\text{-u-\_} \[\text{[bere-ak zakurradog.]} \text{dela} \]
    \text{Peio.\text{\text{ART}}_{\text{sg}}(\text{ABS}) die \text{aux.that}}
    \text{‘Peio said [that his dog died].’}
    \text{[\text{western dialects; Artiagoitia 2003:626]}\]

⇒ Placing the binding index on the possessor in (14) would render it entirely 
mysterious that:
  
  ◦ the possessor in (14) \textbf{requires a local antecedent};
  
  ◦ while the one (15) \textbf{does not}.\(^4\)

\(^3\)Thanks to Karlos Arregi, Aitor Lizardi Ituarte, and Juan Uriagereka, for helpful discussion.

\(^4\)See Amiridze (2003) for related observations about Georgian.
4. Interim Summary

- We have seen that the AAE provides circumstantial evidence against the reduction of anaphoric binding to ϕ-agreement
  - by rendering untenable a timing-based approach to the AAE (straightforward logical incompatibility)
  - and by rendering untenable an encapsulation-based approach to the AAE (evidence from English)
- And we have seen that the AAE also provides direct evidence against such a reduction (evidence from Basque)

⇒ Any proposal that assumes such a reduction (Kratzer 2009, for example) is necessarily wrong.

★ This leaves us with two fairly pressing questions:
  1. What accounts for the AAE?*
  2. What accounts for ϕ-feature-matching between the reflexive and its binder?*

*Now that we’ve seen that anaphoric binding most certainly does not arise via \( F_\phi \)

5. Whence the AAE?

5.1. Against a timing approach more generally

- As noted in section 3.3, the results to that point were compatible with a theory of the following kind:
  1. anaphoric binding and ϕ-agreement have nothing to do with one another; and
  2. the AAE arises because of cyclicity
    - in particular, the ϕ-probe probes the anaphor before the latter has received its ϕ-features
- But notice, now, that given the results of section 3.4, this is at best a kind of “conspiracy”:
  - we wouldn’t expect the ϕ-features on the Basque reflexives —
    - buried, as they are, in the DP-internal possessor
  - to be accessible from the outside
    - even if they made it there “in time” for ϕ-probing

There’s more:

- First, Keine (2017) shows—quite convincingly, in my mind—that the verb phrase (e.g. vP) never constitutes a phase
- In a nutshell, the argument is as follows:
  - both the licensing of wh-in-situ and long-distance agreement (LDA) in Hindi are disrupted by even a single intervening CP boundary

(16) * siita-ne soc-aa [\( \text{CP} \mid \text{ki ravii-ne kis-ko dekh-aa } \)]
Sita-erg think-perf.M.sg that Ram-erg who-acc see-perf.F.sg
Intended: ‘Who did Sita think that Ravi saw?’


(17) larḵo-ne soc-aa/*-ii [\( \text{CP} \mid \text{ki monaa-ne ghazal } \)]
boys-erg think-perf.M.sg/*-perf.F.sg that Mona-erg ghazal.F
gaa-yii thiī sing-perf.F.sg be.perf.F.sg
‘The boys thought that Mona had sung ghazal.’


- but they can cross an unbounded number of verb-phrase peripheries (e.g. in cases of infinitival embedding)

(18) a. tum [inf. kyaawhat kar-naado- ] jaan-te ho?
you [what do-inf.M.sg know-imperf.M.pl be.pres.2pl]
‘What did you know to do?’

b. ? raam [inf. kyaawhat khaa-naa ] phir-se shuruu kar-naa
caa-hā phir-se
want-imperf.M.sg be.pres.3sg
‘What does Ram want to start to eat again?’
[Keine 2017:183]
‘Ram wanted to do something futile.’ (idiomatic)  
[Keine 2017:179]

‘Ram wanted to start to do something futile.’ (idiomatic)  
[Keine 2017:180]

◦ crucially, the examples in (18–19) involve the relevant relation (wh or LDA) crossing multiple transitive verb-phrase boundaries:
  – two transitive verb phrases in each of (18a, 19a),
  and three transitive verb phrases in each of (18b, 19b)
◦ by itself, this might suggest an analysis involving A-movement of the wh-phrase/ LDA-target
  – since A-movement can escape (some) infinitives, but not (most) finite clauses

However, Keine has cleverly selected, in the wh-element kyaa and in
the idiom chunk in (19), elements that are independently known to
resist movement
  – even short, clause-internal scrambling
  (see Keine 2017:182–183, Bhatt & Keine 2017 for details)
⇒ the relations in question are bona fide in-situ relations
⇒ CP is a phase, but a verb-phrase (e.g. vP)—even a transitive one—is not

• Second, we know that ϕ-agreement is (to a limited degree) counter-cyclic
  ◦ in that an instance of ϕ-agreement that would fail on a strictly cyclic
derivation can be rescued by a later operation in the same phase:

(20) [Einverjum stúdent] fínnast t1 tôlvunar ljótar. 
‘Some student finds the computers ugly.’

(21) það fínnast*/fínnast [einverjum stúdent] tôlvunar ljótar. expl. find.sg/*find.pl some student.dat computers.dat ugly
‘Some student finds the computers ugly.’

(22) Hverjum 1 hafa2 strákarnir3 t2 virst t1 t3 vera gáfaðir?
who.dat have.pl boys.the.nom seemed be.inf intelligent
‘Who did the boys seem intelligent to?’
[Holmberg & Hróarsdóttir 2003:999–1000, 1010]

工程项目

• crucially, the intervention-ameliorating movement in (22) (chain 1) is
  C0-related movement
  – thus, the amelioration of intervention in (22) is, to a limited
degree, counter-cyclic
  (given that agreement happens in Icelandic at TP)

• Now, let’s put the last two pieces together:
  1. there is no phase boundary verb-phrase level (VP/vP/VoiceP)
  2. an instance of ϕ-agreement that would fail can be repaired by
     operations that happen later in the same phase
⇒ Anaphoric binding, if it truly transferred ϕ-features from the binder to the
anaphor, could feed successful agreement with the anaphor
  (at least in the maximally local case, where binder and anaphor are in
the same CP)

(23) TP

This shows that a timing-based approach to the AAE is wrong even if
anaphoric binding and ϕ-agreement are formally distinct(!)
5.2. In search of an account of the AAE: encapsulation again

Let’s step back for a moment and ask, in a fairly theory-neutral way:

- What is it that a theory of the AAE needs to deliver?

It needs to deliver an avoidance of (24a) and (24b) holding of a single XP:

(24) a. XP bears an anaphoric index.

b. XP has been target for φ-agreement.

It strikes me that a likely account here is one that appeals to substantives:

- Let us call the kind of elements that bear anaphoric indices AnaphPs (which may or may not, in the end, unify with more familiar projections);

- And let us call the kind of elements that bear φ-features PhiPs (which, again, may or may not, in the end, unify with more familiar projections).

- Let us furthermore suppose that the case of Basque is indicative of something systematic, in that:

  - AnaphP is structurally higher than PhiP (when both are present)

This state of affairs should sound familiar to any practicing syntactician

- it is par for the course to assume that certain substantives cannot “cohabitate” on a single head

- for example, syntactic verbs (say, v or V) are not—perhaps ever—the kinds of things that bear tense;

  - that is the purview of a separate syntactic object, namely T/Infl

- this is not to be confused with the fact that the v/V and T/Infl often end up spelled out as part of the same morphological complex;

- that much, in fact, is going to be true for many cases of AnaphP and PhiP, too (though not in Basque)

- this is precisely why the attempt to unify binding with agreement was not self-evidently moribund from the get-go:

  - when two things are frequently spelled out in the same morphological complex, it takes extra work to show that they are separate projections (cf. Emonds 1970, 1976 on Infl vs. V)

- All that is required to derive the AAE, then, is the assumption that AnaphP (when present) is syntactically opaque

  - on a par with, say, PPs and all non-nominative/non-absolutive cases in many languages (Demirok 2013)

- Now, at this juncture, you might think to yourself: “Well now this is a pretty ad hoc solution to the problem.”

- One thing that can be said in response:

  - given our current results, it seems to be the only remaining contender

- But more interesting, in this regard, is recent work by Middleton (2018):

  - Middleton has looked at 86 languages (from 13 language families)

    - specifically, at the forms used to convey the following four meanings:

(26) a. “ANAPHOR”

Diana λx (x thinks that only Charles λy (y loves y))

b. “DIAPHOR”

only Diana λx (x thinks that Charles λy (y loves x))

c. “EXOPHOR”

only Diana λx (x thinks that Charles λy (y loves z), z = Diana)

d. “PRONOUN”

only Diana λx (x thinks that Charles λy (y loves z), z ≠ Diana)

- and what she shows is:

  - the four forms that each language uses for the four most-embedded nominal expression in (26a–d) are subject to a no-discontinuous-syncretism constraint

  - e.g. English is ABBB (herself, her, her, her)

  - but no language is, e.g., ABCA
moreover, some languages show transparent containment in their morphology
  – e.g. Peranakan Javanese of Semarang (PJS; Cole et al. 2007).

(27) a. ANAPHOR: awake dheen dhwew
    b. DIAPHOR/EXOPHOR: awake dheen
    c. PRONOUN: dheen

⇒ from these two facts (the absence of discontinuous syncretism, and the existence of transparent containment) —
  – it follows that the only system that could capture these facts is one that involves structural containment
    - see Bobaljik (2012) and Bobaljik & Sauerland (2018) for detailed discussion

(28) \[
\begin{array}{c}
\text{ANAPHOR} \\
\text{DIAPHOR} \\
\text{y} \\
\text{EXOPHOR} \\
\text{z} \\
\text{PRONOUN} \\
\text{w}
\end{array}
\]

⇒ We therefore have converging evidence for the kind of structure in (25), above
  – where the anaphoric layer is distinct from, and higher than, the phi-bearing layer

6. Whence \(\varphi\)-matching?

- Given everything said in this talk so far:
  – it cannot be that anaphors match the \(\varphi\)-features of their binders due to syntactic \(\varphi\)-feature agreement (what I have called \(F_\varphi\))

⇒ How does \(\varphi\)-feature-matching between anaphor and binder come about?

- HERE’S A SKETCH OF AN ANSWER:
  – it comes about in the meaning component
    - via the meaning contribution of \(\varphi\)-features
    - in particular, let’s take as a starting point the presuppositional theory of \(\varphi\)-feature meanings (as in Heim & Kratzer 1998)

- Let \(Z(S)\) be a function from a set of \(\varphi\)-features \(S\) to the set of individuals for which \([S]\) is defined
  – roughly: \(Z(S)\) returns the set of individuals that the \(\varphi\)-set \(S\) “refers to”

- Now, if a DP bears a set of \(\varphi\)-features \(S\):
  – the interpretation assigned to the DP is in \(Z(S)\) (definitionally)

⇒ Consequently, if an anaphor bears index \(i\):
  – the individual \(g(i)\) (where \(g\) is the assignment function) must fall within the set identified by \(Z(S)\)

- E.g. if the antecedent carries plural \(\varphi\)-features, and bears index \(i\), then:
  – any bindee bearing index \(i\) will only be interpretable if its set of \(\varphi\)-features \(S\) carves out a set of individuals that includes \(g(i)\)
    - i.e., plurals

- This generally ensures that pairs of expressions bearing the same syntactic index will bear the same sets of \(\varphi\)-feature values.

⇒ Now, at this juncture, you might be saying to yourselves:
  “But what about \(\varphi\)-features that aren’t interpreted (pluralia tantum / grammatical gender on inanimates / etc.)?”

\(^5\) awake literally means “body” in PJS.
These are actually not problematic on the present approach:
- if $S = \{\text{plural}\}$, then by the definition just given, $\llbracket \text{“scissors”} \rrbracket \subseteq Z(S)$
  - and similarly for grammatical gender on inanimates

**Question:** Doesn’t this run counter to the notion of “interpretation function” in the first place?
- after all, features like plural on pluralia tantum or grammatical gender on inanimates don’t seem to be “interpreted” in the usual sense

**Answer:** Whatever our intuitions on this matter are, these conclusions are forced on independent grounds

**Example:**

(29) Speaker A “Where are the scissors?”
Speaker B “They are right here.”
- consider the expression they in Speaker B’s utterance
- it would be logically incoherent to speak of a syntactic relation (Agree, $\phi$, whatever) holding between this expression and the expression the scissors in Speaker A’s utterance
- that is because:
  - syntactic relations are grammatical entities;
  - a grammar is, by definition, a mental object;
  - and minds are, by definition, confined to individual speakers.
  ⇒ for they to refer to the scissors in this instance:
  - it must be the case that $S = \{\text{plural}\} \rightarrow \llbracket \text{“scissors”} \rrbracket \subseteq Z(S)$
  (as just suggested)

The same is true for grammatical gender on inanimates:
- in Kinyarwanda, for example, if one is pointing to a pair of doors, one can say:

(30) ir-a-kingu-ye
  4SUBJ-PAST-OPEN-PREFV
  ‘They are open.’ (‘They have been opened.’)
- but the agreement marker in (30) is ir- because (plural) doors are a member of class 4 (rather than class 2, 6, 8, etc.)
  ⇒ it must be the case that $S = \{\text{class 4}\} \rightarrow \llbracket \text{“doors (pl.)”} \rrbracket \subseteq Z(S)$
  (again, as just suggested)

**Crucially,** this mechanism seems sufficient to account for $\phi$-feature matching between anaphors and their binders
- given that both the anaphor and its binder carry the same index $i$:
  - it follows that $g(i) \in Z(S_{\text{anaphor}})$ and $g(i) \in Z(S_{\text{binder}})$
  - assuming that $\phi$-features create, at a reasonable approximation, a partition over the members of the domain of individuals:
  - it follows from this that $S_{\text{anaphor}} = S_{\text{binder}}$

- Perhaps this mechanism can ultimately be subsumed under, say, Elbourne’s (2013) view of pronouns as definite descriptions, or Merchant’s (2014) view of them as the residue of ellipsis;
- but this is immaterial for our current purposes;

- All that matters for us here is that we know for sure that the mechanism in question is not syntax.

- Just to drive this point home, that $\phi$-feature-matching does not implicate syntax even for “uninterpreted” features like grammatical gender etc. — consider:

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

(32) a. kol exad fe-yej l-o maxberet je-ya-sim
  every one that-exist DAT-3sgM notebook<F> that-3sgM.fut-put
  ot-a/*ot-o
  ba-tik
  acc-3sgF/*acc-3sgM in.the-case
  ‘Everyone who has a notebook<F>, put it.F/*it.M in your bag.’

  b. kol exad fe-yej l-o max$\text{evon}$ je-ya-sim
  every one that-exist DAT-3sgM calculator<M> that-3sgM.fut-put
  ot-a/*ot-a
  ba-tik
  acc-3sgM/*acc-3sgF in.the-case
  ‘Everyone who has a calculator<M>, put it.M/*it.F in your bag.’

- these are instances of Donkey Anaphora; that is:
  - on the relevant reading, the underlined expressions covary
  - and that is 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 relative-clause Complex NP island, inside a Subject Island
this is proof positive that \( \varphi \)-features such as those borne by pluralia tantum, or grammatical gender on inanimates, can be transmitted by a non-syntactic mechanism

- since a syntax that can relate two expressions in the absence of any c-command or locality holding between the two is no syntax at all
- again, whether this dissolves into something more general —
  - e.g. pronouns as definite descriptions (Elbourne 2013),
  - or pronouns as the residue of NP ellipsis (Merchant 2014);
- or needs to be handled in terms of the interpretation function itself, as in the sketch above;
- it doesn’t matter for our present interests
  - all is crucial is the observation that syntax is not what’s responsible.

This echoes certain aspects of Dowty & Jacobson 1988; except that they’re patently wrong about \( \varphi \)-agreement proper —

- \( \varphi \)-agreement is sensitive to case (Bobaljik 2008, a.a.o.); it is subject to a kind of minimalty that can be *ameliorated* by A-movement (see exx. (20–21), repeated here); etc. etc.,

(20) \[ \text{[Einverjum stúdent]}_1 \text{ tölvurnar ljótar} \]
\[ \text{some student.DAT finds.PL computers.the.NOM ugly} \]
\[ \text{‘Some student finds the computers ugly.’} \]

(21) \[ \text{það finnst/*finnst [einverjum stúdent] tölvurnar ljótar.} \]
\[ \text{EXPL finds.SG/*find.PL some student.DAT computers.the.NOM ugly} \]
\[ \text{‘Some student finds the computers ugly.’} \]

- properties which binding, crucially, doesn’t share
  - binding doesn’t care about case
    - in fact, the ability of Icelandic dative subjects to bind subject-oriented anaphors is one of the most striking pieces of evidence that they are indeed subjects (see Zaenen et al. 1985, among many others)
  - nor can A-movement of an intervener ameliorate minimalty violations in binding:

\[
\begin{array}{c}
\text{a. John seems to Mary to like her/*herself.} \\
\text{b. John seems to her/\#to like Mary.}
\end{array}
\]

\( \bullet \) A-moving the downstairs subject out of the way does not allow the experiencer in (33a) to bind the downstairs reflexive

\( \bullet \) this, even though the experiencer is generally able to bind into the infinitive, as the disjoint-reference effect in (33b) attests

\( \bullet \) I think it’s fair to say that no semantic account could capture the kind of minimalty effects associated with \( \varphi \)-agreement — therefore, there’s no hope in going “full Dowty & Jacobson”

7. Conclusion

- We have seen that the Anaphor Agreement Effect (AAE) does not speak in favor of a reduction of anaphoric binding to \( \varphi \)-agreement
  - contrary to previous claims in the literature
- And, in fact, militates quite strongly against such a reduction

- We have seen that timing-based and encapsulation-based theories are outright incompatible with a reduction of anaphoric binding to \( \varphi \)-agreement
- And we have seen that a timing-based approach doesn’t work even if we assume that anaphoric binding and \( \varphi \)-agreement are separate

- Having ruled out a reductionist approach, I presented a non-reductionist, encapsulation-based account of the AAE
  - which seems to converge with some recent findings from morphology (Middleton 2018)

- Finally, I discussed the source of \( \varphi \)-feature-matching between anaphors and their binders
  - suggesting that it requires little more than what is independently necessary to negotiate matching \( \varphi \)-features on pronouns across different utterances (and speakers)
    - as well as in the absence of c-command (in Donkey Anaphora)

\[ \text{See Preminger (2014:129–175) for a more comprehensive discussion.} \]
References


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.


ARRANT CODE: 10131