CLAVIER 09, Modena, 5-7 November 2009

A corpus-based study of indefinite article and invariant tag use in London English

Eivind Torgersen & Costas Gabrielatos

(Lancaster University)

Analysis of spoken London English using corpus tools

PI: Eivind Torgersen

CI: Paul Kerswill

RA: Costas Gabrielatos

Consultant: Sebastian Hoffmann

Abstract

This paper reports on the analysis of the use of indefinite article forms (a/an) in front of vowel sounds, as well as certain established and emerging invariant tags (*yeah, innit, right, okay, you get me*) in spoken London English. The study used the Linguistic Innovators Corpus (LIC; Gabrielatos et al., forthcoming), a 1.3 million word corpus comprising the transcribed and marked-up interview data from the 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.

The analysis of indefinite article forms examined both linguistic and sociolinguistic variables, but only the sociolinguistic variables yielded statistically significant results for use of indefinite article forms. 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 indefinite article form a before vowels seems to have undergone a process of reallocation 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.

For tags, the comparison of LIC and COLT revealed an increase in *yeah* and, in particular, *innit*, and a dramatic increase in *you get me*, but a decrease in the relative frequencies of *right* and *okay*. The analysis of LIC showed that all the innovative tags, such as *innit* and *you get me*, were clearly a feature of young people's speech. In addition, the most innovative tag, *you get me*, was by far most frequent in inner London. The ethnic minority speakers, and male speakers in general, are the most innovative tag users, particularly of *innit* and *you get me*, but the ethnic minority speakers also had high frequencies of *yeah, okay* and *right*, and they were therefore the highest users of tags overall. Overall, there is a difference in tag usage between inner and outer London: the more innovative tags are more frequent in inner London, and the more traditional ones in outer London. The innovative tags *you get me* and *innit* were most frequent, and were used by a larger proportion of speakers, among male, ethnic minority, inner city residents.

We argue that the indefinite article form *a* before vowels and innovative use of tags form part of Multicultural London English (Kerswill et al. 2008), along with other phonological and grammatical features that have already been documented.

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

Research questions and hypotheses

- What factors are good predictors of use?
  - Linguistic? No significant effect
  - Social? Yes

- Predictions:
  Innovative forms will be more frequent among ...
  - non-Anglo male speakers in Hackney.
  - speakers in multicultural friendship groups.

The Linguistic Innovators Corpus (LIC)

No. of words	1.4 million
No. of speakers	118
Data collection period	2005
Data collection method	Sociolinguistic interviews
Age	young=16-18; old=70+
Sex	female; male
Ethnicity	Anglo; non-Anglo
Residence	Inner London (Hackney) Outer London (Havering)
Social class	Working class

Corpus of London Teenage Language (COLT)

- 1993
- 500,000 words
- Self recordings
- Speaker data:
  - Age
  - Sex
  - Residence: Inner London (Hackney, Tower Hamlets, Camden); outer London (Barnet); Hertfordshire
  - Social class

- COLT also contains speech by middle-class recruits and unknown speakers (non-recruits).

→ COLT-2: only working-class recruits.
→ LIC-2 : only young speakers

LIC: Social Factors

Factor	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*)

Methodology (1): Annotation and analysis

- Manual annotation of sorted concordances:
  - Genuine instances of ...
    - indefinite article + vowel-initial token
    - invariant tag use

- Checking recordings for ...
  - transcription errors
  - phonological features

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

Methodology (2): Metrics

	Expression	Utility
Density *(a+vowel)*	Frequency per 100 instances of *a+vowel*	Shows the relative frequency of *a+vowel* use, as opposed to *an+vowel*
Frequency (tags)	Frequency of tag per million words	Shows relative frequency of invariant tag use
Spread	Number of users per 100 speakers	Shows the proportion of speakers using the feature.

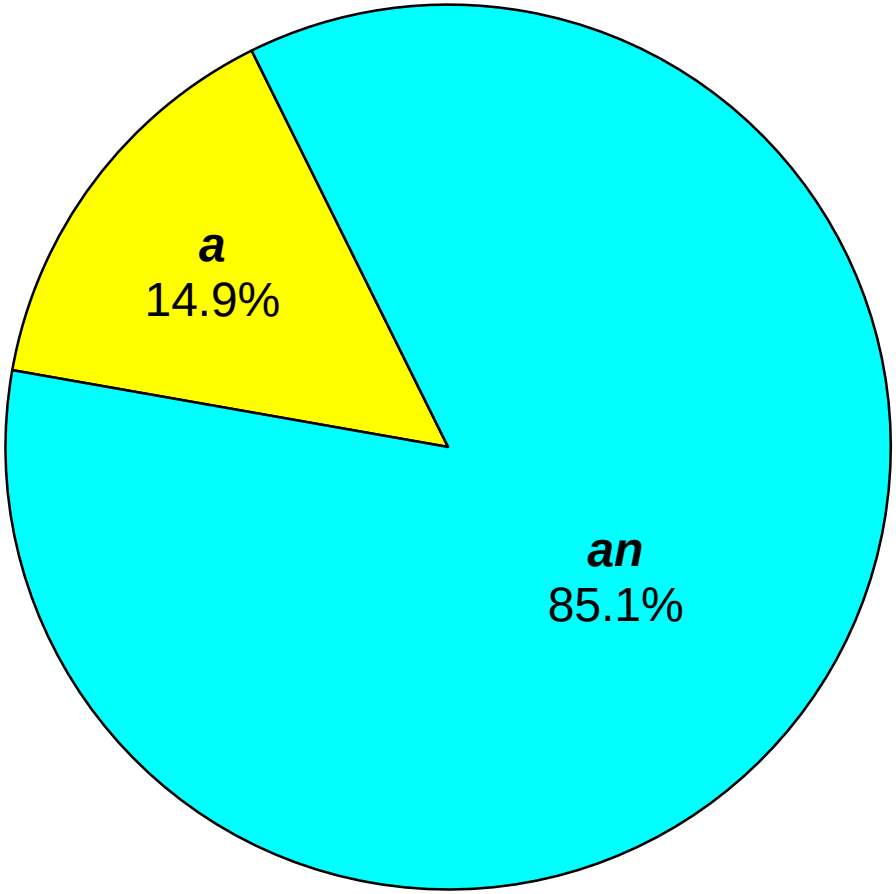
Logistic regression analysis

- Variables considered individually and together
- Pairwise cross-tabulations

Indefinite article

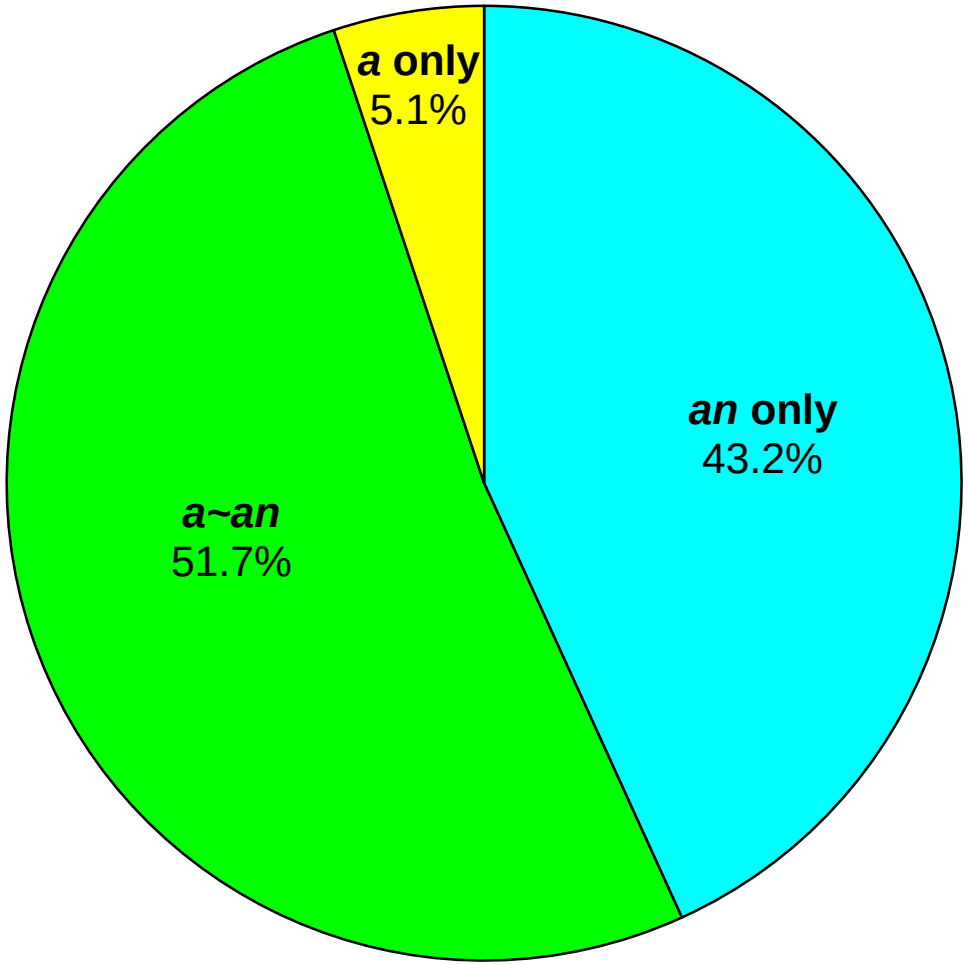
Indefinite article use before vowel sounds: Density

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

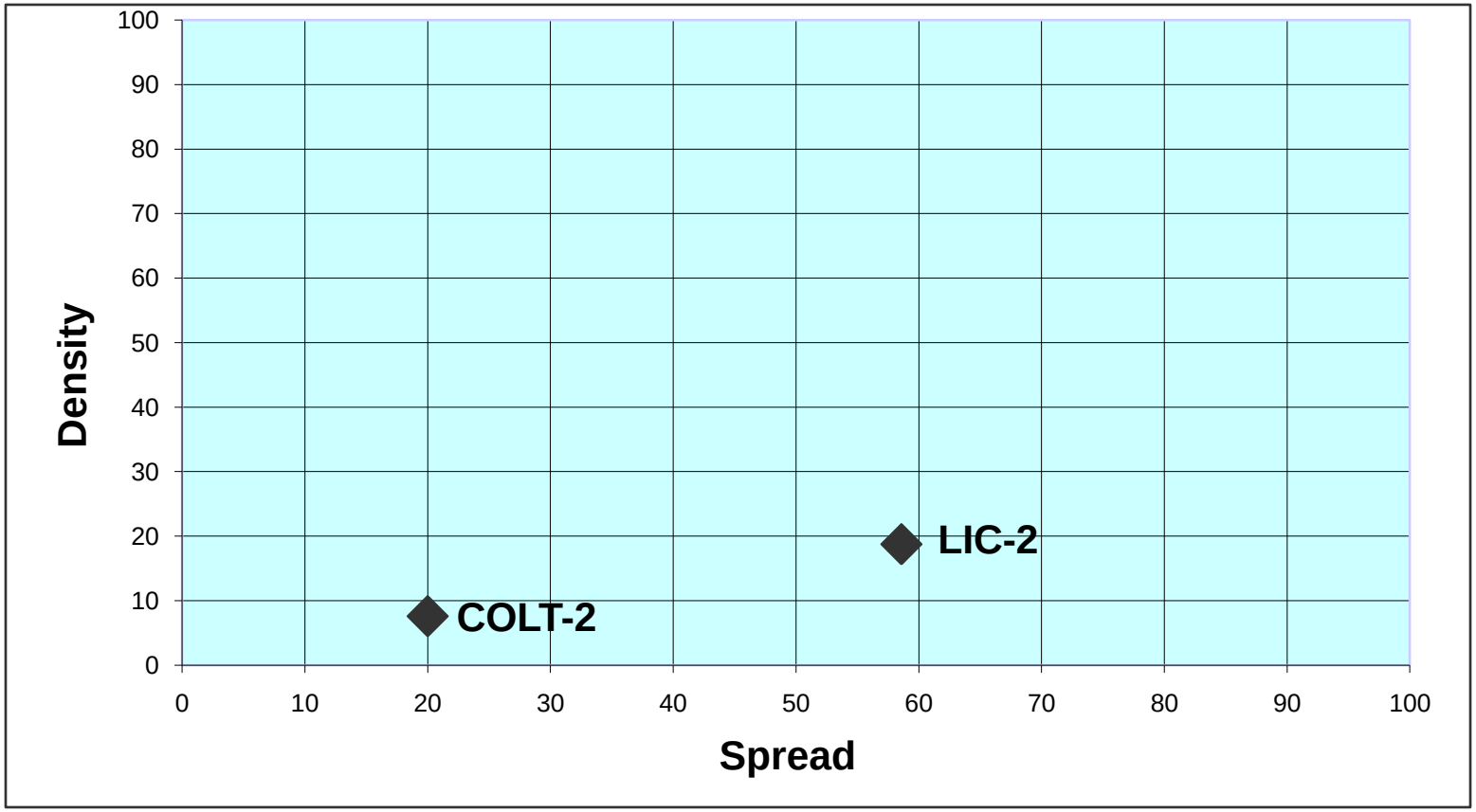


Indefinite article use before vowel sounds: Spread

	Speakers
a~an+vowel	61
a+vowel only	6
an+vowel only	51
Total	118



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



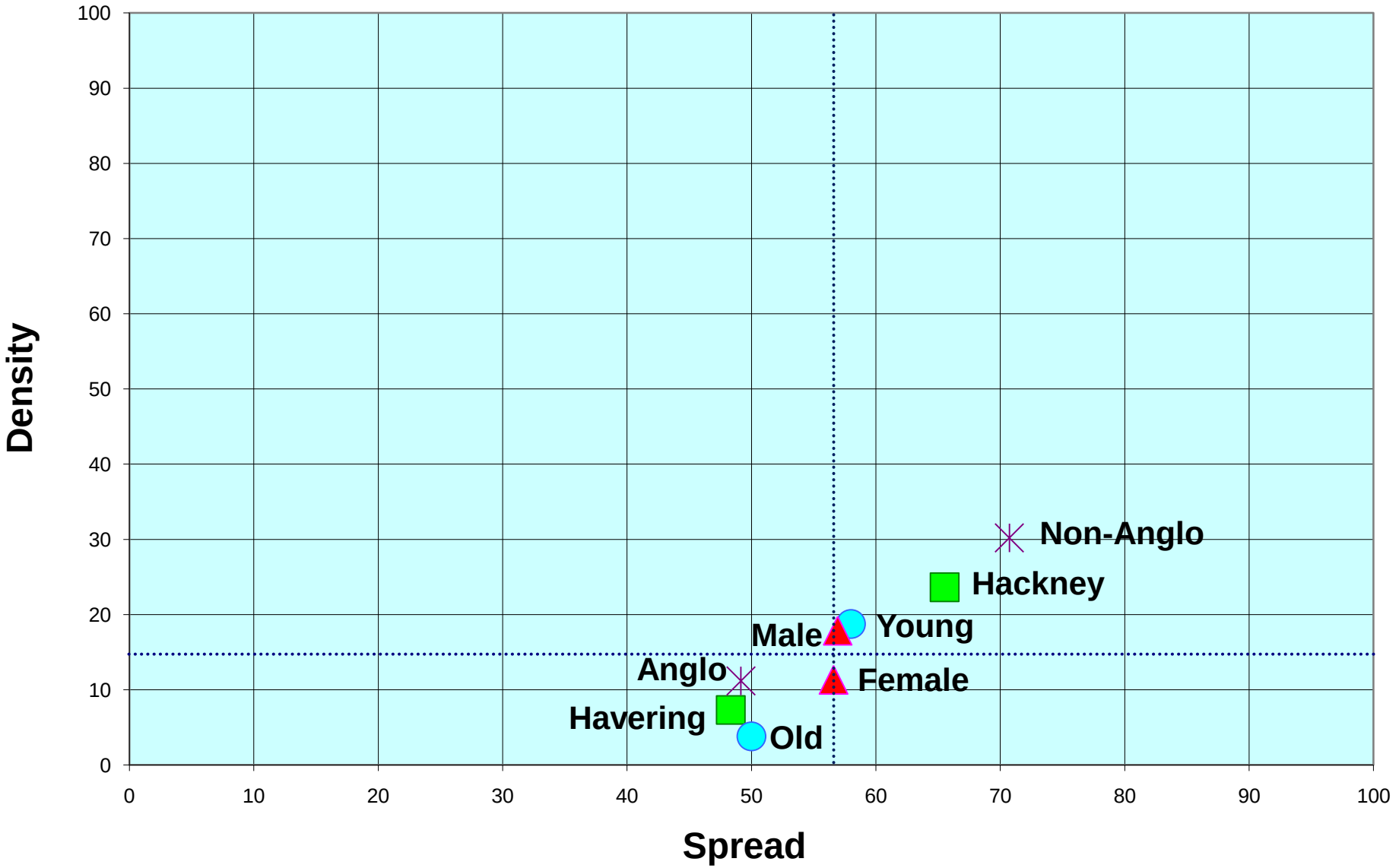
	Freq.	V-initial tokens	Users	Speakers	*Density*	*Spread*
LIC-2	170	907	58	100	*18.7***	*58.0**
COLT-2	9	119	3	15	*7.6*	*20.0*

Effect of social factors on *a+vowel* use

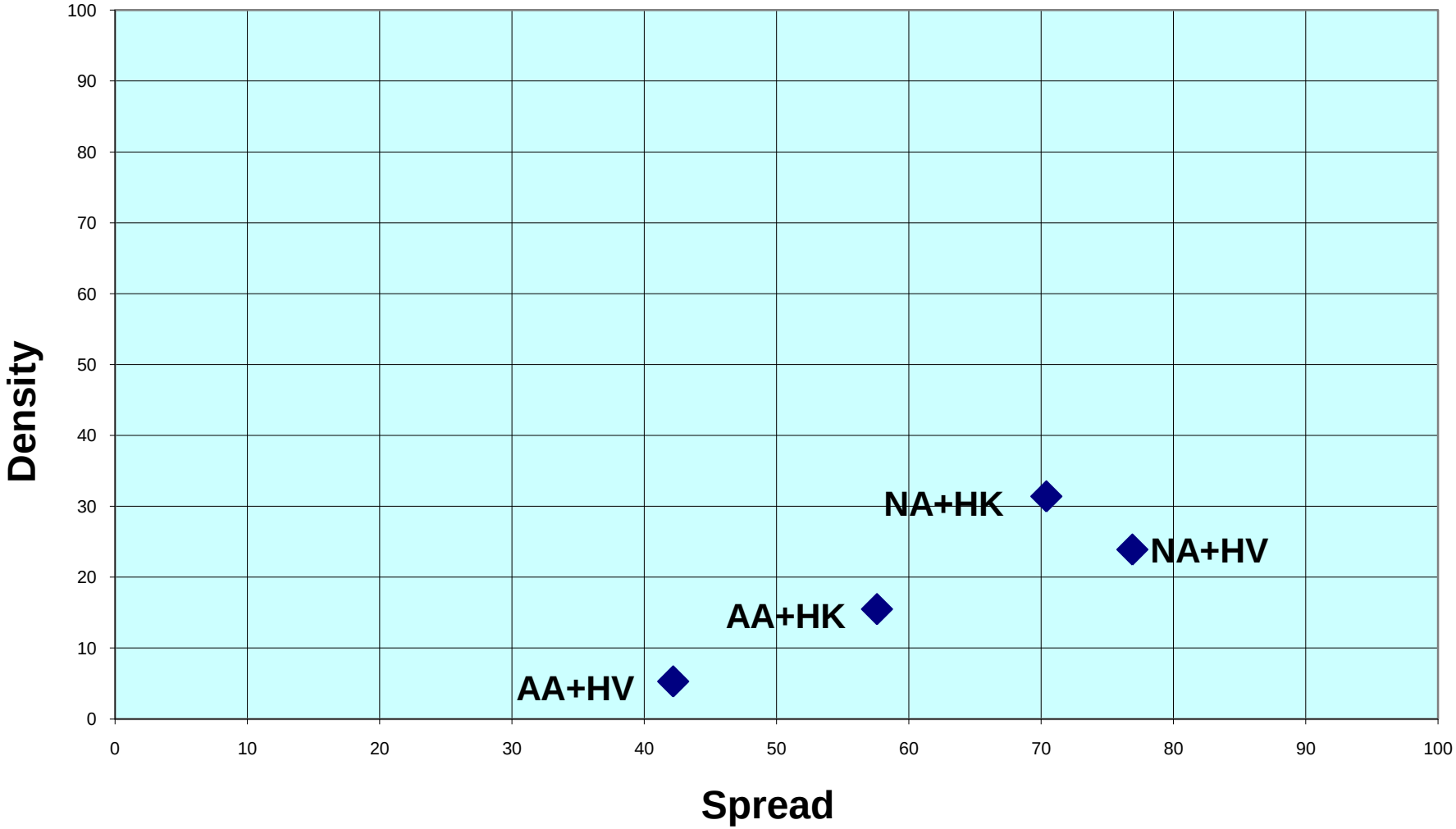
- All individual factors have a significant effect:
  - Age = young
  - Sex = male
  - Ethnicity = non-Anglo
  - Residence = Hackney (inner London)

- When interactions are included in the model ...
  - Ethnicity*Residence (= non-Anglo*Hackney) emerges as a strong predictor.

Individual social factors: density and spread

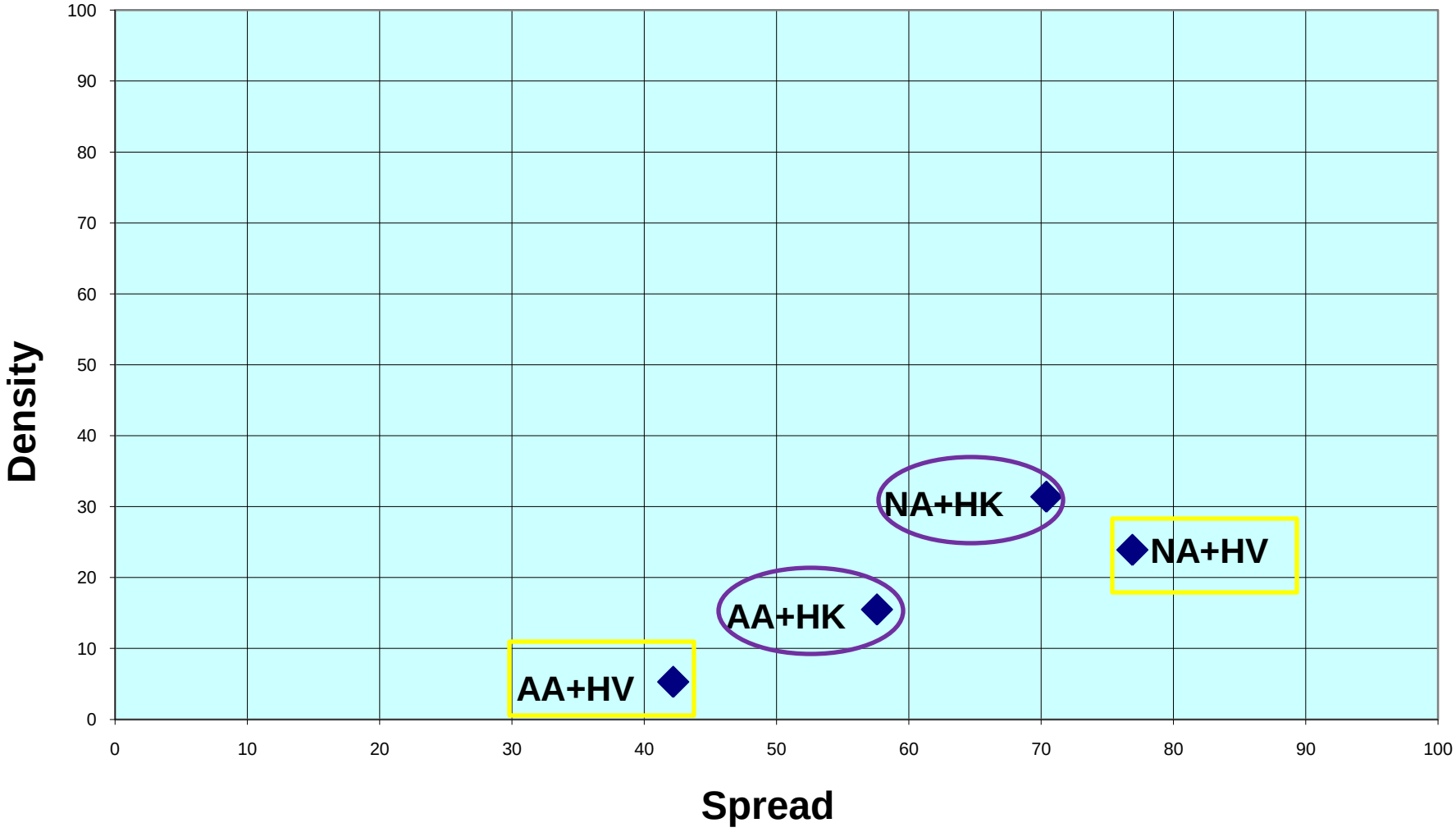


Ethnicity*Residence: density and spread



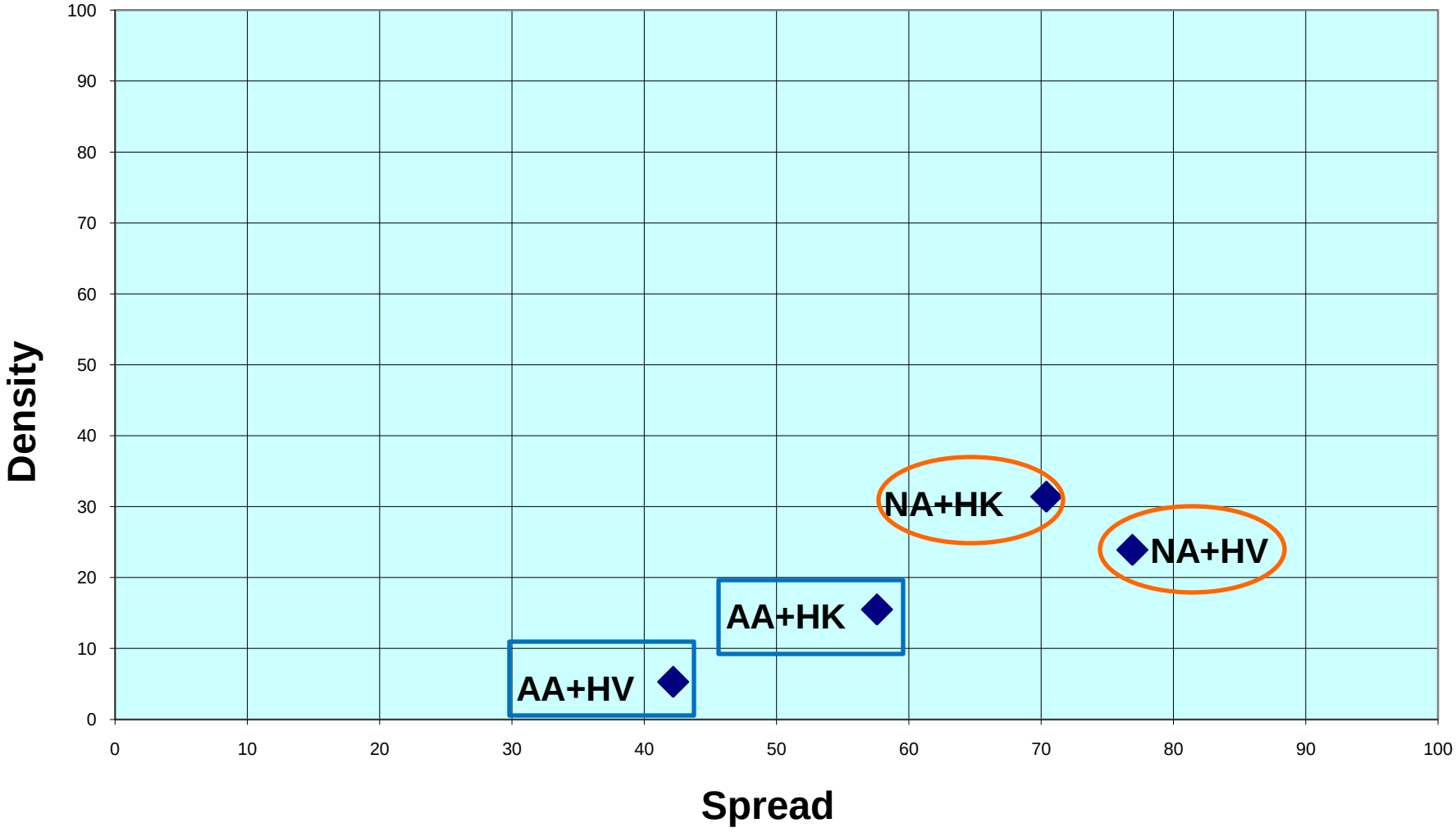
AA=Anglo, NA=non-Anglo, HK=Hackney, HV=Havering

Ethnicity*Residence: density and spread



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

Ethnicity*Residence: density and spread



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

a+vowel: Conclusions

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

- Reallocation of *a+vowel* use 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

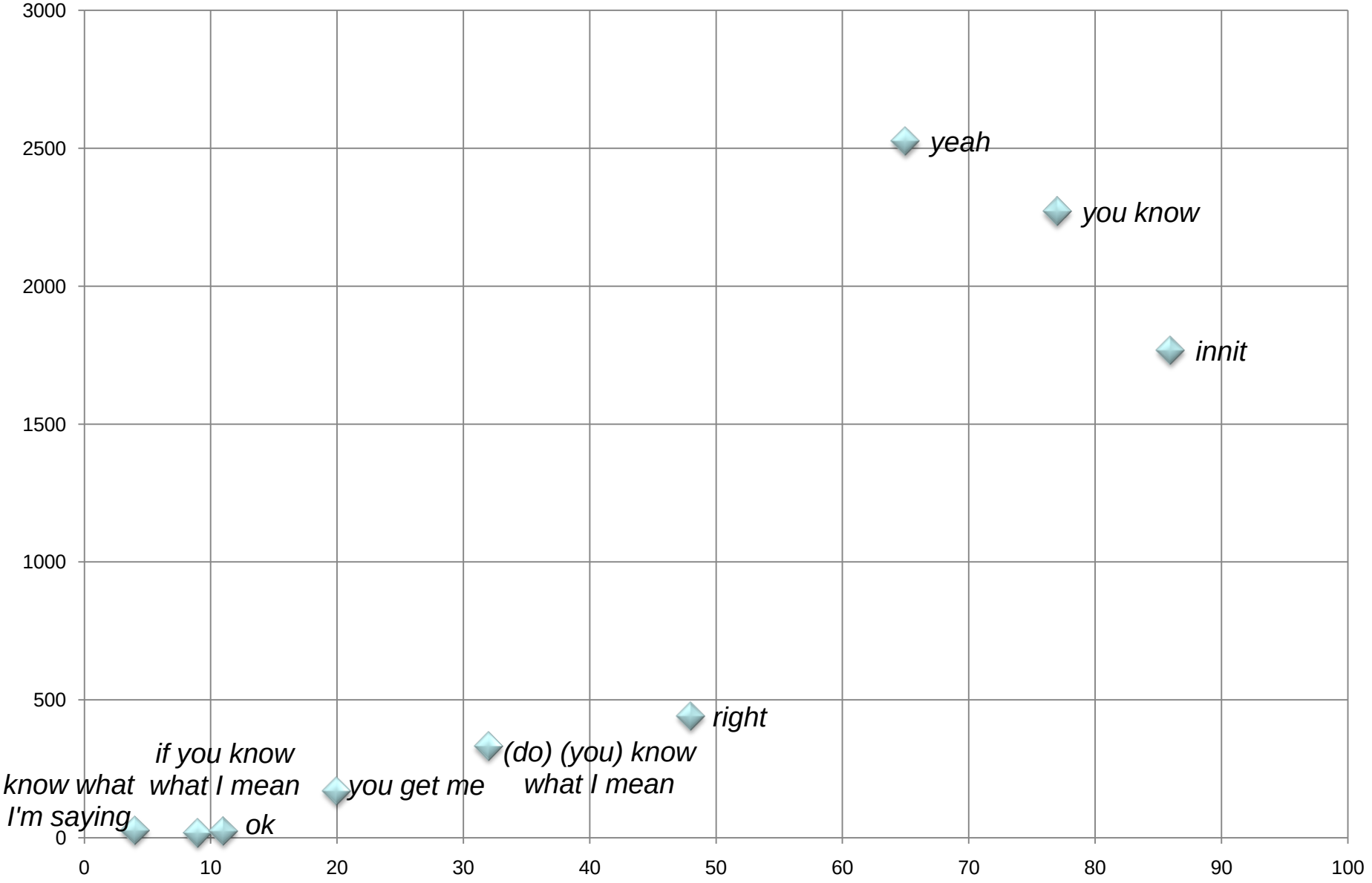
Invariant tags

Invariant tags examined

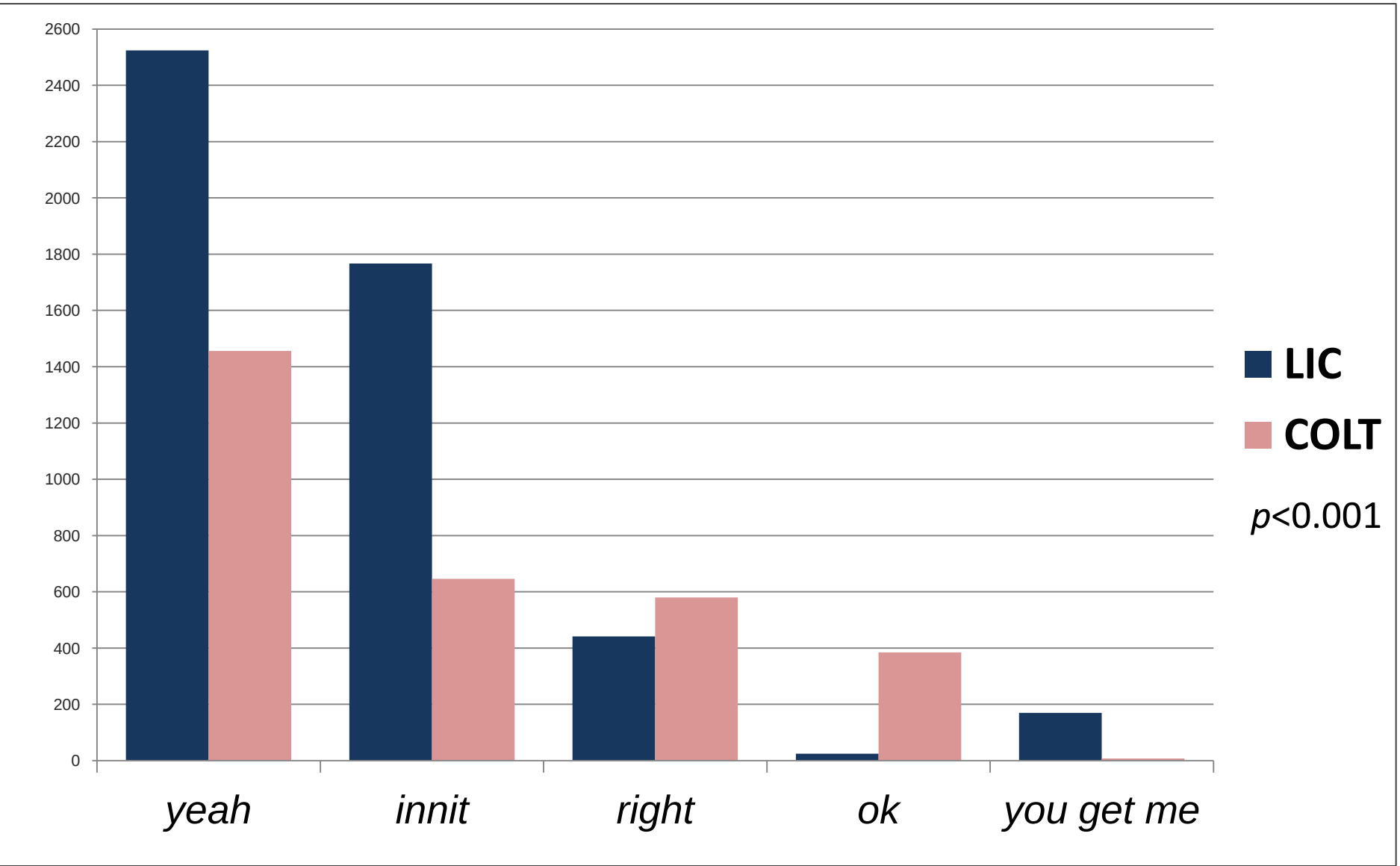
- Simple invariant tags
  - *innit, okay, right, yeah*

- Multi-word invariant tags
  - *You get me*
  - *You know*
  - *(Do)/(If) (you) know what I mean*
  - *(Do) (you) know what I'm saying*

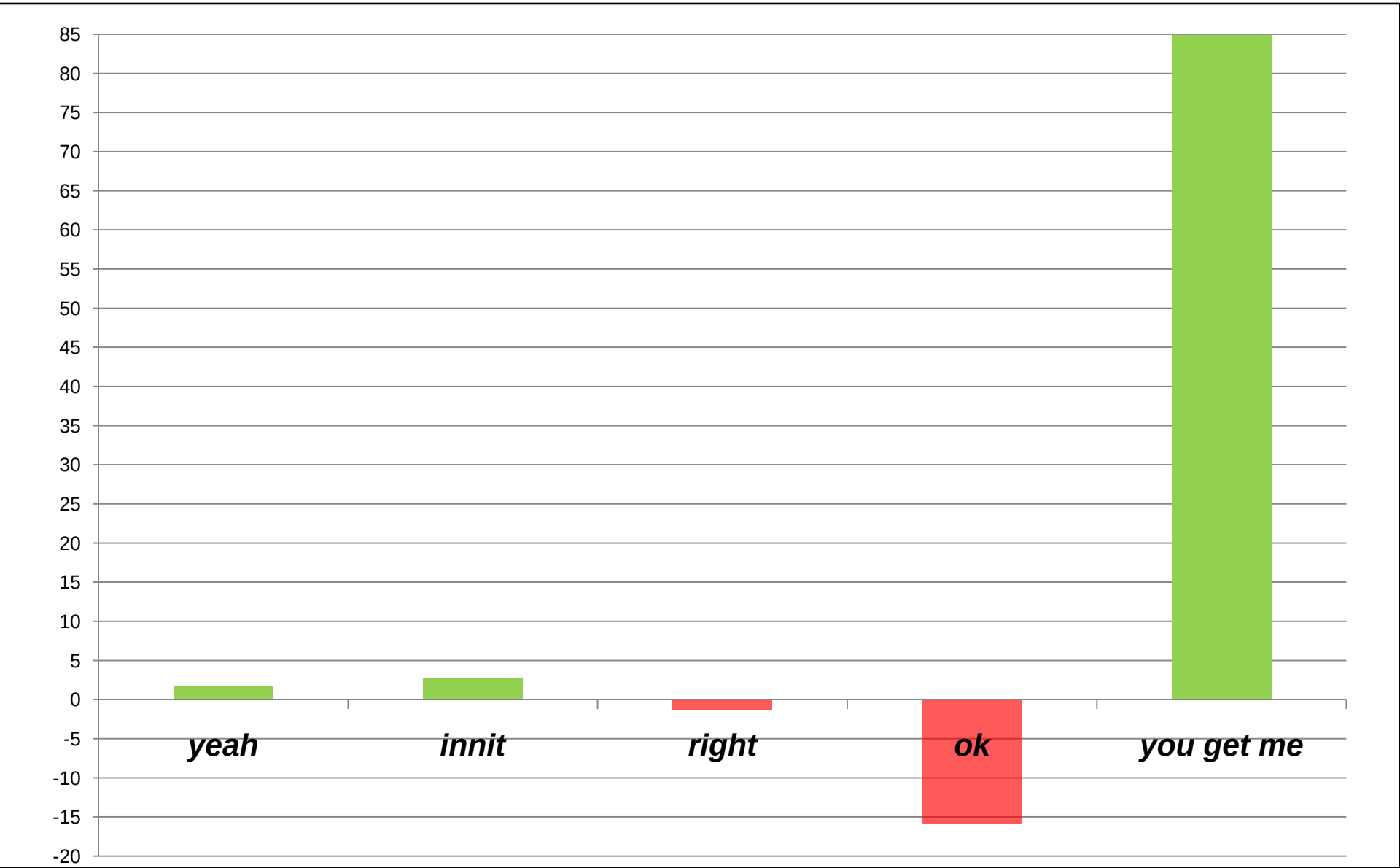
Young LIC speakers: frequency and spread



Young speakers in LIC and COLT: Frequency comparison (per mil. words)



Young speakers in LIC and COLT: Difference ratio LIC/COLT



- **Bold**: both frequency and spread differences are statistically significant.
- Normal: only frequency differences are statistically significant.
- ------ : both frequency and spread are comparable.

	Age	**Sex**	**Ethnicity**	**Residence**
innit	Young	Male	Non-Anglo	------
yeah	Young	------	Non-Anglo	------
you know	Old	Male	Anglo	------
ok	------	**Female**	Non-Anglo	------
right	Young	Female	Non-Anglo	Hackney
(do) (you) know what I mean	Young	Female	Anglo	Havering
if you know what I mean	Young	------	------	Havering
(do) you know what I'm saying	Young	Female	------	Havering
you get me	**Young**	------	**Non-Anglo**	**Hackney**

High-frequency tags

	Age	**Sex**	**Ethnicity**	**Residence**
innit	Young	Male	Non-Anglo	------
yeah	Young	------	Non-Anglo	------
you know	Old	Male	Anglo	------
ok	------	**Female**	Non-Anglo	------
right	Young	Female	Non-Anglo	Hackney
(do) (you) know what I mean	Young	Female	Anglo	Havering
if you know what I mean	Young	------	------	Havering
(do) you know what I'm saying	Young	Female	------	Havering
you get me	**Young**	------	**Non-Anglo**	**Hackney**

Low-frequency simple tags

	Age	**Sex**	**Ethnicity**	**Residence**
innit	Young	Male	Non-Anglo	------
yeah	Young	------	Non-Anglo	------
you know	Old	Male	Anglo	------
ok	------	**Female**	Non-Anglo	------
right	Young	Female	Non-Anglo	Hackney
(do) (you) know what I mean	Young	Female	Anglo	Havering
if you know what I mean	Young	------	------	Havering
(do) you know what I'm saying	Young	Female	------	Havering
you get me	**Young**	------	**Non-Anglo**	**Hackney**

Low-frequency multi-word tags

	Age	**Sex**	**Ethnicity**	**Residence**
innit	Young	Male	Non-Anglo	------
yeah	Young	------	Non-Anglo	------
you know	Old	Male	Anglo	------
ok	------	**Female**	Non-Anglo	------
right	Young	Female	Non-Anglo	Hackney
(do) (you) know what I mean	Young	Female	Anglo	Havering
if you know what I mean	Young	------	------	Havering
(do) you know what I'm saying	Young	Female	------	Havering
you get me	**Young**	------	**Non-Anglo**	**Hackney**

Emerging tag

	Age	Sex	Ethnicity	Residence
innit	Young	Male	Non-Anglo	------
yeah	Young	------	Non-Anglo	------
you know	Old	Male	Anglo	------
ok	------	**Female**	Non-Anglo	------
right	Young	Female	Non-Anglo	Hackney
(do) (you) know what I mean	Young	Female	Anglo	Havering
if you know what I mean	Young	------	------	Havering
(do) you know what I'm saying	Young	Female	------	Havering
you get me	**Young**	------	**Non-Anglo**	**Hackney**

Invariant tags: Conclusions

- Established invariant tags, irrespective of whether they are becoming more or less frequent, have a less marked ethnic distribution.

- Innovative (emerging) tags, like *you get me*, are currently used significantly more frequently within the multi-ethnic networks in which they have probably first emerged.