A corpus-based analysis of indefinite article use in London English

Costas Gabrielatos & Eivind Torgersen
(Lancaster University)

Part of the project:
An analysis of spoken London English using corpus tools

PI: Eivind Torgersen
CI: Paul Kerswill
RA: Costas Gabrielatos
Consultant: Sebastian Hoffmann
Overview

• Link to previous and current projects
• Theoretical considerations
• Data and methodology
• Analysis and conclusions (so far)

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
Theoretical background

- Variation in the use of indefinite article in …
  - British English dialects
  - Child language
  - Contact varieties
  - L2 varieties
Lack of *a/an* alternation in British English dialects

- Wright (1905:71): ‘very few dialects follow the rule of the literary language according to which *an* is used before a vowel’.

- Shorrocks (1999:45) on the Bolton dialect: ‘the indefinite article is pronounced /ə/ or /ən/ before vowels and diphthongs’.

- Wagner (2008:418) on dialects in Southwest England: ‘the indefinite article often occurs as *a* before vowels’.
Survey of English Dialects

• One response: ‘a/an April fool’

• ‘a April fool’ is more frequent than ‘an April fool’ in localities in:
  – The East Midland counties;
  – Somerset (the rest of the Southern counties have not recorded the article or it was undetected);
  – The West Midland counties;
  – The Northern counties

• So basically everywhere!
Lack of *a/an* alternation in child language

- Newton and Wells (1999)

- Percentage of ‘adult’ examples increased from just below 30 at the age of 3 to 80 at the age of 7

- The development follows very much that of the definite article and other morphology
**a/an in African American Vernacular English (AAVE)**

- Labov (1972) *Language in the inner city*:
  - ‘He is a expert’
  - ‘Bell mother got a old beat-up boot’


**a/an in L2 varieties of English**

- Seliger (1979)

- Studied *a/an* alternation among young L2 speakers of English (younger than 13 and adult learners)

- Preference for *a* in both vowel and consonant environments, but used *an* more frequently with increased age
Tower Hamlets (Fox, 2007) (1)

- Tower Hamlets (London borough just South of Hackney)
- 39 speakers
- Age: 12-17
- 9 girls (all of White British origin)
- 30 boys (11 White British, 17 Bangladeshi, 2 mixed)
- Small sample: 94 indefinite article tokens in pre-vocalic position
- Before vowels:
  - an: 65%
  - a: 35%
Tower Hamlets (Fox, 2007) (2)

• Use of a
  – Bangladeshi boys: 75%
  – Mixed race White/Afro-Caribbean boys: 35%
  – White boys: 15%
  – Boys: 44%
  – Girls: 5%
Friendship groups in Tower Hamlets
Research questions and hypotheses

• What variables are good predictors of use?
• Effect of phonological context?

• Non-Anglo boys in Hackney will be the highest users.
• A person with a multicultural friendship group will use a before vowels more often.
Data

• Corpus of approx. 1.4 million words, transcribed orthographically.

• 117 speakers:
  – L1 = English
  – Some bilingual

• 1,241 tokens

• Five variables:
  – Sex: female (53), male (64)
  – Age: old=70+ (18), young=16-18 (99)
  – Ethnicity: Anglo (77), non-Anglo (40)
  – Residence: Hackney (58), Havering (59)
  – Network (only ‘young’): 1 (16), 2 (9), 3 (19), 4 (25), 5 (15)
Methodology (1)

• Concordance analysis and annotation of *a* and *an* (sorted at R1, R2, R3) to identify (and count):

  – instances of ‘indefinite article + adjective/noun’ (continuous or discontinuous)

  – speaker identity of ‘*a* + vowel’ users
off Sue: so that's been abused is it? Grace: yeah /[Maria: it is]/ because a a girl was on a train .. because her bluetooth was on .. and some man

the wrong one Hadiya: it's cool Sue: was it? Hadiya: yeah Sue: is that a a girl's school or mixed? Hadiya: it's a mixed school is a near .. like in

and then pearl harbour i think november or december '41 so they were a a gimmer of hope .. especially in '42 or '43 . when the germans were

it's great but it's a [long flight] Monty: [and i] want to go to india as well . a a goa my daughter went there my middle daughter went there .. they

movement [mm] erm . er in the trade union movement because i'm a a gold member .. retired of the um now the gmb er municipal .. workers

i mean if they'd had done it on the television you'd have thought it was a a good comedy sketch . apparently if you go there .. and you pay with
to sleep something and i always finish up by saying thank you for being a a good audience and those who went to sleep thank you for not snoring

Ollie: oh yeah yeah yeah Sue: what about you? who , you gave me a a group of people . who would you say was your closest friend? ...

mm Talal: like .. it's just .. the Rom the Romford people they they got like a a high tone innit comes out of their mouth like .. like they they bring
quite good and they've tried to do things and they've tried to open ?? of a a history club there you know and .. i'm here this afternoon i go along
you span .. well .. not the entire century like the queen mother did but a a huge chunk of it [Mrs Knight: mm] and the changes that's been [Mrs
[Sue: mm] and erm when i used to steam er blanche in boiling water in a a I had a special saucepan with like a mesh thing which used to be

know years ago you know bits of schrapnel you know so it was it was a a i suppose to children it was er er can i say an exciting time you know
the teachers used to go to him and er/used to but he er his .. he.. suffered a a illness i don't know what it was and his boy took it up but then he
right so she didn't move very far either Arthur: no that's right no no .. cos a a in those days there was a tendency to live near your fam .. i i had what
on the staff [Sue: mm] so he got the pension and [Sue: oh i see] a good a aindex linked pension which proved to be very beneficial [Sue: mm] so

just bsc just before he passed away we went to cuba together .. a a it it was fun [Sue: yeah] it's nice I mean it's ss so it's like the .. so m
.. telly on the ar radio .. on sundays they .. or there's a chap there he gives a a jazz concert and he gets all different places as what you know.

of achievement and that and then /Sue: yeah]/ me and my friends got a a jeep .. Eisa: really? we got in a limo Isabella: I think we fitted fourteen
diseases in supply there he's he's been there a number of years he had a a job in er er in pensions or something of that nature but .. something he
joint income so i mean it's an impossibility to buy anything e even getting a a joint income a joint mortgage [yeah] because if you i i. if your got a

er er municipal properties to housing associations .. and they're making a a killing aren't they i mean you got the peabody you got the guiness you

a problem here in a minute like can you try and help man cut yeh give me a a knife or something yeh cos i'm gonna need to stab through' i mean

in the schools .. bad spelling .. bad writing [mm] .. sometimes you get a a letter from somebody a younger person and you can't read it .. even

Sue: oh right did you get those? Adam: y- actually I wan- I wanted to do a A level but when I was at the exams they told me to go straight to the ..

they're safe and the other people are not .. whereas I've grown up with a a like more of white society .. and more of a like and like some Asians

mind you're saying you need more .. but like if you .. like say you had a a line one night .. and .. you think "i want more and more" but the next
you to make things but .. it's not got a label on it you know it's got to be a a little tick on it [Shirley: designer] or something .. even even the young

where did you go Edie: i went to eastbourne .. and er . i finished up in a a little village called willington that's about . five miles from the coast of
running and then he start . popping it after him .. yeah but .. there was a a a lot of people running/ Sue: this was one of your emr lot was it? Brian:
# Annotation examples

<table>
<thead>
<tr>
<th>Category</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>a</em> is not an article</td>
<td><em>to have a a at least</em></td>
</tr>
<tr>
<td>Repetition</td>
<td><em>a a book</em></td>
</tr>
<tr>
<td>Self–correction</td>
<td><em>have you got a an extended family?</em></td>
</tr>
<tr>
<td></td>
<td><em>it would go like a like two years</em></td>
</tr>
<tr>
<td>Hesitation / fillers</td>
<td><em>a erm</em></td>
</tr>
<tr>
<td>Unfinished utterances</td>
<td><em>... you know what I mean it’s a. you might live in Clapton and you ...</em></td>
</tr>
<tr>
<td>Backchannel interruption</td>
<td>Sue: <em>mm what's /[Fatima: I liked my a]/ your favourite subject at school</em></td>
</tr>
<tr>
<td>Humorous or metalinguistic uses of ‘<em>a</em> + vowel’</td>
<td><em>a orse</em></td>
</tr>
<tr>
<td></td>
<td><em>... you can't say a aeroplane you have to say an aeroplane</em></td>
</tr>
</tbody>
</table>
Methodology (2)

- Tabulation of speaker information:
  - user / non-user
  - tokens and types
  - variable values

- Manual calculation of frequencies and cross-tabulations.
Methodology (3)

- Metrics
  - tokens / speakers  \( (\text{density}) \)
  - users / speakers  \( (\text{spread}) \)
  - Rough indicator:  \( \text{density} \times \text{spread} \)  \( (\text{preference}) \)

<table>
<thead>
<tr>
<th></th>
<th>Speakers</th>
<th>Users</th>
<th>Tokens</th>
<th>Density</th>
<th>Spread</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>64</td>
<td>37</td>
<td>119</td>
<td>1.86</td>
<td>57.8%</td>
</tr>
<tr>
<td>Female</td>
<td>53</td>
<td>30</td>
<td>63</td>
<td>1.18</td>
<td>56.6%</td>
</tr>
<tr>
<td>Total</td>
<td>117</td>
<td>67</td>
<td>182</td>
<td>1.56</td>
<td>57.3%</td>
</tr>
</tbody>
</table>
Low predictive utility

- **Phonological context**
  - No pattern emerged from the examination of use with *a* and *an*.
  - Analysis was in terms of phonemes **not actual phones**.

- **Network score**
  - Correlation of network scores to usage preference symptomatic of the very strong correlation between usage and ethnicity.
  - High network multi-ethnicity not a good predictor overall.

**However**

- Network score good predictor of use among Anglo speakers.
Network score (1)

Average network score of users almost 40% higher than that of non-users.

But

Difference of the users’ network score from average is about 50% smaller than that of non-users.
The seeming correlation between network scores and use is probably due to the correlation between network score and ethnicity (AA=2.75; NA=4.48).
Indefinite article use: tokens

<table>
<thead>
<tr>
<th></th>
<th>Raw freq.</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>a</em> + vowel</td>
<td>182</td>
</tr>
<tr>
<td><em>an</em> + vowel</td>
<td>1,059</td>
</tr>
<tr>
<td>Total</td>
<td>1,241</td>
</tr>
</tbody>
</table>
## Indefinite article use: types

<table>
<thead>
<tr>
<th></th>
<th>Raw freq.</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>a</em> only</td>
<td>48</td>
</tr>
<tr>
<td><em>an</em> only</td>
<td>252</td>
</tr>
<tr>
<td>Both</td>
<td>60</td>
</tr>
<tr>
<td>Total</td>
<td>360</td>
</tr>
</tbody>
</table>

![Pie chart showing article use percentages]

- *an* only: 70%
- *a* only: 13.33%
- Both: 16.67%
Use in terms of individual variables
Interaction of density and spread

- Female
- Male
- Young
- Hackney
- Havering
- Old
- Anglo
- Non-Anglo

Density vs. Spread graph
Comments

• Unlikely that clusters have formed by chance.

• Typical user profile:

  [NON-ANGLO] [YOUNG] [MALE] [HACKNEY RESIDENT]

• [NON-ANGLO] and [HACKNEY RESIDENT] much stronger predictors than [YOUNG] and [MALE].

• [OLD] seems to be a stronger predictor of non-use than [YOUNG] is a predictor of use.
Use in terms of pairwise interactions between variables (cross-tabulations)
Interaction of density and spread

![Graph showing the interaction of density and spread with different categories labeled on the axes and points representing various combinations of density and spread. The categories include HV+Y, NA+Y, AS+O, HV+F, HK+F, HV+M, NA+F, AS+F, Y+F, Y+M, HV+O, AS+Y, HK+Y, NA+HK, NA+M, HK+M, AA+HK, NA+HV, HK+O, O+F, and HV+O. The graph includes grid lines with values for density and spread ranging from 0 to 3.5 and 0 to 100, respectively.]}
Most frequent values in combinations with above-average density and spread

<table>
<thead>
<tr>
<th>Group</th>
<th>Freq. in pairs</th>
<th>Possible combinations</th>
<th>% of possible combinations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-Anglo</td>
<td>4</td>
<td>5*</td>
<td>80%</td>
</tr>
<tr>
<td>Hackney</td>
<td>4</td>
<td>6</td>
<td>67%</td>
</tr>
<tr>
<td>Male</td>
<td>3</td>
<td>6</td>
<td>50%</td>
</tr>
<tr>
<td>Young</td>
<td>3</td>
<td>6</td>
<td>50%</td>
</tr>
<tr>
<td>Female</td>
<td>2</td>
<td>6</td>
<td>33%</td>
</tr>
</tbody>
</table>

* There are no old non-Anglo speakers in the sample.
Pairwise preference
Comments and conclusions

- Ethnicity and residence stronger predictors than age and sex.
- Sequence of predictive utility:
  ETHNICITY > RESIDENCE > AGE > SEX
- Network score not a dependable predictor overall.
- No discernible phonological patterns.
- The analysis assumed that all corpus speakers produced the same or similar number of words.
- The analysis in terms of number of words per speaker may reveal different patterns of use.