Gender differentiation in the Linguistic Innovators Corpus: Grammatical and pragmatic features

Costas Gabrielatos and Eivind Torgersen
Lancaster University

30 March 2010
Lancaster University
Abstract

We report on the use of indefinite article forms (a/an) in front of vowel sounds, as well as certain established and emerging pragmatic markers (PMs), in spoken London English. The study used the Linguistic Innovators Corpus (LIC; Gabrielatos et al., 2010), a 1.3 million word corpus comprising the data from the project, Linguistic Innovators: the English of adolescents in London (Kerswill et al., 2008a), and the Corpus of London Teenage Language (COLT; Stenström et al., 2002).

In the case of a before vowel-initial words (a+vowel), the speakers’ sex was a significant predictor of use, albeit the weakest compared to the other social factors (Gabrielatos et al., 2010). In terms of frequency of use, male speakers were almost twice as likely as female speakers to use a+vowel. Young male speakers were twice as likely to use the pattern, whereas among old speakers there was no significant difference between male and female speakers. However, the inverse picture emerges when the proportion of speakers using the a+vowel pattern is considered. A further notable observation is that, when the extent of use of male and female speakers is compared within age groups (both within LIC and between LIC and COLT), the usage of young male and female speakers appears to be converging.

Patterns more characteristic of male or female speakers were also evident in the use of pragmatic markers (Torgersen et al., under review). More extensively used by male speakers were the emerging PM, you get me, two out of three high frequency PMs (innit, yeah), and the low-frequency PM, right. More extensively used by female speakers were the low-frequency simple PM, ok, and two out of three low-frequency multi-word ones (do you know what I mean, do you know what I’m saying). The rest of the PMs, that is, the high-frequency, you know, and the low-frequency multi-word PM if you know what I mean, showed comparable extent of use in male and female speakers. Although sex was not a very strong predictor (ethnicity and place of residence emerged as the strongest predictors), it was the male non-Anglo speakers who were the most innovative PM users, particularly of innit and you get me. This is at variance with Andersen’s (2001) findings in COLT, namely, that female speakers had the highest frequency of innit, and, therefore, were identified as the innovators.

Overall, our findings show a complex interaction between sex and ethnicity in high-contact speech communities: male non-Anglo speakers have the highest use of innovative features. This has also been found in other studies of grammatical and phonological features in London, using the same data (e.g. Kerswill et al., 2008b).
Why study London English?

• London as the centre of linguistic innovation in British English
  – Diffusion of linguistic features from inner to outer London and beyond

• London as a multicultural city
  – High level of dialect and language contact
Focus

Grammatical feature: *a+vowel*
- A feature of linguistic innovation (Gabrielatos et al., 2010).

Pragmatic markers (PMs)

*innit*
- Innovative pragmatic marker (Cheshire, 1982), but seems to be ... 
  ... becoming established (extent of use)
  ... diversifying functionally

*you get me*
- Emerging pragmatic marker (Torgersen et al., under review).
Linguistic Innovators Corpus (LIC)
(Gabrielatos, et al., 2010)

<table>
<thead>
<tr>
<th>No. of words</th>
<th>1.4 million</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of speakers</td>
<td>118</td>
</tr>
<tr>
<td>Data collection period</td>
<td>2005</td>
</tr>
<tr>
<td>Data collection method</td>
<td>Sociolinguistic interviews</td>
</tr>
<tr>
<td>Age</td>
<td>young=16-18 (average 17); old=70+</td>
</tr>
<tr>
<td>Sex</td>
<td>female; male</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>Anglo; non-Anglo</td>
</tr>
</tbody>
</table>
| Residence             | Inner London (Hackney)  
                          | Outer London (Havering)  |
| Social class          | Working class  |
**Corpus of London Teenage Language (COLT)**
*(Stenström et al., 2002)*

<table>
<thead>
<tr>
<th>No. of words</th>
<th>500,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of speakers</td>
<td>30 recruits + other speakers</td>
</tr>
<tr>
<td>Data collection period</td>
<td>1993</td>
</tr>
<tr>
<td>Data collection method</td>
<td>Self-recordings</td>
</tr>
<tr>
<td>Age</td>
<td>12-17 (average 14)</td>
</tr>
<tr>
<td>Sex</td>
<td>female; male</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>White; ethnic minority</td>
</tr>
</tbody>
</table>
| Residence          | Inner London *(Hackney, Tower Hamlets, Camden)*  
|                    | Outer London *(Barnet)*  
|                    | Hertfordshire          |
| Social class       | Working class; Middle class |
Making LIC and COLT more comparable

**COLT**
- Only young speakers
- Also middle-class recruits
- Also unknown speakers (non-recruits)

→ **COLT-2**: only working-class recruits

→ **LIC-2**: only young speakers
Methodology (1): Annotation and analysis

- Manual annotation of sorted concordances:
  - Genuine instances of ...
    - indefinite article + vowel-initial token
    - pragmatic marker use
  - Checking recordings for ...
    - transcription errors
    - phonological features

- Tabulation of speaker information:
  - user / non-user
  - tokens and types
  - variable values
## Methodology (2): Metrics

<table>
<thead>
<tr>
<th>Expression</th>
<th>Utility</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Density</strong> <em>(a+vowel)</em></td>
<td>Shows the relative frequency of <em>a+vowel</em> use, as opposed to <em>an+vowel</em></td>
</tr>
<tr>
<td><strong>Frequency (PMs)</strong></td>
<td>Shows relative frequency of pragmatic marker use</td>
</tr>
<tr>
<td>Frequency</td>
<td>Frequency of PM per million words</td>
</tr>
<tr>
<td></td>
<td>Frequency per 100 instances of vowel-initial words used with the indefinite article (<em>a</em> or <em>an</em>)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Spread</th>
<th>Shows the proportion of speakers using the feature.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of users per 100 speakers</td>
<td></td>
</tr>
</tbody>
</table>
### a+vowel

#### Predictive utility of social factors

<table>
<thead>
<tr>
<th>Factor</th>
<th>Coefficient</th>
<th>Odds ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (young)</td>
<td>1.404</td>
<td>4.072</td>
</tr>
<tr>
<td>Residence (Hackney)</td>
<td>0.954</td>
<td>2.597</td>
</tr>
<tr>
<td>Ethnicity (non-Anglo)</td>
<td>0.726</td>
<td>2.067</td>
</tr>
<tr>
<td>Sex (male)</td>
<td>0.605</td>
<td>1.832</td>
</tr>
</tbody>
</table>

- Results similar among young speakers.
- All results are highly significant: $p < 0.001$. 
LIC: Use of *a+vowel* (all speakers)

- Male speakers are almost twice as likely to use the pattern.
- However, spread is almost identical.
LIC: Use of *a+vowel* (Anglo speakers)

Focus on age

Old male speakers have comparable density, but higher spread.

Young male speakers have higher density, but lower spread.
LIC: Use of *a+vowel* (Anglo speakers)

Focus on sex

- **Young female speakers** have higher density **and** spread.
- **Young male speakers** have higher density, **but** lower spread.
Usage of Anglo male and female LIC speakers seems to converge.
LIC-2 vs. COLT-2: Use of *a+vowel* by sex
(Anglo + Non-Anglo)

- There is considerable increase in both male and female LIC speakers.
- Increase more marked in female LIC speakers.

→ Male and female use seems to converge.
Pragmatic markers
LIC-2: Extent of use of PMs

![Graph showing the extent of use of PMs with specific terms like "yeah", "you know", and "innit" on the scatter plot.](image-url)
LIC-2: Use of PMs by sex

- **Bold:** both frequency and spread differences are statistically significant.
- **Underline:** only frequency difference are statistically significant - difference is at least 100%.
- **Roman:** only frequency difference are statistically significant - difference clearly below 100%.

<table>
<thead>
<tr>
<th></th>
<th><strong>Male</strong></th>
<th><strong>Female</strong></th>
</tr>
</thead>
</table>
|       | *you get me*  
|       | *innit*  
|       | *yeah*  
|       | *right*  | **Comparable use** |
|       | *you know*  
|       | *if you know what I mean*  | **Emerging** |
|       | **Comparable use** |  
|       | **Female** | **Emerging** |
|       | *ok*  
|       | *(do) (you) know what I mean*  
|       | *(do) you know what I’m saying*  |

**Male speakers:** innovative and high-frequency PMs. weaker preferences

**Female speakers:** low-frequency (+ multi-word) PMs. stronger preferences.
Innit
Grammatical contexts
"Innit": established contexts

• Paradigmatic tag question
  That’s a bit old, *innit*?

• After elliptical host clause
  Every now and then, *innit*?
Innit: innovative contexts

• Non-paradigmatic tag question (Cheshire, 1982)
  They was getting jealous though, *innit*?

• *Innit* as follow-up (Andersen, 2001)
  Zack: But yeah when you're trying to do good, man don't wanna see that though
  Alex: *Innit*
Innit: new functions (invariant tag position ≠ canonical tag)  
(Pichler and Torgersen, 2009)

- After NPs  
  Tina: Yeah, I know I'm a lot smaller than all of them man and who were like "whoa" I mean the sister innit, she's about five times bigger than you. Innit, Mark?

- After subordinate clauses  
  Alan: Or if it got too big, we used to take we we used to have a graveyard next to our school we used to take it there or the cage. That's how it was now if you got beef innit take it to the cage or the graveyard

- Turn-initially, before the proposition over which it has/might have scope  
  Tina: No, I'm not that bad though. But there is (.) innit, there's a scary one in the club.
**Innit: frequency by grammatical context** (young speakers)  
(Pichler & Torgersen, 2009)

- Overall, more frequent in male speech in all localities.
- Non-paradigmatic tokens more frequent in male speech in all localities.
- Follow-up and new functions slightly more frequent in inner London female speech.
Conclusions

\textit{a+vowel}

- Among the four social factors investigated, sex has the smallest effect on its use.
- Observed changes point towards convergence of use by male and female speakers.

Pragmatic Markers

- Extent of use by sex appears to be fairly polarised.
- Inner city male speakers have higher use of emerging and high-frequency PMs.
- Male and female speakers also seem to diverge in what grammatical contexts \textit{innit} is used.
References


