Features, Identity, and ‘Yourself’

Byron Ahn

Princeton University

1. Introduction

English reflexive nominals (a.k.a. ‘reflexive anaphors’) are morphosyntactically composed of two distinct nominals: a pronominal possessor (e.g., my, our) and a √SELF nominal (e.g., self, selves), as argued for in Postal 1966, and more recently in Ahn and Kalin 2018. The precise forms of each nominal in this complex anaphor are generally understood to be determined by the antecedent. As such, each sentence in (1) is understood to be grammatical because the antecedent’s ϕ-features (1.PL or 3.SG.F) are as similar as possible to those of both the pronominal possessor and the √SELF nominal, unlike the sentences in (2).

1

(1) a. We will not embarrass our selves.
   1.PL  1.PL  PL
   b. She is expressing her self.
      3.SG.F  3.SG.F  SG

(2) a. * We will not embarrass one self.
      1.PL  3.SG  SG
   b. * She is expressing your self.
      3.SG.F  2  SG

One basic intuition for patterns like (1)–(2) is that, because the reflexive anaphor is interpretively tied to the referent(s) of the antecedent, the formal features of the anaphor (and

---

*This paper builds on a collaborative work with Laura Kalin, who I must greatly thank. I would also like to thank Kirby Conrod, for discussion of their dissertation work on gender. Finally, thanks to the organizers of NELS49 and to all others for their judgments, feedback, and questions, which made this work better.

1The antecedent and pronoun always differ in some features: e.g., we and our differ in case (NOM and GEN, respectively), and we and selves in person (1st person and 3rd person, respectively).
the nominals within) should match those of the antecedent nominal. In this paper, we will focus on the $\varphi$-features of the pronominal possessor in English $X$-self anaphors:

(3) **Generalization on English Reflexive $\varphi$-Matching (Preliminary)**

In English, the local syntactic antecedent of binding and the pronominal possessor of the reflexive anaphor must maximally match in $\varphi$-features

As noted by Sundaresan (2018), this generalization has been widely assumed to be an empirically true statement and appears as an empirical fact in a number of works, including textbooks (e.g., Haegeman 1994:207, Adger 2003:94, Carnie 2013:10, and Sportiche, Koopman, and Stabler 2013:160). Some version of this generalization is also assumed in a very wide array of approaches to reflexive binding, even beyond the domain of English. (See explicit discussions in, e.g., Pollard and Sag 1992:283–4, Safir 2004:76–78, Hicks 2009:107–8, Kratzer 2009:195.)

There is formidable evidence in favor of formal $\varphi$-matching as a requirement in many languages, but it is an open question of whether this is a universal across languages. This paper focuses on two questions that arise from this framing:

**Question 1:** How valid is the preliminary generalization about English in (3)?

**Question 2:** How should we model this generalization, to the extent it is valid?

What we will find is that, in some contexts, $\varphi$-features in an English anaphor do mismatch the local antecedent of binding. Moreover, $\varphi$-mismatches are constrained such that 3.SG pronouns pattern differently than all other pronouns. These findings undercut an account of reflexive binding in English that requires syntactic $\varphi$-match with the local antecedent.

2. **Feature-Mismatch Phenomena**

Since (3) is widely assumed as a truism, several theoretical approaches to reflexivity have aimed to create accounts in which it is a logical necessity that anaphor-internal $\varphi$-features will match the antecedent of binding. (Such accounts have invoked mechanisms like AGREE, as in Heinat 2006, Reuland 2006, and Rooryck and Vanden Wyngaerd 2011.)

A consequence of these approaches is that the $\varphi$-features in anaphors always match the antecedent of binding. If any language regularly exhibits $\varphi$-mismatch, this would call into question the global validity of (3), and thus any theory that derivest (3) as a consequence.

In the remainder of this section, we review four contexts in English where the pronominal possessor and the local antecedent of binding need not $\varphi$-match. In each case, at least two possible anaphors we be provided, and the $\varphi$-features on the pronominal possessor in one of the reflexive anaphors conflict with those of the local antecedent of binding.

Let us begin with counter-identical (CID) contexts (e.g., Kauf 2017), which are a particular type of swapped-identity contexts: one individual takes on the identity of another. CID contexts can be introduced implicitly in the context, or with overt expressions such as ‘Imagine I were in so-and-so’s position,’ or ‘if I were in so-and-so’s shoes’. In these CID contexts, the anaphor can take multiple forms, while keeping the interpretation constant.
Features, Identity, and ‘Yourself’

(4) [X learned that their child, Y, has a new teacher who is very strict. X says to Y...]
   a. I would behave myself_{Y} in class from now on!
   b. I would behave yourself_{Y} in class from now on! (1.SG > 2)

(5) [On the other side of the apartment next-door is a baby that screams all night]
   If I were [the person moving in next door]...  
   a. ... I would get myself_{Y} some ear plugs.
   b. ... I would get them_{Y} some ear plugs. (1.SG > 3.PL)

This shows that φ-features for both person (π) and number (#) can mismatch between the pronoun and the antecedent of binding. We will discuss more CID data in §3, but a full discussion of CID mismatches is beyond the scope of this paper (see Ahn in prep).²

The next case of φ-mismatch involves so-called “imposters”: nominals that appear to be third person, but which can be interpreted as a speaker or hearer. (See Collins and Postal 2012 for extensive work on this topic.) Where referentially singular, the imposters show that they agree as 3rd person singular on the verb.

(6) [Spoken by a parent to a child]
   a. Mommy and Daddy need some time to them_{Y} selves.
   b. Mommy and Daddy need some time to our_{Y} selves. (3.PL > 1.PL)

(7) [Spoken by a male-identifying individual]
   a. I am a teacher who takes care of him_{Y} self.
   b. I am a teacher who takes care of myself_{Y}. (3.SG > 1.SG)

(8) [Spoken to a female judge]
   a. Does Your Honor doubt her_{Y} self?
   b. Does Your Honor doubt yourself_{Y}? (3.SG > 2.SG)

Collins and Postal (2012) argue that the φ-features in the anaphor need not match the binder (contra (3)), but can match an “ultimate antecedent”.³ This sort of data shows that π-mismatch between the pronoun and the antecedent of binding is regularly available.

Our third φ-mismatch context concerns quantified antecedents. Quantified NPs in English can range over plural entities of any person while agreeing with the verb as 3.SG.

(9) [Spoken by a woman in a group of women]
   a. Each of us can choose for herself.
   b. Each of us can choose for our_{Y} selves / them_{Y} selves. (3.SG > 1.PL / 3.PL)

²Other works considering anaphors in CID contexts (e.g., Lakoff 1996, Anand 2007, Kamholz 2012, Kauf 2017) explore how anaphors (that φ-match the local antecedent of binding) are interpreted (e.g., considering whether ‘If I were you, I’d be looking at myself’ is a looking-at-addresssee action or looking-at-speaker action). On the other hand, this work explores which morphological forms of the anaphor are available for a fixed interpretation. As far as I know, this question has not been investigated up to this point.

³According to Collins and Postal, the ultimate antecedent is represented syntactically, but is distinct from the local antecedent of binding in cases like (6)–(8).
Byron Ahn

(10) [Spoken to a group of men]
   a. At least one of you has perjured himself.
   b. At least one of you has perjured your / themselves.  \((3.\text{SG} > 2 / 3.\text{PL})\)

(11) [Spoken about a mixed-gender group]
   a. ?# Every one of my students has praised himself.
   b. Every one of my students has praised themselves.  \((3.\text{SG} > 3.\text{PL})\)

This data shows that π- and #-mismatch between the pronoun and the antecedent of binding is regularly available. In fact, many have the intuition that, outside of prescriptive norms, (11) requires a #-mismatch between the pronominal possessor and the antecedent of binding, whenever the individuals ranged over are not all male.\(^4\)

As a final case of ϕ-mismatch, similar to quantified NPs, group NPs (aka “collective nouns”) are nominals that refer to a collection of individuals. However, in some varieties, the patterns of agreement with the verb are more complex than the quantified NPs above. First consider contexts where group NP antecedents agree with the verb as singular; all investigated varieties of English allow ϕ-mismatches like (b) examples below.

(12) a. The U.N. finds itself in a difficult position.
   b. The U.N. finds themselves in a difficult position.  \((3.\text{SG} > 3.\text{PL})\)

(13) a. The football team is sorting the issue out itself.
   b. The football team is sorting the issue out themselves.  \((3.\text{SG} > 3.\text{PL})\)

In addition to allowing (12) and (13), some varieties of English (e.g., many BrE varieties) also allow the verb to agree as plural with group NPs. However, in when the verb agrees as plural, ϕ-mismatches with the anaphor are restricted (see Smith 2017).

(14) a. % The U.N. find themselves in a difficult position.
   b. * The U.N. find itself in a difficult position.  \((*3.\text{PL} > 3.\text{SG})\)

(15) a. % The football team are sorting the issue out themselves.
   b. * The football team are sorting the issue out itself.  \((*3.\text{PL} > 3.\text{SG})\)

In this section, we have seen that the ϕ-features of the pronominal possessor can regularly mismatch the antecedent with respect to π and # features. This suggests that the preliminary generalization in (3) is not a hard-and-fast rule. Instead, this sort of data will be critical for understanding the nature of how derivations with reflexive anaphors in En-

---

\(^4\)In the data below, I present themselves as the 3.PL anaphor mismatch; for some, themself is preferred as the 3.PL anaphor in these contexts. The themself anaphor does not show that them needs to have a [#.SG] specification, since the # feature of selves is independent of the # feature of the pronoun. (See fn.5.) This is more obvious in cases like ourself, which is quite well-attested (e.g., “We’re boxing ourself into a corner”, NPR All Things Considered Nov. 1 2004). Further, treating them as [#.PL] will be consistent with the finding that them does not pattern with 3.SG anaphors, regarding the availability of ϕ-mismatch (see §3.1).
5 At the same time, not just any mismatch is possible, as already discussed for certain varieties, in (14) and (15). Additionally across other English varieties, there are contexts where \( \phi \)-mismatches are impossible, which we turn to now, and which will be critical for determining how to model \( \phi \)-(mis)match in reflexive contexts.

3. Blocked Mismatches

In this section, we will find evidence that certain bound pronoun in reflexive anaphors do not support \( \phi \)-mismatch with their local antecedent, due to their syntactic \( \phi \)-values. In addition, some anaphor-internal pronouns allow \( \phi \)-mismatch with their antecedent in principle, but mismatch is constrained interpretationally, rather than syntactically.\(^6\)

3.1 \( 3.\text{SG} \) Bound Pronouns

When the pronominal in an English reflexive anaphor is \( 3.\text{SG} \), \( \phi \)-mismatch with the local antecedent is impossible. (Note: this is not \([\pi:3] \) vs. \([\pi:1/2] \), but rather \( 3.\text{SG} \) vs. others.) The natural question to ask is: what is different about \( 3.\text{SG} \) pronouns that would lead to this split in behavior with \( \phi \)-(mis)matching? Let us begin by noting that only \( 3.\text{SG} \) pronouns are necessarily specified for gender (\( \gamma \)) features in English.\(^7\) This difference is shown in the table below, which shows accusative pronouns in English, as organized by \( \phi \)-features.

<table>
<thead>
<tr>
<th></th>
<th>#:SG</th>
<th></th>
<th>#:PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>( [\pi:1] )</td>
<td>me</td>
<td></td>
<td>us</td>
</tr>
<tr>
<td>( [\pi:2] )</td>
<td>you</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( [\pi:3] )</td>
<td>him</td>
<td>( [\gamma:1,#:SG] )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>her</td>
<td>( [\gamma:F] )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>it</td>
<td>( [\gamma:INANIM] )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>one</td>
<td>( [\gamma:GENERIC] )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>them</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Broadly speaking, the contexts that support him are more narrowly defined than the contexts that support them.\(^8\) The natural question now is: why would this block \( \phi \)-feature mismatch?

\(^5\)Also, the \( \sqrt{SELF} \) and antecedent can \( \phi \)-mismatch, (i), and so can the \( \sqrt{SELF} \) and pronominal, (ii).

(i) Everyone credits them selves.\(^1\)

(ii) We all need to ask our self [a very serious question]. \( (ABC \ \text{Nightline, June 13, 1997}) \)

The latter suggests that the features of the maximal DP for an anaphor (whose head N in ourself would be the \([\pi:3, \#:SG] \) self nominal) can differ from the features of the pronoun (our in ourself, which is \([\pi:1, \#:PL] \)), a finding made in investigations of other languages as well (cf. Greek clitic doubling, Iatridou 1988, Anagnostopoulou and Everaert 1999; Selayarese agreement, Woolford 1999).

\(^6\)It could be that there is syntactic feature matching/valuation, if we change some assumptions regarding (i) how/when feature-matching syntactic operation applies, and/or (ii) the nature of \( \phi \)-features on the antecedent. Thus, to make syntactic feature matching/valuation viable in what I am calling “mismatches” in this paper, multiple sets of \( \phi \)-features would need to be present on the same antecedent (and there would need to be constraints on when those features are available). See Smith 2017 for such an analysis in the context of group NPs, but which does not obviously extend to the other cases of mismatch discussed here.

\(^7\)Unlike \( 1/2/3.\text{PL} \) pronouns, ‘gender neutral’ nominals like friend (cf. Bjorkman 2017), and even ‘gendered’ nominals like cowgirl or Kaity (cf. Ackerman 2019).

\(^8\)A reminder: they is taken as grammatically \( 3.\text{PL} \), even when referring to a single entity. See fn. 4.
The γ features on pronouns, where they exist, are licensed by an appropriate antecedent in the context, via syntactic and/or interpretive mechanisms (e.g., Ackerman 2019, Conrod 2019, and Sigurðsson to appear). The first, syntactic way to license γ on a 3.SG pronoun is to ensure that a syntactic antecedent with a γ feature (e.g., a gendered pronoun), if one exists, has the same γ-feature as the pronoun. A second, interpretive way would be to search the context for an accessible coferent/covariant antecedent, and consider whether the pronoun’s γ feature is compatible with it. Either way, a γ feature on the 3.SG pronoun requires additional derivational steps, with these steps leading to ϕ-match.

(17) Gender-Driven ϕ-Feature Matching (English)
When gender (γ) is specified on a pronoun in an English reflexive anaphor, the pronoun must not conflict with the ϕ-features of the local antecedent of binding.

The derivational steps that give rise to (17) will not be spelled out in any further detail here. Instead, for our purposes what matters is (i) that these steps only take place for γ-specified pronouns in English (i.e., 3.SG pronouns), and (ii) that they will block ϕ-features of the 3.SG pronoun from conflicting with those of the antecedent.

This can help explain why a 3.SG pronominal in an anaphor cannot π-mismatch its antecedent in a CID, as in (18), even though CIDs otherwise allow π-mismatches (cf. (5)).

(18) [On the other side of the apartment next-door is a baby that screams all night]
If I were the man moving in next door...
   a. ...I would get my self some ear plugs.
   b. #$I would get him self some ear plugs. (#1.SG > 3.SG.M)

Similarly, though contexts like ’each of us can...’ and ’we can each...’ in (19) are mutually entailing, only the latter leads to a π-mismatch when the anaphor contains a 3.SG pronoun. We predict a mismatch in that latter context to be unacceptable, and it is:

(19) [Spoken by a woman in a group of women]
   a. Each of us can choose for her self.
   b. #$We can each choose for her self. (#1.PL > 3.SG)

As a third case, we also understand the facts in (20) with ‘singular they’ and generic one:

(20) a. One / They must always behave them selves.
   b. One / #$They must always behave one self. (#3.PL > 3.SG.GENERIC)

---

9One possibility is that the γ features are ‘in-born’ at merge, and what is needed is a comparison against the antecedent’s interpretation (Cooper 1983, Heim 2008, Conrod 2019). Another possibility is that the 3.SG pronoun is deficient such that its γ features are determined by the features of the context or the antecedent nominal (both of which are represented syntactic ally; Sigurðsson to appear). I will not take up this issue here. However, it should be noted that the former would mean all anaphors in English can come fully ϕ-specified, while the latter would mean anaphors in English exhibit a gender-based split as to whether they can.
Though *them* and *one* can corefer, as in (20a), it is impossible when the 3.SG *one* occurs as the pronominal in an anaphor, as in (20b). This is because *one*, unlike *them*, comes with the specified gender feature ([\(\gamma:\text{GENERIC}\)]), blocking mismatch with the antecedent.

### 3.2 Pronominal Appropriateness

The constraint in (17) can explain the asymmetry between anaphors containing a 3.SG pronominal possessor versus all anaphors, whereby the former cannot \(\varphi\)-mismatch the antecedent. However, not all blocked \(\varphi\)-mismatches reduce to (17).

To explore this, let us turn to discourse-anaphoric pronouns, as in (21). Here, pronoun choices lead to (in)felicity because of how pronouns are interpreted in context (see, e.g., Heim and Kratzer 1998, Sauerland 2003, Ackerman 2019):

(21)  
\[
\begin{align*}
    \text{a. } \sqrt{\text{Each of us}}_9 \text{ said that } \text{we}_9 \text{ are going.} \\
    \text{b. } \# \text{I}_4 \text{ said that } \text{you}_4 \text{ are going.} \\
    \text{c. } \# \text{The children}_7 \text{ said that } \text{she}_7 \text{ is going.}
\end{align*}
\]

Similarly, the pronominal in an anaphor is subject to the same sorts of constraints concerning interpretation in context. This is summarized in (22), predicting the patterns in (23).

(22)  
**Contextually-Determined \(\varphi\)-Feature Appropriateness (English)**

The pronoun in an English reflexive anaphor must be able to be appropriately construed as referring to / varying with its antecedent, at a conceptual level.

(23)  
\[
\begin{align*}
    \text{a. } \sqrt{\text{Each of us}}_9 \text{ convinced } \text{our}_9 \text{ selves to go.} \\
    \text{b. } \# \text{I}_4 \text{ convinced } \text{your}_4 \text{ self to go.} \\
    \text{c. } \# \text{The children}_7 \text{ convinced } \text{her}_7 \text{ self to go.}
\end{align*}
\]

The main idea is that the \(\varphi\)-features of the pronominal possessor in an anaphor are subject to the same sorts of constraints of contextual appropriateness as any other pronoun.

We now make a prediction: where there is variation in speakers’ views of appropriateness in using particular discourse-anaphoric pronouns, the same variation should be attested in reflexive contexts. Consider what happens with ‘they’ in contexts like (24):

(24)  
\[
\text{% Every mother thinks their child is special} \quad \text{ (USA Today, Apr 1997)}
\]

While some speakers allow these examples where the antecedent of a *they* pronoun is a quantified expression ranging over individuals of the same gender (here female), some speakers disprefer *their* in these contexts (instead preferring a gendered pronoun). In this way, there is variation in the types of contexts in which *they* is allowed in (cf. Bjorkman

---

2017, Ackerman 2019, Conrod 2019). The prediction here, based on (22), is that if and only if a speaker allows examples like (24), the will also accept examples like (25):

(25) %Every mother tells them selves that they have a special child

Similar predictions are made with regard to the usage of they to refer to specific individuals (e.g., non-binary individuals, or individuals whose gender the speaker does not disclose).

Additionally, this approach will allow us to capture the contrast between (a) and (b) in (26) on the basis of interpretation in context:

(26) [Spoken by a woman in a group of women]
   a. Each of us can choose for herself.
   b. #Each of us can choose for himself. (\#3.SG > 3.SG.M)

If nominals like each of us have gender \(\varphi\)-features, (26b) could be explained by (17); but (26b) is also interpretively inappropriate even if such nominals lack grammatical gender. That is, it is the context that precludes himself, since all the relevant individuals are women.

Lastly, compare the good \(\varphi\)-mismatch in (5) with the bad one in (27), where all that is manipulated is whether or not a discourse referent has been introduced.

(27) [On the other side of the apartment next-door is a baby that screams all night]
   If I were moving in next door...
   a. ... I would get myself some ear plugs.
   b. #... I would get them selves some ear plugs.

The infelicity of (27b) highlights the role of contextual appropriateness: when the context doesn’t introduce the appropriate type of antecedent (or indeed no antecedent at all, as in (27)), then mismatch is inappropriate.

Given that \(\varphi\)-features play a role in determining the which pronouns are interpretively felicitous in context, it must be that the \(\varphi\)-features in reflexive anaphors are visible at LF, and are not simple morphological decorations.

4. Conclusions

Having investigated the empirical landscape of English more carefully, we can return to our original questions from §1. Addressing Question 1, we can deem the preliminary generalization in (3) to be too coarse-grained. It does not predict any \(\varphi\)-mismatches between the pronominal possessor of a reflexive anaphor and its local antecedent of binding, counter to fact (cf. §2). As for Question 2, addressing this required looking at where \(\varphi\)-mismatch data is (im)possible (cf. §3), and we uncovered the following two generalizations.

\[\text{[There is variation in how a reflexive containing a ‘them’ reflexive anaphor surfaces: themselves, themself (see fn. (4)), and theirselves are all attested.]}\]

\[\text{[This raises important questions on how to get appropriate interpretations in so-called ‘fake indexical’ contexts (cf. Kratzer 2009). This is outside of the scope this paper, but must ultimately be addressed.]}\]
Features, Identity, and ‘Yourself’

(17) Gender-Driven φ-Feature Matching (English)
When gender (\(\gamma\)) is specified on a pronoun in an English reflexive anaphor, the pronoun must not conflict with the \(\phi\)-features of the local antecedent of binding.

(22) Contextually-Determined φ-Feature Appropriateness (English)
The pronoun in an English reflexive anaphor must be able to be appropriately construed as referring to / varying with its antecedent, at a conceptual level.

As described in (17), some pronouns in reflexive anaphors appear to always \(\phi\)-match the local antecedent in English (3.SG ones). As a result, English reflexive binding exhibits a grammatical split (3.SG vs. other) with respect to whether \(\phi\)-mismatch is possible, due to how gender features (only found on 3.SG pronouns) are licensed and interpreted in English.

Because of this split, if attention is paid only to contexts with a 3.SG pronoun in the reflexive anaphor, \(\phi\)-matching would appear to be obligatory in English, and (3) would be a good explanandum for a theory of binding. However, as there are contexts that support \(\phi\)-mismatch, and so (3) is insufficient. Our investigation implies that there may be two grammatical routes for determining the appropriate \(\phi\)-features in a reflexive anaphor in English: one that allows \(\phi\)-mismatch and the other that blocks it. I leave it as an open question as to how to model each of these grammatical routes. Instead of trying to model this in this work, I suggest that the statements in (17) and (22) supersede the statement in (3), as they are better candidates for the explananda for a theory of binding in English.

References

Ackerman, Lauren. 2019. Syntactic and cognitive issues in investigating gendered coreference. lingBuzz/004064.
Ahn, Byron. In Preparation. If I were you, I would be able to bind yourself. ms., Princeton.
Byron Ahn


Sigurðsson, Halldór Ármann. To appear. Gender at the edge. LI.


Byron Ahn
bta@princeton.edu