Plurality and Binding within the English *th...sel{f/ves}* Paradigm

Dennis Ryan Storoshenko

University of Calgary

June 3, 2019
English *th*-self forms are...complicated.

With three possible “case” pronouns and two possible “number” variants, there are six possible forms:

<table>
<thead>
<tr>
<th>Case</th>
<th>Singular</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nom</td>
<td><em>theyself</em></td>
<td><em>themselves</em></td>
</tr>
<tr>
<td>Acc</td>
<td><em>themself</em></td>
<td><em>themselves</em></td>
</tr>
<tr>
<td>Gen</td>
<td><em>themselves</em></td>
<td><em>themselves</em></td>
</tr>
</tbody>
</table>

Disclaimer

This presentation is going to ignore -*selfs*, though it exists.
These are not all interchangeable

Neither variable is strictly associated with any particular type of antecedent

Different grammars appear to carve up the functional load borne (prescriptively) by *themselves* in different ways (with *they* forms being outliers)
Outline

1. Today's Puzzle
2. Prior Discussion
3. New Twitter Corpus
4. Results
5. Discussion
Singular *they*

- This is partly wrapped up in a discussion of what exactly the *th-* pronouns denote.
- Social changes aside, singular uses of the *th-* pronouns abound in contexts where definiteness is uncertain.

**Example (Bjorkman (2017))**

They had the wrong number.
(To someone else in the room after quickly ending a phone call)

**Example (Bodine (1975))**

Somebody left their sweater.

- These can both be felicitous even when the speaker has a reasonably good idea about the gender of the referent.
Bjorkman goes on to report that for some speakers, use of a *th*-form is obligatory in bound variable contexts.

The account is that the *th*-pronouns are less featurally-rich than gendered pronouns, and thus emerge as defaults.

This dovetails into the Déchaine and Wiltschko (2002a) account that increased featural specification makes a pronoun less likely to function as a bound variable.

**Question**

How does this relate to reflexive forms?
In Déchaine and Wiltschko (2002b), this is extended to the English -self paradigm

1st and 2nd person: full DPs

- 1st and 2nd person: full DPs
  \[
  \text{DP} \quad \text{DP}
  \]
  \[
  \text{D} \quad \Phi_P
  \]
  \[
  \text{my} \quad \Phi
  \]
  \[
  \text{NP} \quad \text{NP}
  \]

3rd person: ΦP only

- 3rd person: ΦP only
  \[
  \text{ΦP}
  \]
  \[
  \Phi
  \]
  \[
  \text{him} \quad \text{NP}
  \]
  \[
  \text{self}
  \]

The possessive morphology signals an extra layer of definiteness not present in 3rd person

- DPs undergo assigned co-reference, ΦPs are semantically bound
Implicit Predictions

- This makes a clear prediction for a form like *hisself*: assigned co-reference, not variable binding
- Compare to Cheshire et al. (1993) where there is just described as dialectal paradigm levelling
- *herself* and *itself* are a bit of a wash for testing this
- And, we would expect the same thing for *themselves* vs *theirseives*
In results presented at CLA 2013 and LSA 2015, with two different Twitter data sets, the story was the same:

- \textit{hisself} behaves exactly as predicted
- \textit{themselves} stubbornly allows bound variable antecedents

Nagging Question

Why????
One pattern that jumped out was that most of the “unexpected”
data came from the Commonwealth, or from parts of the US where
this variation is uncommon

Maybe it is paradigm levelling for some people, or a snowclone (a
template being misapplied without knowledge of the underlying
grammar)

This is not a simple binary choice; English in North America gives us
six possible forms

But, the data collection was tiresome, so I left it alone. Even though
some isolated speakers had exactly the predicted pattern.
Outline

1. Today's Puzzle
2. Prior Discussion
3. New Twitter Corpus
4. Results
5. Discussion
Easier-to-use tools for scraping twitter now exist. In this case, twitteR (Gentry, 2015), which, via R (R Core Team, 2017) makes this all a lot faster.

As a first step, collect the most recent 5000 tokens of each form (June 2018).

The search window is roughly one week, if the frequency is lower than the value selected, you just get all the possible data.

No regional filters used.

Eliminating retweets and obvious robots, a manageable corpus can be collected.
Corpus Coding

- Number of the antecedent (Sg/Pl)
- Animacy of the antecedent (Animate/Inanimate)

Function of the form:
- Argument
- Adnominal Modifier
- Adverbial Modifier
- By Phrase
- Exempt
- Metalinguistic/Robot

Type of Antecedent:
- Definite DP
- Indefinite DP
- Bound Element (Control and Relative Clauses)
- Weak Quantifiers (Numerals, *most* N, *a lot of* N, etc...)
- Strong Quantifiers (*every* N, *some* N, *wh*...)

Dennis Ryan Storoshenko (UofC)
Outline

1. Today's Puzzle
2. Prior Discussion
3. New Twitter Corpus
4. Results
5. Discussion
<table>
<thead>
<tr>
<th></th>
<th>Argument</th>
<th>Adverbial</th>
<th>Adnominal</th>
<th>By Phrase</th>
</tr>
</thead>
<tbody>
<tr>
<td>theyself</td>
<td>90.70%</td>
<td>1.69%</td>
<td>0.28%</td>
<td>5.29%</td>
</tr>
<tr>
<td>themselves</td>
<td>80.82%</td>
<td>6.85%</td>
<td>8.81%</td>
<td>1.96%</td>
</tr>
<tr>
<td>theirselves</td>
<td>88.51%</td>
<td>4.04%</td>
<td>1.86%</td>
<td>4.97%</td>
</tr>
<tr>
<td>theyselves</td>
<td>91.04%</td>
<td>0%</td>
<td>1.49%</td>
<td>7.46%</td>
</tr>
<tr>
<td>themselvess</td>
<td>89.18%</td>
<td>4.58%</td>
<td>4.02%</td>
<td>1.80%</td>
</tr>
<tr>
<td>theirselves</td>
<td>91.82%</td>
<td>3.48%</td>
<td>1.31%</td>
<td>2.18%</td>
</tr>
</tbody>
</table>

- *them* forms are the most versatile overall...
- ...but appear least frequently in a *by*-phrase.
- Same basic trends in *-self* vs *-elves*
<table>
<thead>
<tr>
<th></th>
<th>Singular</th>
</tr>
</thead>
<tbody>
<tr>
<td>theyself</td>
<td>23.10%</td>
</tr>
<tr>
<td>themself</td>
<td>57.73%</td>
</tr>
<tr>
<td>theirselves</td>
<td>36.34%</td>
</tr>
<tr>
<td>theyselves</td>
<td>2.99%</td>
</tr>
<tr>
<td>themselves</td>
<td>13.04%</td>
</tr>
<tr>
<td>theirselves</td>
<td>12.10%</td>
</tr>
</tbody>
</table>

- Definitely not the case that *th-* denotes plurality
- The plural suffix is only relatively plural, not absolute
- In both cases, the *they* forms are markedly different
- *themselves* and *themselves* are fairly close though
## Bound Variable Usage

<table>
<thead>
<tr>
<th>Bound Variable Usage</th>
<th>StrongQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>themselves</td>
<td>1.49%</td>
</tr>
<tr>
<td>themselves</td>
<td>5.55%</td>
</tr>
<tr>
<td>themselves</td>
<td>4.96%</td>
</tr>
</tbody>
</table>

- Speakers needing a singular bound variable are adopting *th-self* forms

**Example**

- Everybody expose *themselves* in due time.

- Everyone ... is putting *themself* in the right side of the playing-field.

- I just seen somebody hawkin loogies on *themselves*.

- *themselves* is the outlier
Recall that bound variable contexts facilitate singular *they*

Example

...he’s just just someone calling *themselves* a vegan to get a rise out of us.

For the *they* forms, singular use seems more dependent on bound variable contexts

Not much difference here between *them* and *their*
Outline

1. Today's Puzzle
2. Prior Discussion
3. New Twitter Corpus
4. Results
5. Discussion
Bjorkman’s discussion of simple pronouns is reflected here with reflexive forms: there’s a relationship between variable binding and singularity.

The difference is that this is now reflected in two ways:
1. The proportion of singular interpretations arising from binding
2. The proportion of -self forms emerging as singular compared to -sel/ves

That Nagging Question
What about the DP/ΦP stuff?
The Twist Ending

- *themselves* is not doing what was predicted, but *themselves* is
- There is one other dimension on which these two are alike, and distinct from *themselves*

<table>
<thead>
<tr>
<th></th>
<th>% Animate</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>they</em></td>
<td>97.01%</td>
</tr>
<tr>
<td><em>them</em></td>
<td>94.31%</td>
</tr>
<tr>
<td><em>thir</em></td>
<td>98.87%</td>
</tr>
</tbody>
</table>

- There is a significant animacy contrast between *themselves* and *themselves*
- If a rejection of inanimates is a reflection of possession, this makes some sense for the genitive form
My proposal is that the *they* forms are historically *their*.

Of all three pronoun forms, the *they* cases are most clearly sociolinguistically coherent; they belong to AAE.

This happens to be a community that should also have *themselves* (Green, 2002).

This variety is known to have a post-vocalic de-rhoticization process (Pollock and Berni, 1996).

This variety has *hisself* but not *heself*.

These speakers are preserving a DP structure for *themselves*. 
- *theirselfs* is widely spread around the world
- This could indeed be paradigm levelling when considering a global dataset
- Both *themselves* and *theirselfs* seem to be retaining more of a sense of plurality...are users considering them to be alternatives to *themselves*?
- *themselves* appears to be majority singular usage, and is the only one with a sense of conscious use as a non-binary form
- A more regionally controlled dataset (collected but not yet analyzed) should shed more light on this
- I still have no idea what to do with the functional differences

Thanks!
Thanks to the audience at the LSA in 2019 who saw a partial presentation of this data, colleagues for presentation feedback, and all the people whose tweets are caught up in my dataset. Some were worth a follow.


This is unique tokens in a pool of 5000 (suitable measure of salt)

For the -self forms:

\[ \text{themselves} > \text{themselves} > \text{themselves} \]

For the -selves forms:

\[ \text{themselves} > \text{themselves} > \text{themselves} \]

-themselves so infrequent that stats get problematic

-themselves is most frequently discussed metalinguistically

-themselves has the highest frequency of robot usage