A corpus-based study of invariant tags in London English

Eivind Torgersen & Costas Gabrielatos
Lancaster University

Multicultural London English: The emergence, acquisition and diffusion of a new variety (2007–10)

Investigators:
Paul Kerswill (Lancaster University)
Jenny Cheshire (Queen Mary, University of London)

Research Associates:
Sue Fox, Arfaan Khan, (Queen Mary, University of London)
Eivind Torgersen (Lancaster University)

Funded by the Economic and Social Research Council
http://www.lancs.ac.uk/fss/projects/linguistics/innovators/
http://www.lancs.ac.uk/fss/projects/linguistics/multicultural/
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
Why study invariant tags?

• Frequent in spoken language

• Frequent in young people’s speech
  – Young people are linguistic innovators
  – We’re interested in linguistic innovation

• An innovative tag user = A linguistic innovator?
Invariant tags

• Part of a tag question

• Anchor and tag
  – Canonical tag
    • *It’s cold, isn’t it?*
    • *They’re late, aren’t they?*
  – Invariant tag
    • *It’s cold, innit?*
    • *They’re late, innit?*

• *innit* is an invariant tag
We examined:

• Simple invariant tags
  – *innit, okay, right, yeah*

• Multi-word invariant tags: (elliptical) clauses
  – *You get me*
  – *You know*
  – *(Do)/(If) (you) know what I mean*
  – *(Do) (you) know what I’m saying*
Some examples

• but he’s been here for her innit?

• they just ain’t got nothing innit?

• she’s coming up for sixty now yeah?

• no the thing right? I I didn’t mind right? but the thing that pissed me off is the that she brung the fight into the house

• he makes you laugh but he’s just annoying if you know what I mean? but he’s just he will come behind me once yeah? he come behind me and he’s got my hood on my jacket and stuck it over my head
Linguistic Innovators Corpus (LIC)

• 2005/2008
• 1,3 million words
• London boroughs:
  – Hackney (inner London)
  – Havering (outer London)
• Speaker data:
  – age, gender, ethnicity, friendship network, social class (all working class)
• Sociolinguistic interviews
Sociolinguistic variables in LIC

• Age
  – Young (16-19)
  – Old (60+)

• Place of residence:
  – Hackney (inner London)
  – Havering (outer London)

• Sex

• Ethnicity
  – Anglo
  – Non-Anglo
Corpus of London Teenage Language (COLT)

- 1993
- 500,000 words
- London boroughs:
  - Hackney, Tower Hamlets, Camden (inner London)
  - Barnet (outer London), Hertfordshire
- Speaker data:
  - age, gender, social class
- Self recordings
Metrics

• **Frequency**
  – occurrences per million words

• **Spread**
  – Ratio (%) of tag users to speakers
Young LIC speakers: frequency and spread

- innit
- yeah
- you know
- (do) (you) know
- right
- you get me
- (do) (you) know
- what I mean
- if you know
- what I mean
- I'm saying
- know what
- ok
- 0 10 20 30 40 50 60 70 80 90 100
Young speakers in LIC and COLT: Comparison of frequencies

yeah
innit
right
ok
you get me

LIC
COLT
Difference ratio: LIC/COLT
(without you get me)

-20 -15 -10 -5 0 5

yeah innit right ok
Difference ratio LIC/COLT
(including you get me)
Looking at tags:
Use by sociolinguistic variables
<table>
<thead>
<tr>
<th></th>
<th>Age</th>
<th>Sex</th>
<th>Ethnicity</th>
<th>Residence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>innit</strong></td>
<td>Young</td>
<td>Male</td>
<td>Non-Anglo</td>
<td>---</td>
</tr>
<tr>
<td><strong>ok</strong></td>
<td>---</td>
<td>Female</td>
<td>Non-Anglo</td>
<td>---</td>
</tr>
<tr>
<td><strong>right</strong></td>
<td>Young</td>
<td>Female</td>
<td>Non-Anglo</td>
<td>Hackney</td>
</tr>
<tr>
<td><strong>yeah</strong></td>
<td>Young</td>
<td>---</td>
<td>Non-Anglo</td>
<td>---</td>
</tr>
<tr>
<td><strong>you get me</strong></td>
<td>Young</td>
<td>---</td>
<td>Non-Anglo</td>
<td>Hackney</td>
</tr>
<tr>
<td><strong>you know</strong></td>
<td>Old</td>
<td>Male</td>
<td>Anglo</td>
<td>---</td>
</tr>
<tr>
<td><strong>(do) (you) know what I mean</strong></td>
<td>Young</td>
<td>Female</td>
<td>Anglo</td>
<td>Havering</td>
</tr>
<tr>
<td><strong>if you know what I mean</strong></td>
<td>Young</td>
<td>---</td>
<td>---</td>
<td>Havering</td>
</tr>
<tr>
<td><strong>(do) you know what I’m saying</strong></td>
<td>Young</td>
<td>Female</td>
<td>---</td>
<td>Havering</td>
</tr>
</tbody>
</table>

- **Bold:** both frequency and spread differences are statistically significant.
- **Normal:** only frequency differences are statistically significant.
- ‘---’ : both frequency and spread are similar/comparable.
innit
okay
yeah
you get me
you know
(do) (you) know what I mean
(do) (you) know what I’m saying
If you know what I mean
Looking at sociolinguistic variables:
Tag usage in sociolinguistic groups
## Tags and age

<table>
<thead>
<tr>
<th>Freq. + Spread</th>
<th>Young</th>
<th>Old</th>
<th>No clear preference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>you get me</em></td>
<td>--</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Freq. only</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>innit</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>right</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(do) (you) know what I mean</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>if you know what I mean</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(do) you know what I’m saying</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>yeah</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>you know</td>
<td></td>
<td></td>
<td><em>ok</em></td>
</tr>
</tbody>
</table>

- Vast majority of tags more frequently used by young speakers.
# Tags and sex

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Female</th>
<th>No clear preference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Freq. + Spread</strong></td>
<td>--</td>
<td>•  <em>ok</em></td>
<td></td>
</tr>
<tr>
<td><strong>Freq. only</strong></td>
<td>• <em>innit</em></td>
<td>• <em>right</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• <em>you know</em></td>
<td>• <em>(do) (you) know what I mean</em></td>
<td>• <em>you get me</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• <em>(do) you know what I’m saying</em></td>
<td>• <em>if you know what I mean</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• <em>yeah</em></td>
</tr>
</tbody>
</table>

- Females show preference for a larger number of tags.
- Males show preference for the high frequency tags.
## Tags and ethnicity

<table>
<thead>
<tr>
<th></th>
<th>Anglo</th>
<th>non-Anglo</th>
<th>No clear preference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freq. + Spread</td>
<td>--</td>
<td>• you get me</td>
<td></td>
</tr>
<tr>
<td>Freq. only</td>
<td>• you know (do) (you) know what I mean</td>
<td>• innit • ok • right • yeah</td>
<td>• if you know what I mean • (do) you know what I’m saying</td>
</tr>
</tbody>
</table>

- Non-Anglos ...
  - show preference for a larger number of tags.
  - have a significantly higher frequency of all simple tags
  - show significantly higher frequency and spread for the innovative tag, *you get me*. 
### Tags and inner and outer city

<table>
<thead>
<tr>
<th></th>
<th>Hackney</th>
<th>Havering</th>
<th>No clear preference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Freq. + Spread</strong></td>
<td>• <em>you get me</em></td>
<td>--</td>
<td></td>
</tr>
<tr>
<td><strong>Freq. only</strong></td>
<td>• <em>right</em></td>
<td>• <em>(do) (you) know what I mean</em></td>
<td>• <em>innit</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• <em>if you know what I mean</em></td>
<td>• <em>ok</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• <em>(do) you know what I’m saying</em></td>
<td>• <em>you know</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• <em>yeah</em></td>
</tr>
</tbody>
</table>

- *you get me* and *right* characteristic of hackney
- The multi-word *you know* tags are most frequent in Havering.
- Comparable preference for most simple tags.
## Friendship network (1)

<table>
<thead>
<tr>
<th>Tag</th>
<th>Comparison of average scores of users/non-users</th>
<th>Two-dimensional comparison: number of users/non-users with high/low scores</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>innit</em></td>
<td>No significant effect</td>
<td>No significant effect</td>
</tr>
<tr>
<td><em>ok</em></td>
<td>No significant effect</td>
<td>No significant effect</td>
</tr>
<tr>
<td><em>right</em></td>
<td>No significant effect</td>
<td>No significant effect</td>
</tr>
<tr>
<td><em>yeah</em></td>
<td>No significant effect</td>
<td>No significant effect</td>
</tr>
<tr>
<td>Tag</td>
<td>Comparison of average scores of users/non-users</td>
<td>Two-dimensional comparison: number of users/non-users with high/low scores</td>
</tr>
<tr>
<td>-------------</td>
<td>-------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
| you get me  | • Users have a 30% higher average network score than non-users | • Users of *you get me* can be expected to belong to high-density multi-ethnic friendship networks.  
• No safe predictions can be made regarding the score of non-users.  
→ High network score does not predict use.  
→ Low network score predicts non-use.  
→ Tag still emerging. |
Innovative tags and linguistic innovation

• Established tags, irrespective of whether they are becoming more or less frequent, are widespread enough to no longer depend on multi-ethnic interactions.

• On the contrary, new (innovative) tags, like you get me, are currently used significantly more frequently within the multi-ethnic networks in which they have probably first emerged.
Tags as part of the bigger picture

• Innovation in the invariant tag system is in line with previous findings on innovation in phonology and grammar.

• Inner-city, non-Anglo males are in the lead

• The frequency of new tags is highest in high-density multi-cultural friendship groups