What motivates verbal agreement variation with collective-headed subjects? Evidence from synchrony and diachrony

Yolanda Fernández Pena
yolanda.fernandez@uvigo.es
University of Vigo (Spain)
5 May 2016
Outline

1. Introduction
2. Aims
3. Corpus-based study
   1. Methodology
   2. Data analysis
      1. Regional variation
      2. Syntactic complexity
      3. (C)overt morphology
      4. Semantics
      5. Idiomatisation
4. Conclusions
5. References
1. Introduction

**Agreement and Collective Nouns**

Collective noun:

“morphologically singular nouns designating a group of (in)animates”

(Dekeyser 1975: 35fn.1)

family, police, committee...

number, group, majority...

(1a) The *crowd*$_{SG}$ here *is*$_{SG}$ really thick despite the weather.

(1b) the *crowd*$_{SG}$ *are*$_{PL}$ on their feet, roaring and waving their arms.

(1c) *The crowd*$_{SG}$ *of cockneys*$_{PL}$ *were*$_{PL}$ singing along* [BNC: BPA 62]

Morphologically motivated overrides/mismatches
2. Aims

✓ Explore variation of **verbal agreement** in number with **collective nouns** taking of-PPs

(2) A large $\text{group}_{\text{SG}}$ of $\text{people}_{\text{SG}}$ was$_{\text{SG}}$ standing just beyond the wrought-iron gate.

(3) A $\text{group}_{\text{SG}}$ of $\text{parents}_{\text{PL}}$ were$_{\text{PL}}$ standing in the corner

✓ Focus:

✓ Present-Day British and American English
✓ syntactic/structural/formal explanation for subject-verb agreement variation

✓ Further issues:

✓ “Inner Englishes”
✓ Semantic and lexical determining factors
✓ Late Modern English
3. Corpus-based study

• 3.1. Methodology

• 3.2. Data analysis

  1. Regional variation
  2. Syntactic complexity
  3. (C)overt morphology
  4. Semantics
  5. Idiomatisation
3.1. Methodology

**DATA RETRIEVAL:**

‘$N_{\text{COLL}}$-of-$N_{\text{PL}}$’ subject + verb inflected for number

- **$N_{\text{COLL}}$**
  - 23 *singular* collective nouns

  - *band*  
  - *crowd*  
  - *majority*  
  - *series*  
  - *batch*  
  - *flock*  
  - *minority*  
  - *set*  
  - *bunch*  
  - *gang*  
  - *number*  
  - *shoal*  
  - *class*  
  - *group*  
  - *pack*  
  - *swarm*  
  - *clump*  
  - *herd*  
  - *party*  
  - *troop*  
  - *couple*  
  - *host*  
  - *rash*  

(Biber et al.’s 1999: 249 ‘quantifying collectives’
Huddleston and Pullum et al.’s 2002: 503 ‘number-transparent nouns’

---

---
3.1. Methodology

**DATA RETRIEVAL:**

‘$N_{\text{COLL}}$-of-$N_{\text{PL}}$’ subject + verb inflected for number

- $N_{\text{COLL}}$
  - 23 singular collective nouns

- $N_{\text{PL}}$
  - oblique noun

(i) $NN_2$ = overtly-marked plural N (boys, things...)

A group of $\text{boys}_{NN_2} / \text{girls}_{NN_2}$

A bunch of $\text{things}_{NN_2} / \text{cases}_{NN_2}$

(ii) the non-overtly-marked plural N people

A group of $\text{people}$
3.1. Methodology

**CORPORA:**

- **British National Corpus (BNC)**
  100 million words (1970s-1993)

- **Corpus of Contemporary American English (COCA)**
  over 450 million words (1990s-2012)

- **Global Web-based English Corpus (GloWbE)**
  1,9 billion words (2012-13)
  20 varieties of English

- **Corpus of Historical American English (COHA)**
  400 million words
  1810-2009
3.1. Methodology

PDE BrE & AmE (adapted from Fernández-Pena 2015)

<table>
<thead>
<tr>
<th></th>
<th>Abs. Freq.</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SG Verb</td>
<td>PL Verb</td>
</tr>
<tr>
<td>With of-PP</td>
<td>1,593</td>
<td>3,014</td>
</tr>
<tr>
<td>Without of-PP</td>
<td>51,892</td>
<td>13,383</td>
</tr>
</tbody>
</table>

(χ²(1)=4832.62, p<0.0001; Crammer’s V = 0.263)

- With of-DEP: (4a) a **group** of British **skiers were** horrified to see a man [BNC: CCK 737]

- Without of-DEP: (4b) **when a whole group is having** a go [BNC: ATAW_non_ac_soc_science]

Table 1

*Verbal agreement with collective nouns with of-PPs and without of-PPs in the BNC and COCA*
3.2.1. Regional variation

Collective nouns with of-PPs

British English and American English show similar tendencies

Figure 1.
Verbal agreement with collective nouns with of-PPs in the BNC and COCA

<table>
<thead>
<tr>
<th></th>
<th>COCA</th>
<th>BNC</th>
</tr>
</thead>
<tbody>
<tr>
<td>PL</td>
<td>1775</td>
<td>1239</td>
</tr>
<tr>
<td>SG</td>
<td>999</td>
<td>594</td>
</tr>
</tbody>
</table>
Table 2. Verbal agreement in the varieties of GloWbE

<table>
<thead>
<tr>
<th>VARIETY</th>
<th>SG Verb</th>
<th>PL Verb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jamaica</td>
<td>33.15</td>
<td>66.85</td>
</tr>
<tr>
<td>Nigeria</td>
<td>35.98</td>
<td>64.02</td>
</tr>
<tr>
<td>Ireland</td>
<td>36.03</td>
<td>63.97</td>
</tr>
<tr>
<td>Tanzania</td>
<td>39.44</td>
<td>60.56</td>
</tr>
<tr>
<td>Great Britain</td>
<td>42.74</td>
<td>57.26</td>
</tr>
<tr>
<td>Sri Lanka</td>
<td>43.98</td>
<td>56.02</td>
</tr>
<tr>
<td>Kenya</td>
<td>44.59</td>
<td>55.41</td>
</tr>
<tr>
<td>Ghana</td>
<td>44.59</td>
<td>55.41</td>
</tr>
<tr>
<td>Pakistan</td>
<td>44.71</td>
<td>55.29</td>
</tr>
<tr>
<td>Canada</td>
<td>47.63</td>
<td>52.37</td>
</tr>
<tr>
<td>South Africa</td>
<td>48.10</td>
<td>51.90</td>
</tr>
<tr>
<td>Australia</td>
<td>48.63</td>
<td>51.37</td>
</tr>
<tr>
<td>New Zeland</td>
<td>49.16</td>
<td>50.84</td>
</tr>
<tr>
<td>Singapore</td>
<td>49.73</td>
<td>50.27</td>
</tr>
<tr>
<td>Malaysia</td>
<td>50.60</td>
<td>49.40</td>
</tr>
<tr>
<td>India</td>
<td>50.84</td>
<td>49.16</td>
</tr>
<tr>
<td>United States</td>
<td>52.18</td>
<td>47.82</td>
</tr>
<tr>
<td>Philippines</td>
<td>52.78</td>
<td>47.22</td>
</tr>
<tr>
<td>Bangladesh</td>
<td>53.29</td>
<td>46.71</td>
</tr>
<tr>
<td>Hong Kong</td>
<td>55.46</td>
<td>44.54</td>
</tr>
</tbody>
</table>

9/20 >55% PL Verbal forms

11/20 <53% PL Verbal forms
### 3.2.1. Regional variation

<table>
<thead>
<tr>
<th>Variety</th>
<th>SG Verb</th>
<th>PL Verb</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ireland</td>
<td>178 (36.03%)</td>
<td>316 (63.97%)</td>
<td>494</td>
</tr>
<tr>
<td>Great Britain</td>
<td>795 (42.74%)</td>
<td>1,065 (57.26%)</td>
<td>1,860</td>
</tr>
<tr>
<td>Canada</td>
<td>291 (47.63%)</td>
<td>320 (52.37%)</td>
<td>611</td>
</tr>
<tr>
<td>Australia</td>
<td>228 (48.10%)</td>
<td>246 (51.90%)</td>
<td>474</td>
</tr>
<tr>
<td>New Zealand</td>
<td>206 (49.16%)</td>
<td>213 (50.84%)</td>
<td>419</td>
</tr>
<tr>
<td>United States</td>
<td>911 (52.18%)</td>
<td>835 (47.82%)</td>
<td>1,746</td>
</tr>
<tr>
<td>TOTAL</td>
<td>2,609 (46.56%)</td>
<td>2,995 (53.44%)</td>
<td>5,604</td>
</tr>
</tbody>
</table>

Table 3. Verbal agreement in the inner varieties in GloWbE
3.2.1. Regional variation

Figure 2.

Pearson residuals in the inner varieties in GloWbE
### 3.2.1. Regional variation

<table>
<thead>
<tr>
<th>Variety</th>
<th>SG Verb</th>
<th>PL Verb</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Great Britain</td>
<td>795 (42.74%)</td>
<td>1,065 (57.26%)</td>
<td>1,860</td>
</tr>
<tr>
<td>United States</td>
<td>911 (52.18%)</td>
<td>835 (47.82%)</td>
<td>1,746</td>
</tr>
</tbody>
</table>

(χ²(1) = 31.78, p < .0001; Crammer’s V = 0.0944)
3.2.2. Syntactic complexity

- **Syntactic distance** (Corbett 1979; Levin 2001)

  Distance increases the likelihood of finding plural agreement
Meaning (not form) is kept activated

Collective $N_{SG}$ + of (...) $N_{PL}(...)$ .......................... $V_{PL}$

- Morphologically SG
- Conceptually PL
3.2.2. Syntactic complexity

<table>
<thead>
<tr>
<th>Structure of the of-PP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. (of) BARE NP</strong></td>
</tr>
<tr>
<td>[a group of <em>boys/people</em>] + V</td>
</tr>
<tr>
<td><strong>2. (of) PREMOD + NP</strong></td>
</tr>
<tr>
<td>[a group of <em>young boys/people</em>] + V</td>
</tr>
<tr>
<td><strong>3. (of) BARE NP + POSTMOD</strong></td>
</tr>
<tr>
<td>[a group of <em>boys/people</em>] <em>from the UK</em> + V</td>
</tr>
<tr>
<td><strong>4. (of) PREMOD + NP + POSTMOD</strong></td>
</tr>
<tr>
<td>[a group of <em>young boys/people</em>] <em>from the UK</em> + V</td>
</tr>
</tbody>
</table>

Table 4.

*Syntactic configurations of of-dependents in the BNC and COCA*
### 3.2.2. Syntactic complexity

<table>
<thead>
<tr>
<th>Structure of the of-PP</th>
<th>SG Verb</th>
<th>PL Verb</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Abs.freq.</td>
<td>%</td>
</tr>
<tr>
<td>1. (of) BARE NP</td>
<td>478</td>
<td>31.95</td>
</tr>
<tr>
<td>2. (of) PREMOD + NP</td>
<td>605</td>
<td>34.99</td>
</tr>
<tr>
<td>3. (of) BARE NP + POSTMOD</td>
<td>265</td>
<td>33.46</td>
</tr>
<tr>
<td>4. (of) PREMOD + NP + POSTMOD</td>
<td>245</td>
<td>41.53</td>
</tr>
</tbody>
</table>

Bare NP vs PREMOD + NP + POSTMOD  
($\chi^2(1)=16.7$, p<0.0001; Cramer’s V = 0.0906)

Table 5.  
Verbal agreement in bare, preMod and postMod of-dependents in the BNC and COCA
3.2.2. Syntactic complexity

Pearson residuals in the structure of of-PP in the BNC and COCA
3.2.3. (C)overt morphology

<table>
<thead>
<tr>
<th>Type of $N_{PL}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overtly-marked (-s)</td>
</tr>
<tr>
<td>a group of <em>boys</em></td>
</tr>
<tr>
<td>Non-overtly-marked (-ø)</td>
</tr>
<tr>
<td>a group of <em>people</em></td>
</tr>
</tbody>
</table>

Table 6.  
*Type of $N_{PL}$ in the BNC and COCA*
### 3.2.3. (C)overt morphology

<table>
<thead>
<tr>
<th>Structure of the of-PP</th>
<th>NN2 (-s)</th>
<th>people</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SG Verb</td>
<td>PL Verb</td>
</tr>
<tr>
<td>1. (of) BARE NP</td>
<td>38.61</td>
<td>61.39</td>
</tr>
<tr>
<td>2. (of) PREMOD + NP</td>
<td>36.60</td>
<td>63.40</td>
</tr>
<tr>
<td>3. (of) BARE NP + POSTMOD</td>
<td>35.48</td>
<td>64.52</td>
</tr>
<tr>
<td>4. (of) PREMOD + NP + POSTMOD</td>
<td>40.66</td>
<td>59.34</td>
</tr>
</tbody>
</table>

(χ²(3), p<0.0001); Crammer’s V = 0.2815

Table 7. Verbal agreement (%) with NN2 and people in bare, preMod and postMod of-dependents in the BNC and COCA.
3.2.4. Semantics

COLLECTIVE NOUN

ANIMACY (Dekeyser 1975, Levin 2001)

(5) And to be a fair traded product, the charter says that erm the company who are selling the products should have an input at that level.

(Humanness (Levin 2001)

(6) We have a tremendous population here that have not discovered what’s on their doorstep.

Influence of semantics of the oblique noun on verbal agreement
3.2.4. Semantics

Figure 4.
Frequency (%) of (in)animate and (non)human oblique nouns in the BNC and COCA
3.2.4. Semantics

\( \chi^2 (1) = 368.5, \ p < 0.0001; \ \text{Crammer’s V} = 0.2672 \)

<table>
<thead>
<tr>
<th></th>
<th>Animate-Human</th>
<th>Animate-NonHuman</th>
<th>Inanimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>PL Verb</td>
<td>2087</td>
<td>119</td>
<td>899</td>
</tr>
<tr>
<td>SG Verb</td>
<td>835</td>
<td>202</td>
<td>1021</td>
</tr>
</tbody>
</table>

(7) A small crowd of **people have** gathered by the door [COCA: FIC MovElf]

(8) A flock of **seabirds lands** in our garden [BNC: CA5 1757]

(9) This batch of **cars was** transferred onto South Metropolitan tracks [BNC: CBK 1744]

Figure 5.
Verbal agreement in relation to the animacy/humanness of the oblique noun in the BNC and COCA
3.2.4. Semantics

VERB MEANING


(10) *The committee comprises/consists of / has eight members.*

\[ \text{*comprise/consist of/have} \quad (\text{Biber et al. 1999: 189}) \]

\[ \Rightarrow \text{Influence of semantics of the verb on verbal agreement} \]
3.2.4. Semantics

VERB MEANING (only BNC)

(Levin 1993)

PL AGR ≥ 60%
- prototypical human reference
- most frequent semantic types:
  possession \((get, give)\)  send/carry
  existence \((live, gather)\)  communication \((say, ask)\)

PL AGR ≤ 59%
- less straightforward connection with human reference
- most frequent semantic types:
  change of state \((increase, rise)\)  appearance \((come, appear)\)
  motion \((run, follow)\)
3.2.5. Idiomatisation

A number of

VERBAL AGREEMENT:

plural agreement →

MEANING: ✓ Quantificational meaning

(11a) [A number of] these papers\textsubscript{pl} appear\textsubscript{pl} to have been a correspondence between this gentleman and his more zealous brethren. [1827 FIC TennesseanANovel]

(11b) From this, [a number of] important consequences\textsubscript{pl} follow\textsubscript{pl} [1990 N ThinkingSociologically]

(11c) [a number of] his poems\textsubscript{pl} have\textsubscript{pl} been sold during those same years [1962 NF EnglishLiterature]
### 3.2.5. Idiomatisation

**A group of**

#### VERBAL AGREEMENT:

- plural agreement
- **BUT** singular AGR has increased →

#### MEANING:  

**Quantificational meaning**

(12a) [a group of] *children_{pl} were_{pl} joyfully embracing the knees [1843 FIC LettersFromNew-]*

(12b) [A group of] *developers_{pl} have_{pl} sued CBS in federal court [1996 NEWS AP]*

(12c) [A group of] *hags_{pl} begin_{pl} beating a poor child [1872 NF Saunterings]*

---

**Figure 7.**  
*Verbal agreement with a group of + plural oblique in COHA*
3.2.5. Idiomatisation

*A majority of*

VERBAL AGREEMENT:

plural agreement →

MEANING: ✓ Quantificational meaning

(13a) [A majority of] cases_{pl} are_{pl} affected in that way [1887 MAG Century]

(13b) [a majority of] Congressmen_{pl} don't_{pl} want to get involved [1971 NEWS WallStJrnl]

(13c) [A majority of] failing thrifts_{pl} have_{pl} no directors and officers liability insurance [1990 NEWS NYT]

Figure 8.

*Verbal agreement with a majority of + plural oblique in COHA*
3.2.5. Idiomatisation

**The majority of**

VERBAL AGREEMENT:

plural agreement →

<table>
<thead>
<tr>
<th></th>
<th>1810s-1850s</th>
<th>1860s-1900s</th>
<th>1910s-1950s</th>
<th>1960s-2000s</th>
</tr>
</thead>
<tbody>
<tr>
<td>plural verb</td>
<td>41</td>
<td>132</td>
<td>214</td>
<td>242</td>
</tr>
<tr>
<td>singular verb</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

**MEANING:** Berg (1998: 54); “function of a quantifier such as *most*”

(14a) the majority of men_{pl} are_{pl} imperfectly educated [1877 NF HistoryConflict]
(14b) the majority of persons_{pl} do_{pl} not believe in their existence [1913 FIC TTembarom]
(14c) the majority of savages_{pl} possess_{pl}s this instinct in a much more perfect form [1897 MAG NorthAmRev]
4. CONCLUSIONS

- Significance of *of*-dependency (with *of*-PP vs. without *of*-PP)
- Main observations:
  - Regional variation: significant variability, discrepancies with literature
  - Distance/complexity: no significant trigger of plural agreement → decreasing tendencies
  - Morphology: overt vs. covert morphology → differences with increasing distance
  - Semantics: influence of animacy and humanness
  - Idiomatisation: fixation of syntactic patterns across time
ACKNOWLEDGEMENTS

Department of Linguistics and English Language (University of Lancaster)

UCREL: University Centre for Computer Corpus Research on Language (Lancaster)

Englisches Seminar (University of Zurich)

David Tizón-Couto (University of Vigo)

Financial support:

• Spanish Ministry of Education (FPU, research grant FPU13/01509)

• the Spanish Ministry of Science and Innovation and the European Regional Development Fund (grants no. FFI2013-44065-P, FFI2014-51873-REDT)

• LVTC research group (University of Vigo)
REFERENCES

Primary sources
- BNC = British National Corpus: [http://bncweb.lancs.ac.uk/bncwebSignup/user/login.php](http://bncweb.lancs.ac.uk/bncwebSignup/user/login.php)
- COCA = Corpus of Contemporary American English: [http://corpus.byu.edu/coca](http://corpus.byu.edu/coca)
- COHA = Corpus of Historical American English: [http://corpus.byu.edu/coha](http://corpus.byu.edu/coha)

Secondary sources
- Berg, Thomas. 1998. The resolution of number conflicts in English and German agreement patterns. Linguistics 36, 41–70.
REFERENCES


What motivates verbal agreement variation with collective-headed subjects?
Evidence from synchrony and diachrony

Yolanda Fernández Pena
yolanda.fernandez@uvigo.es
University of Vigo (Spain)
5 May 2016