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

Costas Gabrielatos & Eivind Torgersen
(Lancaster University)

This paper reports on the analysis of the use of indefinite article forms (*a/an*) in front of vowel sounds in spoken London English, which formed a part of the completed project *Analysis of spoken London English using corpus tools* (funded by the British Academy). The study used the Linguistic Innovators Corpus (LIC), a 1.4 million word corpus comprising the transcribed and marked-up interview data from the Lancaster/Queen Mary ESRC-funded project, *Linguistic innovators: the English of adolescents in London* (Kerswill et al. 2008), as well as the Corpus of London Teenage English (COLT) (Stenström et al. 2002).

The research methodology combined approaches and techniques from sociolinguistics and corpus linguistics. Variables were examined individually and in cross-tabulations, using both manual/semi-automated and automated techniques (logistic regression analysis). The former analysis took account of the frequency of the *a+vowel* pattern relative to the number of opportunities for a choice between *a* or *an* (i.e. vowel-initial words preceded by the indefinite article) and the proportion of speakers who used the pattern.

The study examined both linguistic and sociolinguistic variables, but only the sociolinguistic variables yielded statistically significant results. This suggests that the linguistic variables play a minor role, if any at all, in the choice between *a* or *an* in front of a vowel sound. The sociolinguistic variables comprised the speakers' sex, age, ethnicity and place of residence, as well as the ethnic make-up of the friendship networks. In particular the speakers' ethnicity and place of residence, emerged as the strongest predictors of the use of *a* before vowels.

The comparative analysis of LIC and COLT showed an almost three-fold increase in the use of *a* before vowel-initial words by young speakers (19% and 8% respectively). Equally striking is the three-fold change in the proportion of young speakers who use the *a+vowel* pattern (58% and 20% respectively). More specifically, in LIC, the majority of speakers (52%) alternate between *a* and *an*, 43% use *an+vowel* only, and 5% use *a+vowel* only. In contrast, the vast majority of COLT speakers (85%) use only *an+vowel*, with a small minority (15%) alternating between *a+vowel* and *an+vowel* - no COLT speaker uses *a+vowel* only.

The indefinite article form *a* before vowels seems to have undergone a process of reallocation (Britain & Trudgill 1999) in which its sociolinguistic status has been realigned. While the form *a* in front of vowels earlier seemed to have been avoided, either because it was socially stigmatised or only formed a part of child language and L2 varieties, it is now frequently found among adolescent speakers in inner London. We argue that the indefinite article form *a* before vowels forms part of Multicultural London English (Kerswill et al. 2008), along with other phonological and grammatical features that have already been documented.

References

Britain, David & Peter Trudgill. (1999). Migration, new-dialect formation and sociolinguistic refunctionalisation: Reallocation as an outcome of dialect contact. *Transactions of the Philological Society* 97(2). 245-256.

Kerswill, Paul, Jenny Cheshire, Susan Fox & Eivind Torgersen. (2008). *Linguistic innovators: the English of adolescents in London*. Final report submitted to the ESRC.

Stenström, Anne-Brita, Gisle Andersen & Kristine Hasund. (2002). *Trends in teenage talk.* Amsterdam: Benjamins.

ICAME 30, Lancaster, 27-31 May 2009

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

Costas Gabrielatos & Eivind Torgersen

(Lancaster University)

Part of the project:

Analysis of spoken London English using corpus tools

PI: Eivind Torgersen

CI: Paul Kerswill

RA: Costas Gabrielatos

Consultant: Sebastian Hoffmann

Supported by The British Academy
THE NATIONAL ACADEMY FOR THE HUMANITIES AND SOCIAL SCIENCES

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


HERTFORDSHIRE
BUCKS
ESSEX
BERKSHIRE
KENT
SURREY
Enfield
Barnet
Harrow
Haringey
Waltham Forest
Redbridge
Havering
Hillingdon
Brent
Camden
Isling-ton
Hackney
Barking
Newham
Ealing
West-Minster
City
Tower Hamlets
Hammersmith
Ken & Chel.
South-wark
Greenwich
Hounslow
Lambeth
Bexley
Richmond
Wandsworth
Lewisham
Merton
Kingston
Bromley
Sutton
Croydon
SCALE
0
Miles
10
0
Km
10

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’

Child language, L2 varieties, contact varieties

- Increase of *a~an* alternation with increased age in both L1 and L2.
- Increase of *a~an* alternation for AAVE speakers in multi-ethnic friendship groups.
- Decreased *a~an* alternation for white speakers in multi-ethnic friendship groups.

Tower Hamlets (Fox 2007)

- Tower Hamlets (London borough just South of Hackney)
  - *an+vowel:* 65%
  - *a+vowel:* 35%

- Use of *a+vowel*
  - Bangladeshi boys: 75%
  - Mixed race White/Afro-Caribbean boys: 35%
  - White boys: 15%

  - Boys: 44%
  - Girls: 5%

Research questions and hypotheses

- What variables are good predictors of use?
  - Linguistic
  - Sociolinguistic

- Predictions:
  - Non-Anglo boys in Hackney will be the highest users.
  - A person in a multicultural friendship group will use *a+vowel* more often.

The Linguistic Innovators Corpus (LIC)

Data collected	2005
Data collection method	Sociolinguistic interviews
No. of speakers	118
Age	young=16-18 old=70+
Sex	female male
Ethnicity	Anglo non-Anglo
Residence	Inner London (Hackney) Outer London (Havering)
Social class	Working class

Variables (1)

Linguistic variables

- Initial vowel sound (phoneme)
- Stress pattern (stress-bearing syllable)
- Number of syllables

Variables (2)

Sociolinguistic variables

Variable	Values and number of speakers		
Age	old=70+ *(18)*	|	young=16-18 *(100)*
Sex	female *(53)*	|	male *(65)*
Ethnicity	Anglo *(77)*	|	non-Anglo *(41)*
Residence	Hackney *(58)*	|	Havering *(60)*

Variables (3)

Friendship network score

Score	Ethnicity composition	Speakers
1	all friends same ethnicity as self	*16*
2	up to 20% of a different ethnicity than self	*9*
3	21-40% of a different ethnicity than self	*19*
4	41-60% of a different ethnicity than self	*26*
5	61-80% of a different ethnicity than self	*30*

Methodology (1)

- Analysis and annotation of sorted concordances of *a* and *an* :
  - genuine instances of ‘indefinite article + vowel-initial token
  - socio-demographic details of *a/an*+vowel users

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

Methodology (2)

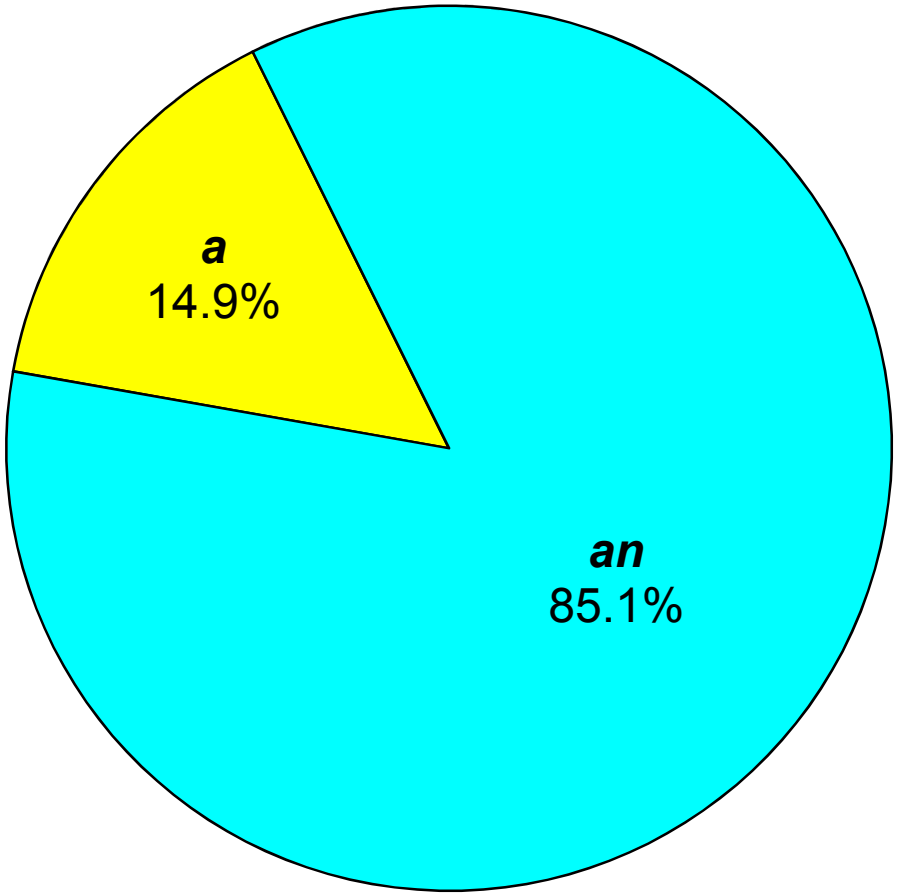
	Expression	Utility
Density	Frequency per 100 instances of indefinite article + vowel-initial word	Shows the relative frequency of *a+vowel* use, as opposed to *an+vowel*.
Spread	Number of users per 100 speakers	Shows the proportion of speakers using *a+vowel*.

Logistic regression analysis

- Variables considered individually and together
- Cross-tabulations

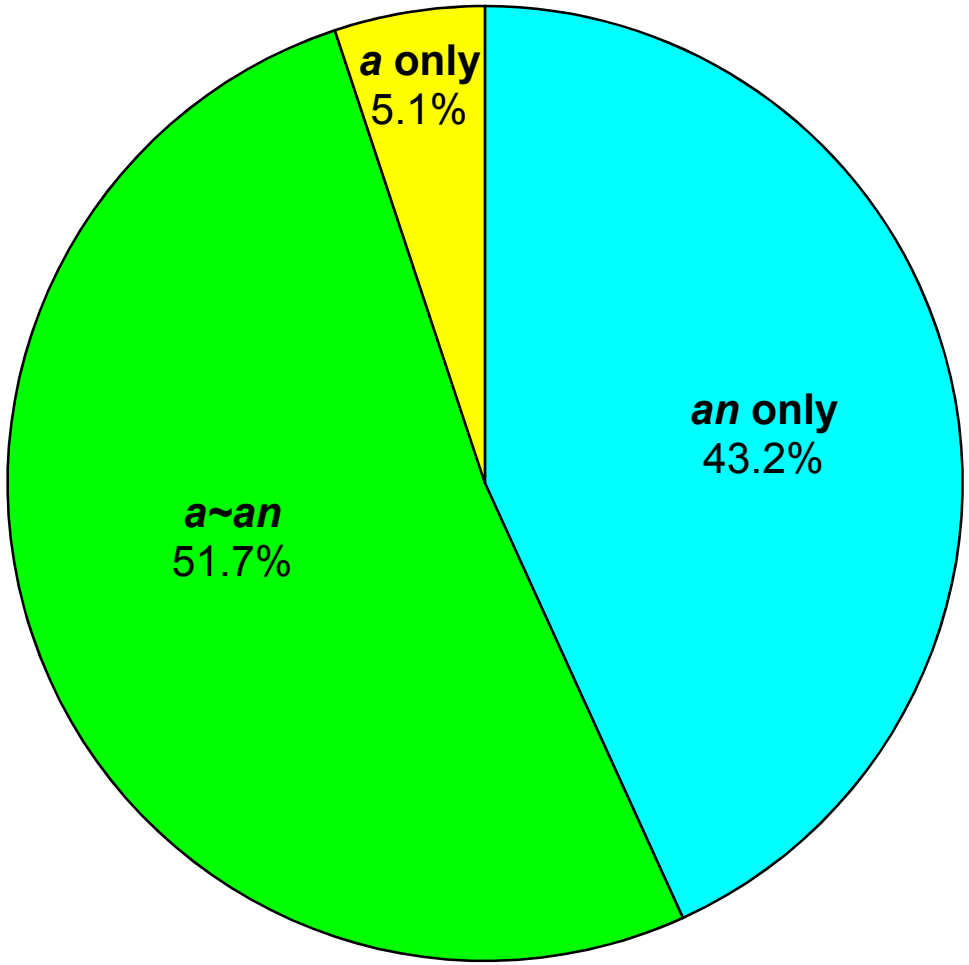
Indefinite article use: density

	Raw freq.
***a* + vowel**	182
***an* + vowel**	1042
Total	1224

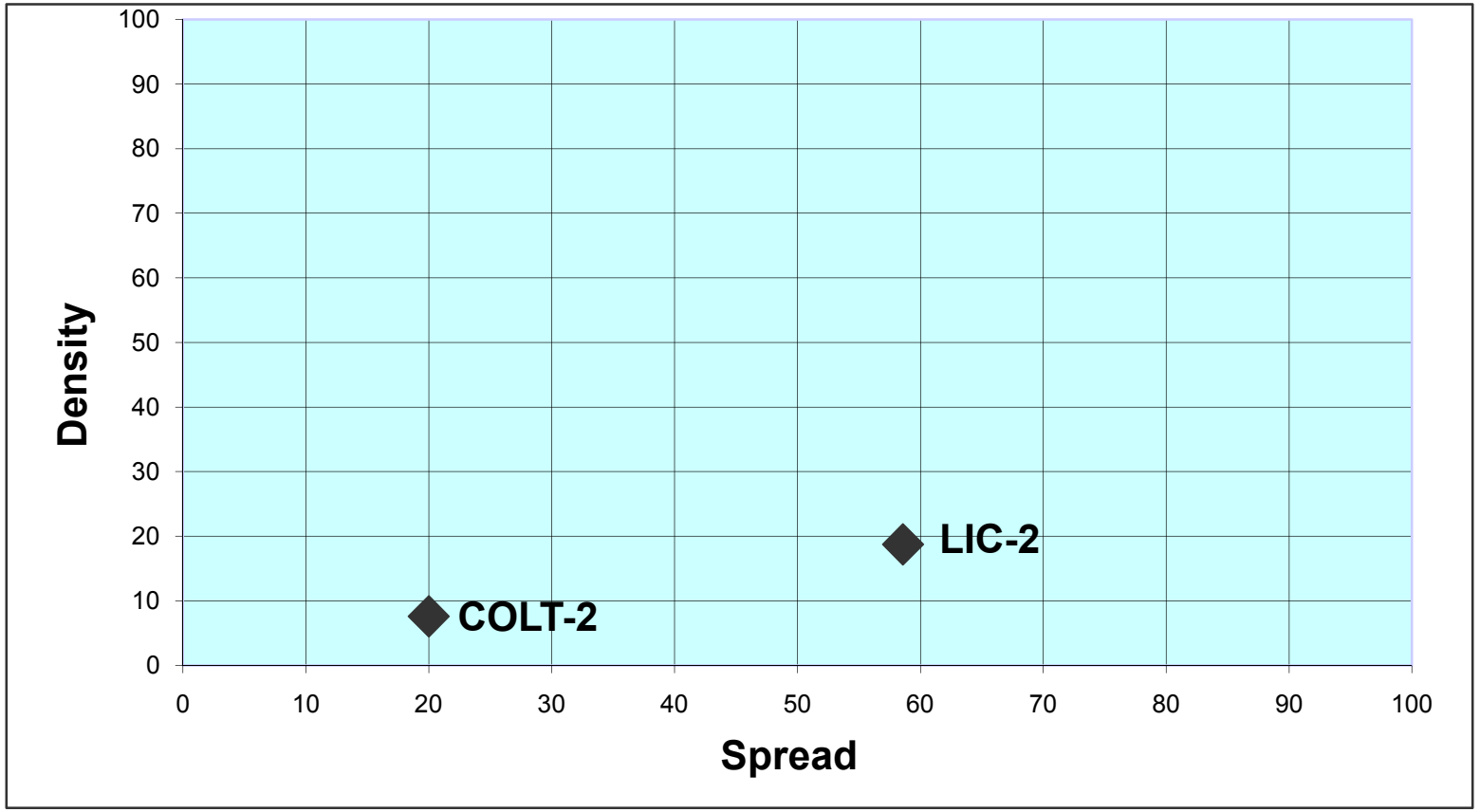


Indefinite article use: spread

	Speakers
Alternation (*a* ~ *an+vowel*)	61
No alternation (*a+vowel* only)	6
No alternation (*an+vowel* only)	51
Total	118



a+vowel: LIC-2 vs. COLT-2



	Freq.	V-initial tokens	Users	Speakers	*Density*	*Spread*
LIC-2	170	907	58	100	*18.7***	*58.0**
COLT-2	9	119	3	15	*7.6*	*20.0*

Effect of linguistic variables

- Stress pattern
- Number of syllables

→ No effect on use of *a+vowel*

  - *main effects*
  - *interactions*

- Initial vowel sound (phoneme)

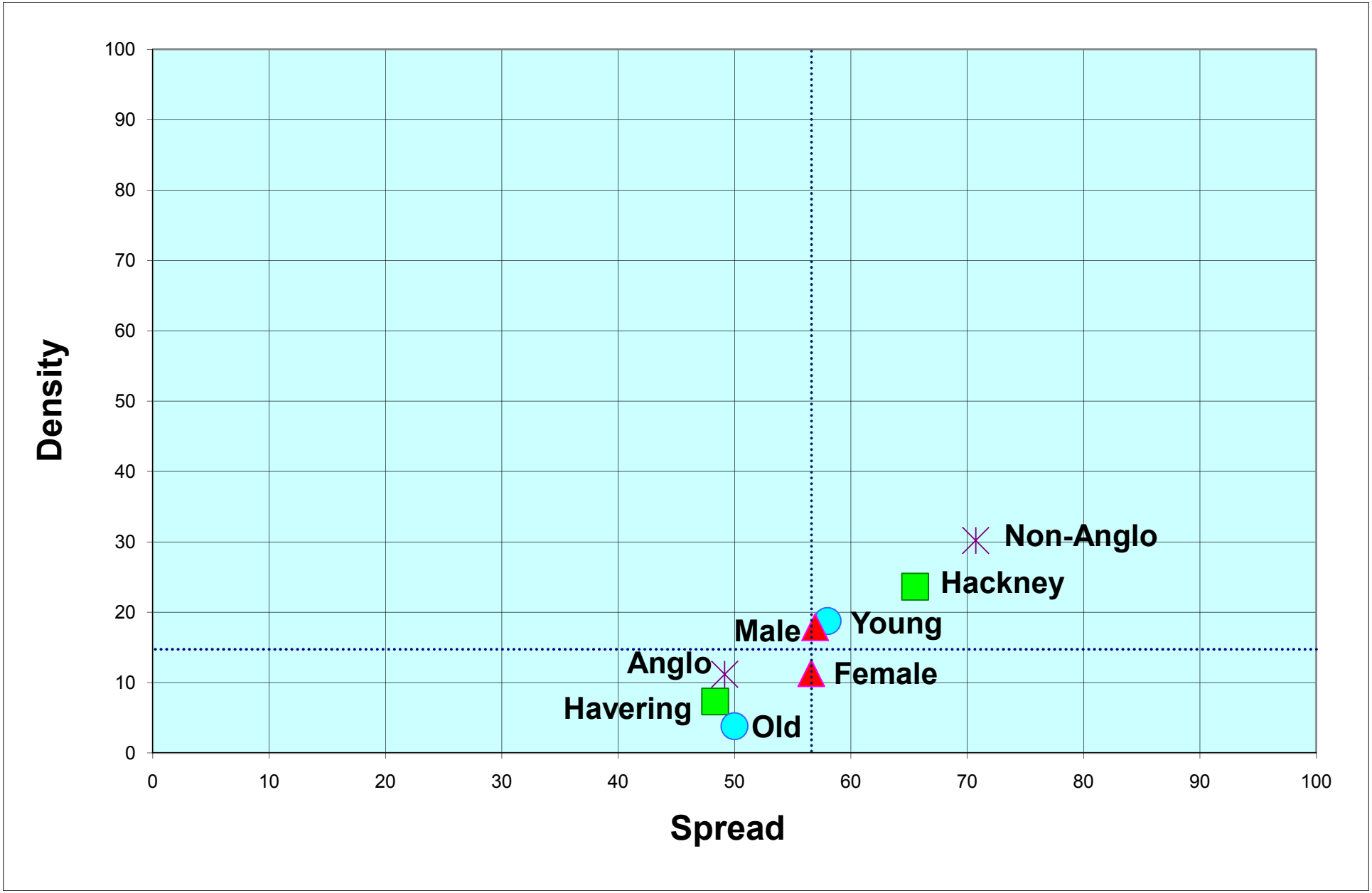
→ Only /aʊ/ predicts *a+vowel* ($p<0.001$)

→ No significant effect when regression model included all three linguistic variables.

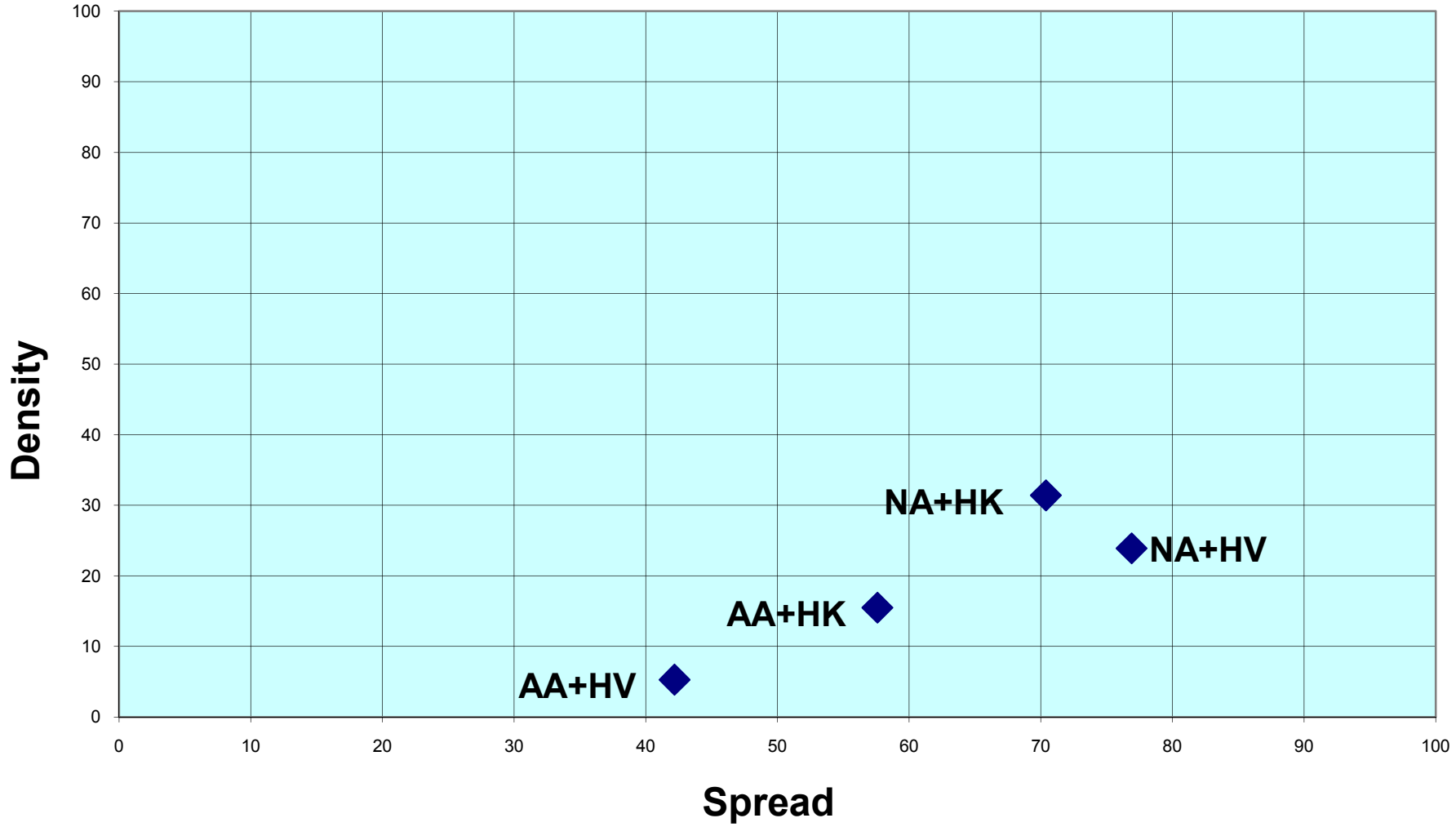
Effect of sociolinguistic variables

- All individual variables have a significant effect on *a+vowel* use
  - Age = young
  - Sex = male
  - Ethnicity = non-Anglo
  - Residence = Hackney
- When interactions are included in the model
  - Ethnicity*Residence (non-Anglo*Hackney) emerges as a strong predictor.

Individual variables: density and spread



Ethnicity*Residence: density and spread



AA=Anglo, NA=non-Anglo, HK=Hackney, HV=Havering

Network score (1)

Average scores

- Non-Anglo=*4.48;* Anglo=*2.75*;
- Hackney=*4.42*; Havering=*2.48*

Prediction

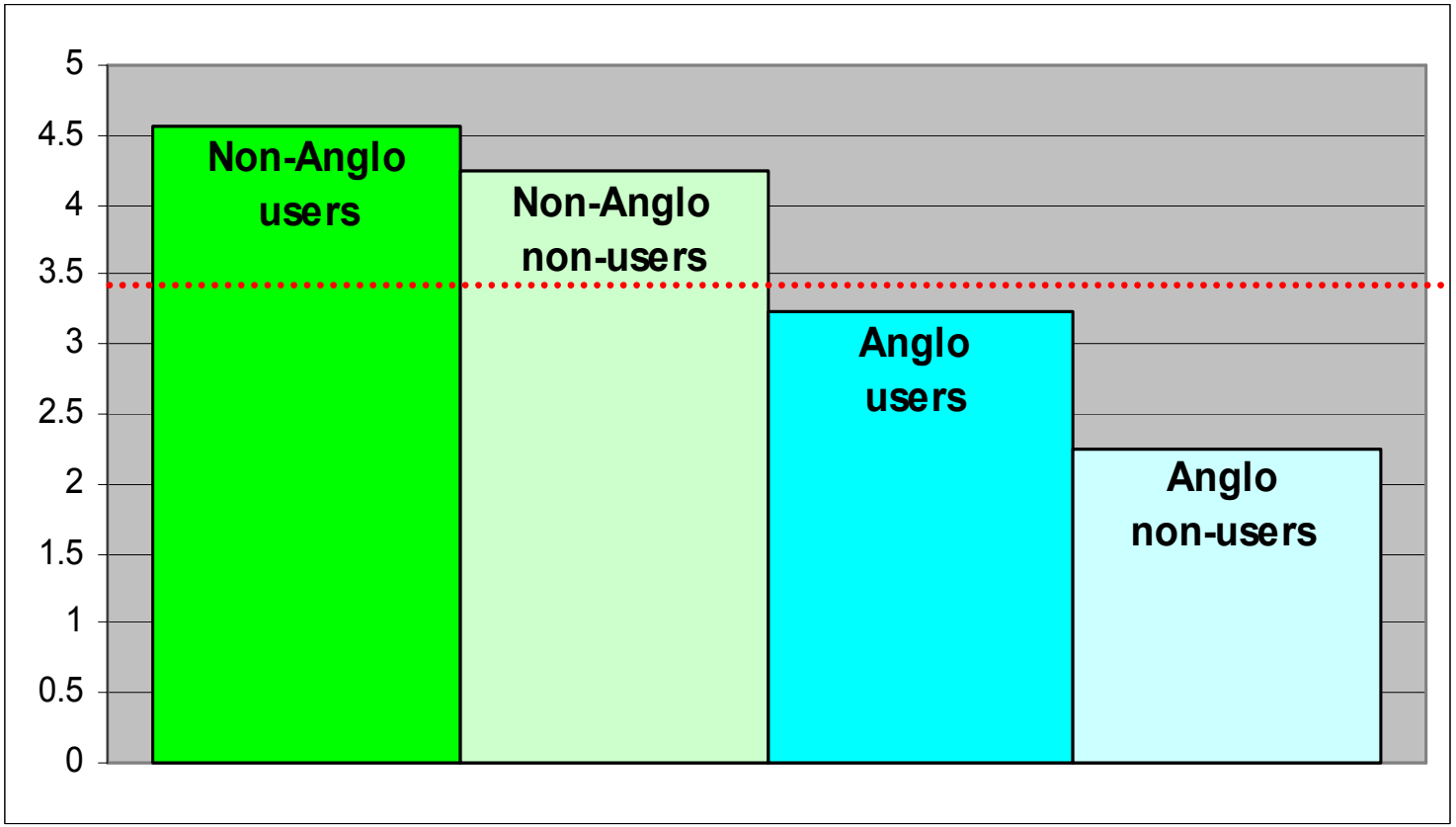
- Correlation between network score and *a+vowel* use

Why?

- NA and HK show higher preference for *a+vowel*
- High score networks by definition have high proportion of NA speakers

→ If so, AA and NA users of *a+vowel* should have higher average scores than non-users.

Network score (2)



Network score (3)

- Correlation of network scores to usage preference probably symptomatic of …
  - the correlation between network score and ethnicity.
  - the very strong correlation between usage and ethnicity (non-Anglo).

- Overall, friendship network multi-ethnicity not a good predictor of *a+vowel* use.

However

- Network score good predictor of use among Anglo speakers.

Conclusions

- Strongest predictors of *a+vowel*:
  - Age (young)
  - Ethnicity*Residence (non-Anglo*Hackney)

- Reallocation of indefinite article form *a+vowel* due to extensive dialect contact in inner London:
  - *a+vowel* used in informal styles among young speakers

- *a+vowel* a feature of Multicultural London English - along with several other phonological and grammatical features already documented

Appendix

Annotation examples

Category	Example
a is not an article	▪ *to have a a at least*
Repetition	▪ *a a book*
Self-correction	▪ *have you got a an extended family?* ▪ *it would go like a like two years*
Hesitation / fillers	▪ *a erm*
Unfinished utterances	▪ *… you know what I mean it's a. you might live in Clapton and you …*
Backchannel interruption	▪ Sue*: mm what's /*[Fatima*: I liked my a*]*/ your favourite subject at school*
Humorous or metalinguistic uses of '*a* + vowel'	▪ *a orse* ▪ *… you can't say a aeroplane you have to say an aeroplane*

Indefinite article use: types

	Raw freq.
***a* only**	41
***an* only**	253
Both	63
Total	357