Only Cool People Tweet Theirselves: Variation in the English Reflexive Paradigm

Dennis Ryan Storoshenko

Department of Linguistics
Yale University

2013 Meeting of the CLA
June 1-3, 2013

Yale
Variation in the Reflexive System

- Some representative examples from COCA:

  **Example**

  He presented *hisself* to me in a very unprofessional manner.
  ...the American people need to wean *theirselves* off student loans.

- I will refer to these as “non-accusative variants” (NAVs)

**The Big Question**

What do we know about these variants?
In this paper, I address three research questions:

1. What is the social/geographic distribution of the variants? *Who uses them, and where?*

2. Are there any conditions on the usage of the variants? *Do the variants have the same distribution as their accusative equivalents?*

3. What are the theoretical implications of the observed usage? *Do we gain any insights into the nature of pronominal systems from this undertaking?*
Déchaine and Wiltschko (2002a, 2002b) describe three different internal structures for pronouns (including reflexives):

For each different syntactic structure, reflexivity is derived via distinct semantics:

<table>
<thead>
<tr>
<th>Table: Reflexive Typology</th>
</tr>
</thead>
<tbody>
<tr>
<td>D-Reflexive</td>
</tr>
<tr>
<td>Semantics</td>
</tr>
</tbody>
</table>
D&W note that the uneven Standard English reflexive paradigm suggests different categorial statuses:
- First and Second Person: possessive syntax $\rightarrow$ DP
- Third Person: accusative syntax $\rightarrow$ $\varphi$P

**Corollary**

Having possessive syntax, *hisself* and *theirselves* should be DPs. Reflexivity derives therefore from assigned co-reference rather than semantic binding.

**Prediction**

NAVs should resist quantificational binding.
This project has also been a test-bed for exploring the usefulness of twitter as a tool for “remote fieldwork”.
1 The Issue
2 Distribution
3 Conditions on Use
4 Theoretical Implications
Outline

1. The Issue
2. Distribution
3. Conditions on Use
4. Theoretical Implications
hisself and theirselves are documented in American English as a feature of the southeast back to the 19th Century (Cassidy and Houston Hall, 1991).

Figure: Hisself distribution in DARE (Cassidy and Houston Hall, 1991)
The same forms appear in the dictionary of Smoky Mountain English (Montgomery and Hall, 2004).

These two sources are dictionary resources only, citing NAVs only as alternatives to their accusative counterparts.

All but one of the dictionary sentences has a referential antecedent.


NAVs are also present in regional dialects across the UK (Cheshire et al., 1993).
Collecting a Twitter Corpus

- Collect 24 hours of instances of *hisself* and *theirselves*.
- A first-pass inspection was carried out to remove re-tweets and obvious robots.
- For each tweet, attempted to ascertain the location of the tweeter.
  - Geotagged tweets
  - Location in profile
  - Inference from photos/text
- Only 80% of the accounts could later be verified.

**Table: Raw vs. Verified Tweets**

<table>
<thead>
<tr>
<th></th>
<th>Raw</th>
<th>Verified</th>
</tr>
</thead>
<tbody>
<tr>
<td>hisself</td>
<td>742</td>
<td>606</td>
</tr>
<tr>
<td>theirselves</td>
<td>505</td>
<td>404</td>
</tr>
</tbody>
</table>
US is by far the leader, with the UK coming second, and Canada a distant third.

- US: 872
- UK: 54
- Canada: 11

Some “clusters” in the Philippines, Malaysia, and the Netherlands.

Only 8 could not be narrowed down to at least a country

The remaining 65 are scattered around the world, mostly in former British holdings.
UK Results as expected based on Cheshire et al.
- Primarily younger speakers (potential sample bias).
- Mostly caucasian, some Middle Eastern individuals in the London area.
- Relatively few in major centres though.

Canadian results relatively sparse
- Spread from BC to the Maritimes
- 7 out of 11 are in Ontario, the only cities with multiple tweets are Toronto and Windsor.
- Mostly caucasian speakers, but otherwise ethnically diverse.
Looking at the US

- 40 out of 50 states represented, with varying degrees of frequency.
- Densest concentration is in the southeast, as described in DARE
- With the exception of California, there is a very clear east/west divide
- Ethnic division is stark: 80% African American, 17% Caucasian, 3% Other. If anything, Green underplays this as a feature of AAE.
- Little geographic overlap between ethnic groups, except in California, New York, New Jersey, and 2 states with much lower numbers (Kentucky and Oklahoma)
Tweet Density by State

Figure: Verified Tweet Counts by State
Figure: Ethnic Diversity of Tweeters
Figure: Tweets by Ethnicity (Green = African American, Blue = Caucasian) in PA

- Pittsburgh and Philadelphia each conceal one blue pin.
Overall, the twitter corpus identifies three distinct usage communities:

1. Young speakers in the UK
2. Rural caucasian speakers in the US
3. African American speakers in the US who are the only users in major urban areas. (with earlier noted exceptions)
One Interesting Example from World English

Figure: Tweet from an ESL School in India

Question

Why the possessive variant in this context?
1. The Issue

2. Distribution

3. Conditions on Use

4. Theoretical Implications
Historical Examples

- From DARE:

  **Example**
  
  He has hurt hisself mighty bad.
  That fellow that called hisself Mayor Gumby.
  I was aimin’ for him to take ’em hisself.
  Somebody got hisself a good meal.

- From Smoky Mountain English

  **Example**
  
  Ole Bullin hisself looked down at me...
  ...he was join’ the have one if he had to go on by hisself.
  “Since Doc Hornsby killed hisself,”
  ...he could jump off the bridge and kill hisself one night.

- All examples have clear referential antecedents.
### Checking Against COCA

#### Table: Antecedents for *hisself* in COCA 1990-2010

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pronoun</td>
<td>114</td>
<td>59.3%</td>
</tr>
<tr>
<td>Proper Name</td>
<td>36</td>
<td>18.8%</td>
</tr>
<tr>
<td>Definite DP</td>
<td>28</td>
<td>14.6%</td>
</tr>
<tr>
<td>Indefinite DP</td>
<td>8</td>
<td>4.2%</td>
</tr>
<tr>
<td>Relativized Subject</td>
<td>5</td>
<td>2.6%</td>
</tr>
<tr>
<td>Numeral</td>
<td>1</td>
<td>0.52%</td>
</tr>
</tbody>
</table>

- No instances of the schema `every [nn*] [v*] hisself`
Table: Antecedents for *theirselves* in COCA 1990-2010

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pronoun</td>
<td>18</td>
<td>46.2%</td>
</tr>
<tr>
<td>Proper Name</td>
<td>1</td>
<td>2.6%</td>
</tr>
<tr>
<td>Definite DP</td>
<td>6</td>
<td>15.4%</td>
</tr>
<tr>
<td>Indefinite DP</td>
<td>1</td>
<td>2.6%</td>
</tr>
<tr>
<td>Relativized Subject</td>
<td>1</td>
<td>2.6%</td>
</tr>
<tr>
<td>Numeral</td>
<td>1</td>
<td>2.6%</td>
</tr>
<tr>
<td>Bare Plural</td>
<td>4</td>
<td>10.2%</td>
</tr>
<tr>
<td>“People”</td>
<td>2</td>
<td>5.1%</td>
</tr>
<tr>
<td>Coordinated Phrase</td>
<td>2</td>
<td>5.1%</td>
</tr>
<tr>
<td>Anybody</td>
<td>1</td>
<td>2.6%</td>
</tr>
<tr>
<td>Exempt</td>
<td>2</td>
<td>5.1%</td>
</tr>
</tbody>
</table>

No instances of the schema

all [nn*] [v*] theirselves
The NAV is lost with a quantified antecedent.
Looking at the Twitter Corpus

- In the twitter corpus, the only instances of *hisself* with a strong quantifier antecedent are in one fixed expression:
- *every man for hisself* (3 instances: 2 US, 1 Middle Eastern)
- The situation for *theirselves* is murkier:

**Example**

Twitter’s just proof that everyone talks to theirselves. Every girl in my class introduced theirselves & said their fav artist is taylor swift #typical
It bugs me how some people nickname theirselves.

- In all, 54 out of the 404 verified instances of *theirselves* have some sort of quantificational antecedent.
- *Hiss*elf strongly resists quantificational antecedents.
- *Theirse*lves is...messy.

**Checking The Original Prediction**

Halfway confirmed; *hiss*elf behaves according to the D&W prediction, but *theirse*lves does not.

**Next Step**

Explaining the difference.
Outline

1. The Issue
2. Distribution
3. Conditions on Use
4. Theoretical Implications
It has been suggested that the NAVs are simple paradigm levelling.

Perhaps the NAVs are an innovation, showing a change underway to a level reflexive paradigm.

We could imagine a DM--esque account where $\varphi$-features are spelled out identically in [Gen] and [Refl] contexts.

There are two problems with this:

1. Doesn’t explain why *hisself* and *theirselves* are different.
2. Doesn’t explain why the dictionaries and COCA match the D&W predictions for *theirselves* (only the twitter data is problematic).

Next Step

Look closer at that twitter data.
54/404 *themselves* had a quantificational antecedent.

The country of origin for these tweets does not match the overall corpus:

- USA: 36 (12% of total US *themselves*)
- UK: 9 (20% of total UK *themselves*)
- Canada: 2 (20% of total Canada *themselves*)
- Other Countries: 7

There is a disproportionate representation of the UK in these unexpected tweets.
A Tale of Two Birminghams

- Birmingham, UK
  - In the past year, 2 *hisself* tweets, 14 *themselves* tweets
  - Two quantificational antecedents for *themselves* (*someone* and *everyone*)

- Birmingham, AL
  - In the past week, 8 *hisself* tweets, 6 *themselves* tweets
  - One quantificational antecedent for *themselves* (*anybody*)

Usage in the two countries is clearly different

- In the same period, Birmingham, AL has two *thirself* and one *theyself*

- Birmingham, UK has three *thirself*, no *theyself*
Thinking back to D&W, *theyself* is a good $\varphi$P, if we can show *they* being bound quantificationally:

**Example**

Most of the presenters thought they were doing well.

**Explanation**

The different countries exploit the available pronominal inventory differently.

**Lingering Question**

We still don't know why *themselves* is showing up with a quantified antecedent at all.
Looking at the US Examples

The 36 “unexpected” tweets are from a mix of Caucasian and African American tweeters.

Most of the caucasian speakers are either from urban areas (Cleveland, Pittsburgh, New Jersey) or peripheral states (Minnesota, Alaska)

No examples at all from Georgia or Louisiana, one each from Mississippi and Alabama

The Takeaway

For both Caucasian and African American tweeters, most of these examples come from areas outside their core constituency.

Potential Explanation

These examples could be the result of either imperfect learning, or a snowclone effect.
Baltin et al. (2013) note that English pronouns can function as either D or \( \varphi \) heads.

Example

Every football team hopes that it will win the Grey Cup.
Every farmer who owns a donkey beats it.

With independent evidence that *their* in isolation can be a \( \varphi \) head, speakers may construe *themselves* as \( \varphi P \), not DP.

Example

Each student submitted their homework on time.

*themselves* may be either \( \varphi P \) or DP.
So why is *hisself* so well-behaved, when we can just as easily show that *his* can also be a $\varphi$ head?

- While *hisself* can be ambiguous between $\varphi$P and DP, *himself* is only $\varphi$P. In a bound variable context, it is the most specified form, and we can appeal to blocking.

- [3PL] pronouns in general are subject to varying conditions of use (Larry Horn, p.c.)

- This, combined with the fact our speakers are from outside established communities may make them more prone to error.

- A lack of clear data for *themselves* in peripheral communities may be feeding a situation in which the distinction is not sharp enough for the blocking to be learned.

- In the core communities, where usage is more frequent, the blocking is acquired.
Conclusions and Future Work

- There is clear evidence that *hisself* and *theirselves* are used differently in the UK and the US.
- Most likely, this is an interaction with a richer reflexive paradigm in the US.
- The D/ϕ duality of *their* (and *his*) extends into the reflexive construction.
- While *hisself* can be a DP or ϕP reflexive, it can be argued that *himself* blocks *hisself* in bound variable contexts. Insufficient data prevents this from happening with *theirselves*.
- More systematic comparisons of the Birmingham type will shore up the inter and intra-national distinctions.
- More efforts need to be made to elicit judgments from tweeters (active rather than passive data collection).
Thanks to all the members of the Yale Grammatical Diversity Project for their support and suggestions during this work, and to SFU’s XSyn Lab for listening to an earlier version of this presentation. Thanks also to all my unsuspecting data sources.


