

# **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)

No. of words	1.4 million
No. of speakers	118
Data collection period	2005
Data collection method	Sociolinguistic interviews
Age	young=16-18 (average 17); 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)

(Stenström et al., 2002)

No. of words	500,000
No. of speakers	30 recruits + other speakers
Data collection period	1993
Data collection method	Self-recordings
Age	12-17 (average 14)
Sex	female; male
Ethnicity	White; ethnic minority
Residence	Inner London (Hackney, Tower Hamlets, Camden) Outer London (Barnet) Hertfordshire
Social class	Working class; Middle class

HERTFORDSHIRE



Barnet

ESSEX

Camden

Hackney

Havering

Tower  
Hamlets

# 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

	Expression	Utility
<b>Density</b> <i>(a+vowel)</i>	Frequency per 100 instances of vowel-initial words used with the indefinite article ( <i>a</i> or <i>an</i> )	Shows the relative frequency of <i>a+vowel</i> use, as opposed to <i>an+vowel</i>
<b>Frequency</b> (PMs)	Frequency of PM per million words	Shows relative frequency of pragmatic marker use
<b>Spread</b>	Number of users per 100 speakers	Shows the proportion of speakers using the feature.

***a+vowel***

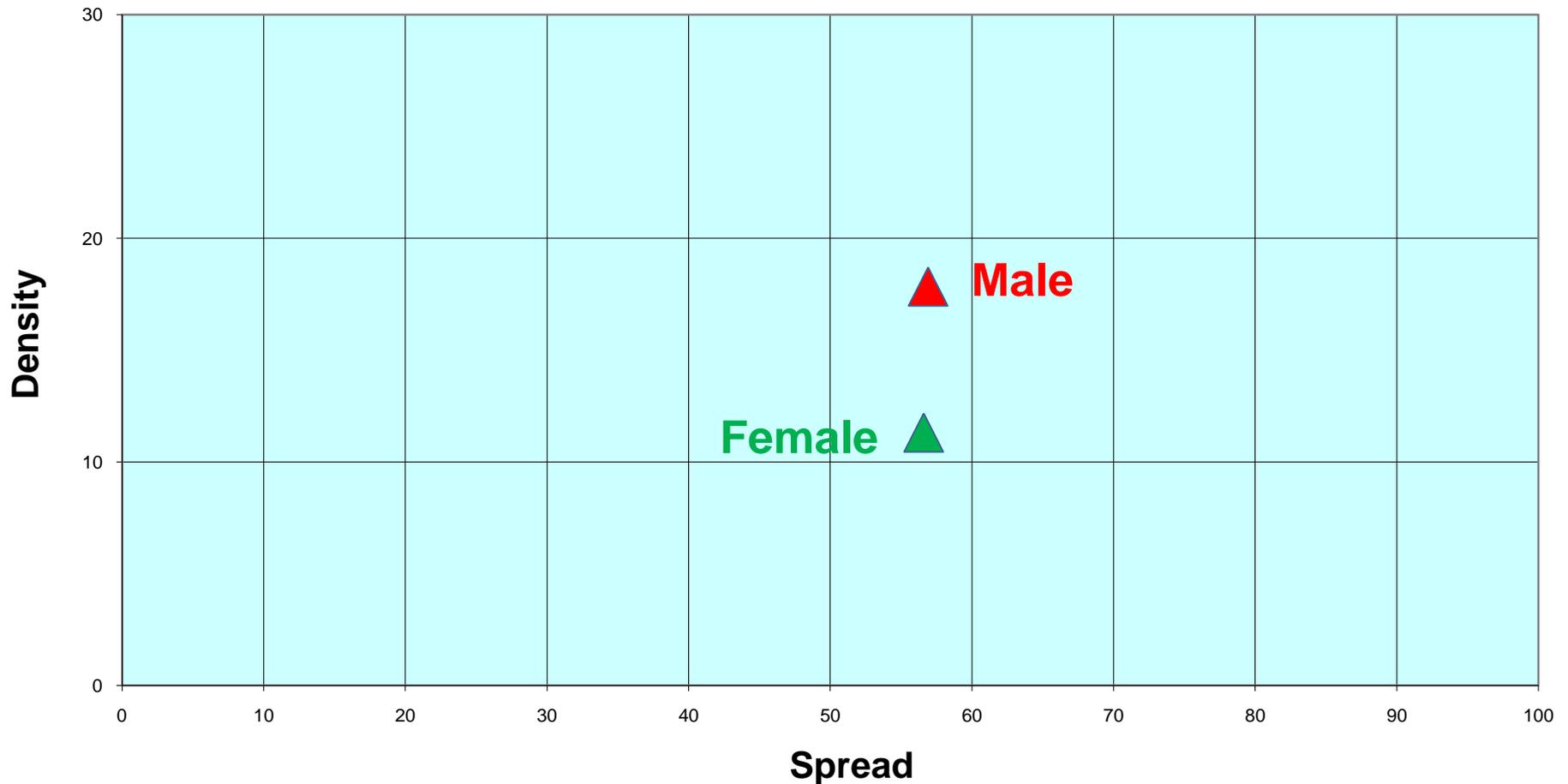
## *a+vowel*

### Predictive utility of social factors

	Coefficient	Odds ratio
Age (young)	1.404	<b>4.072</b>
Residence (Hackney)	0.954	<b>2.597</b>
Ethnicity (non-Anglo)	0.726	<b>2.067</b>
Sex (male)	0.605	<b>1.832</b>

- Results similar among young speakers.
- All results are highly significant:  $p < 0.001$ .

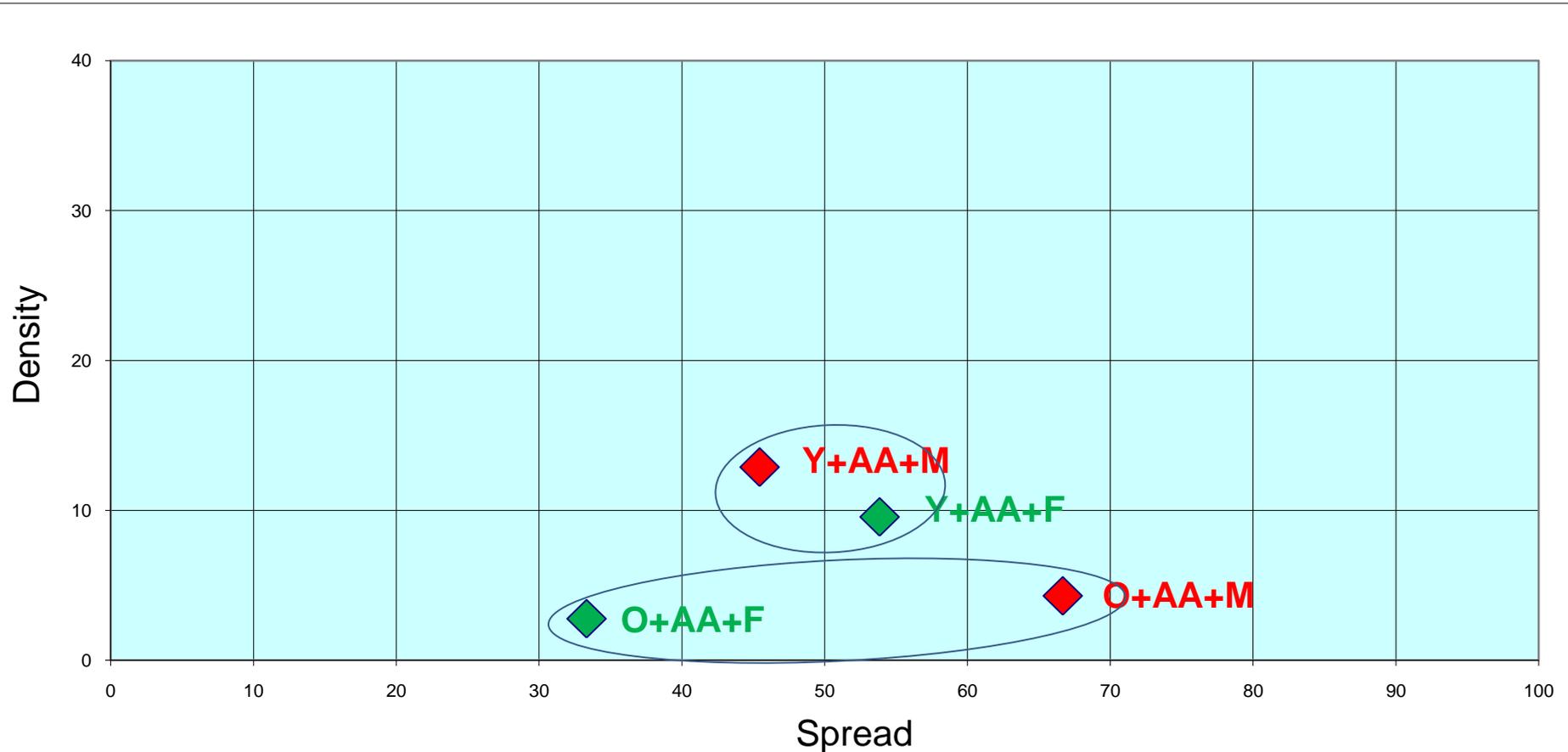
# LIC: Use of *a+vowel* (all speakers)



- ❖ 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

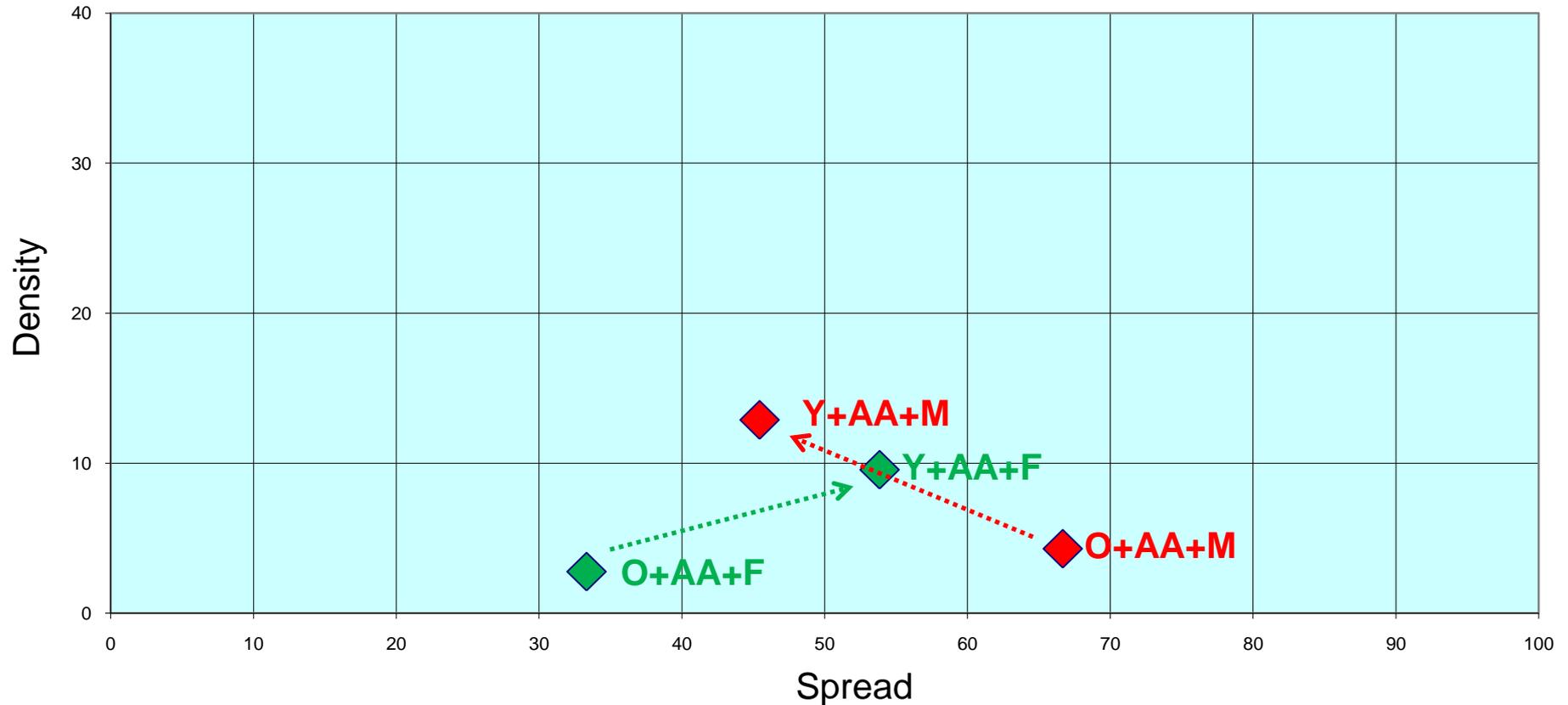


**Old male speakers** have comparable density, but higher spread.

**Young male speakers** have higher density, but lower spread.

# LIC: Use of *a+vowel* (Anglo speakers)

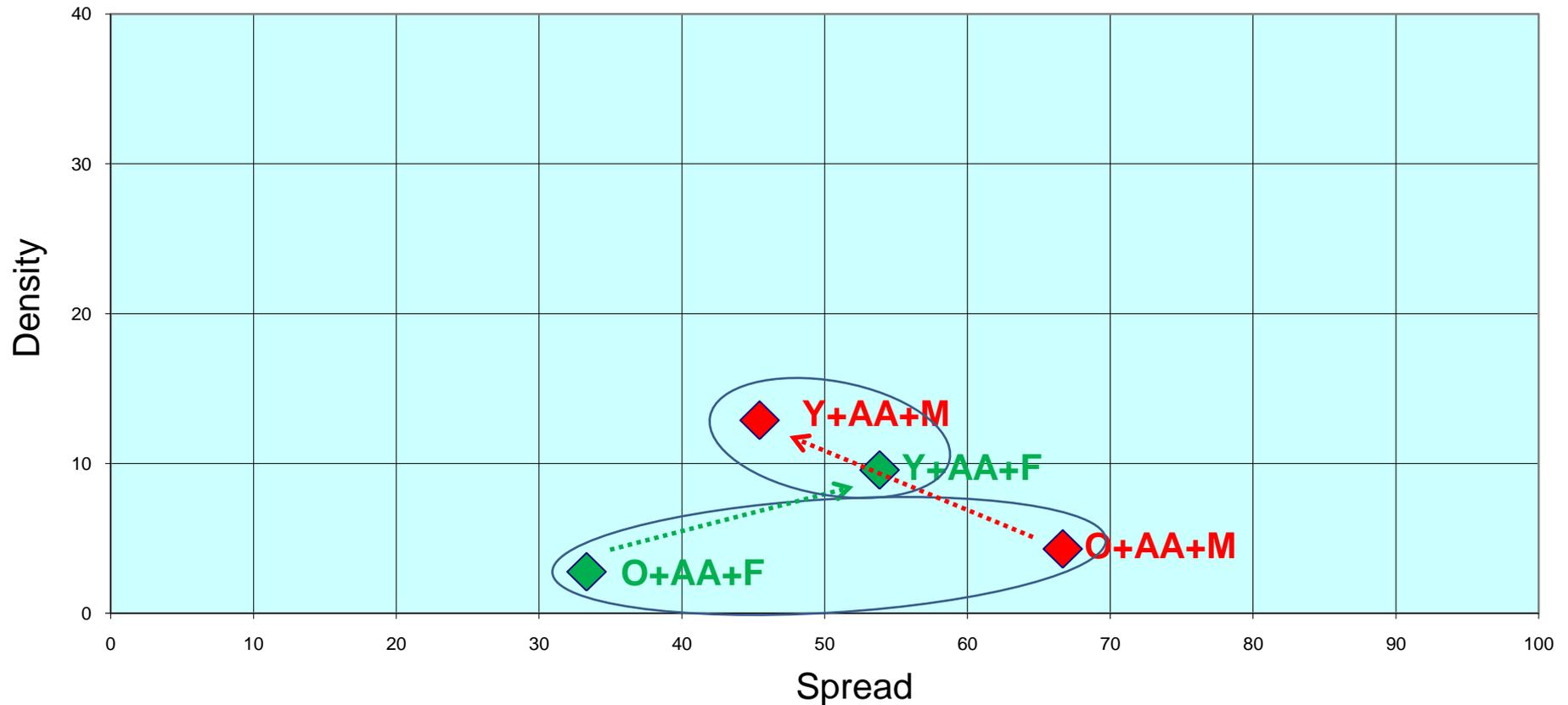
## Focus on sex



- ❖ **Young female speakers** have higher density and spread.
- ❖ **Young male speakers** have higher density, but lower spread.

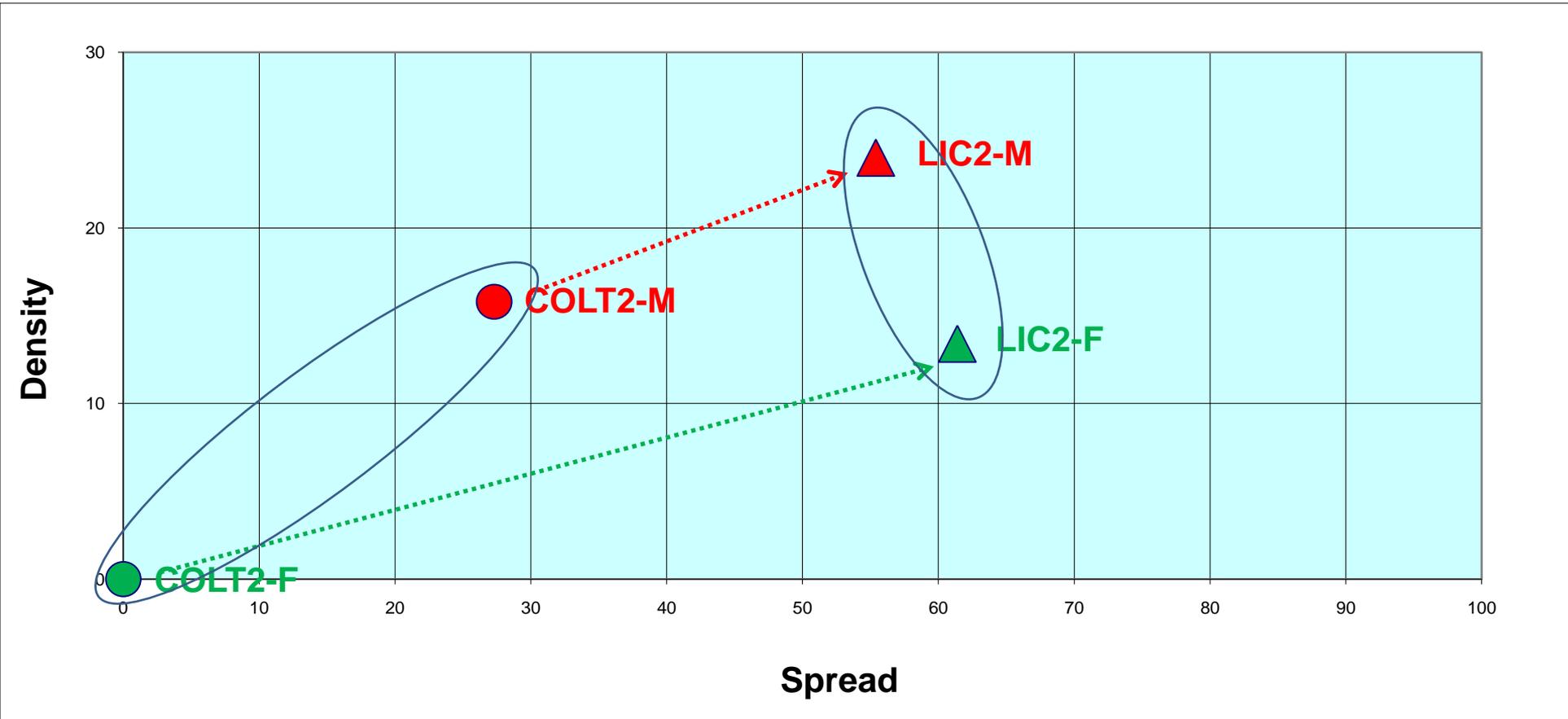
# LIC: Use of *a+vowel* (Anglo speakers)

## Focus on sex



**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)

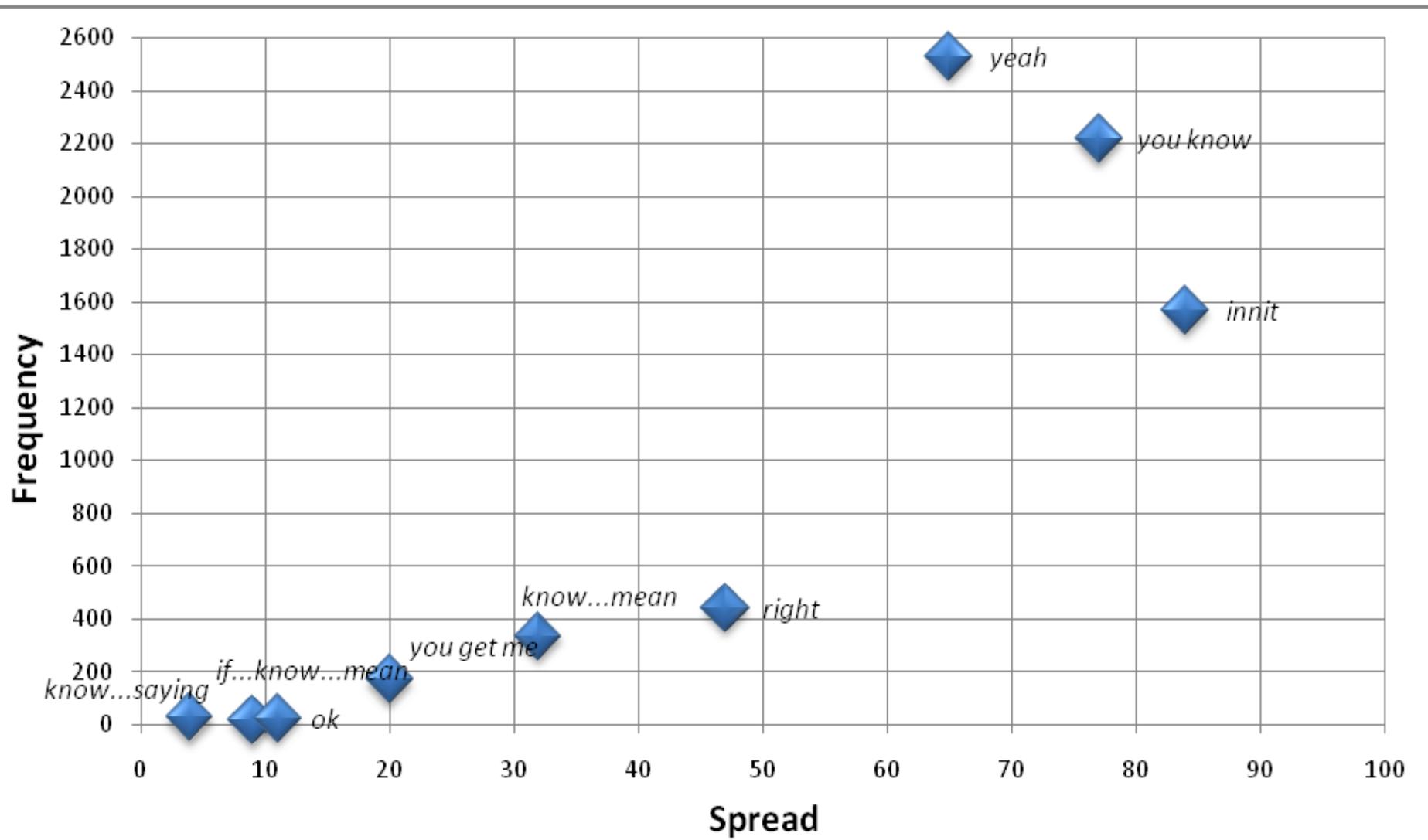


- ❖ There is considerable increase in both male and female LIC speakers.
- ❖ Increase more marked in female LIC speakers.

➔ **Male and female use seems to converge.**

# Pragmatic markers

# LIC-2: Extent of use of PMs



# 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%.

<b>Male</b>	<i><b>you get me</b></i> <i><b>innit</b></i> <i><b>yeah</b></i> <i><b>right</b></i>	<ul style="list-style-type: none"> <li>• <b>Emerging</b></li> <li>• <b>2/3 of high-frequency</b></li> <li>• <b>One low-frequency</b></li> </ul>
<b>Comparable use</b>	<i>you know</i> <i>if you know what I mean</i>	<ul style="list-style-type: none"> <li>• One high-frequency</li> <li>• One low-frequency</li> </ul>
<b>Female</b>	<b><u>ok</u></b> <b><u>(do) (you) know what I mean</u></b> <b><u>(do) you know what I'm saying</u></b>	<ul style="list-style-type: none"> <li>• <b>Only low frequency</b></li> </ul>

**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?

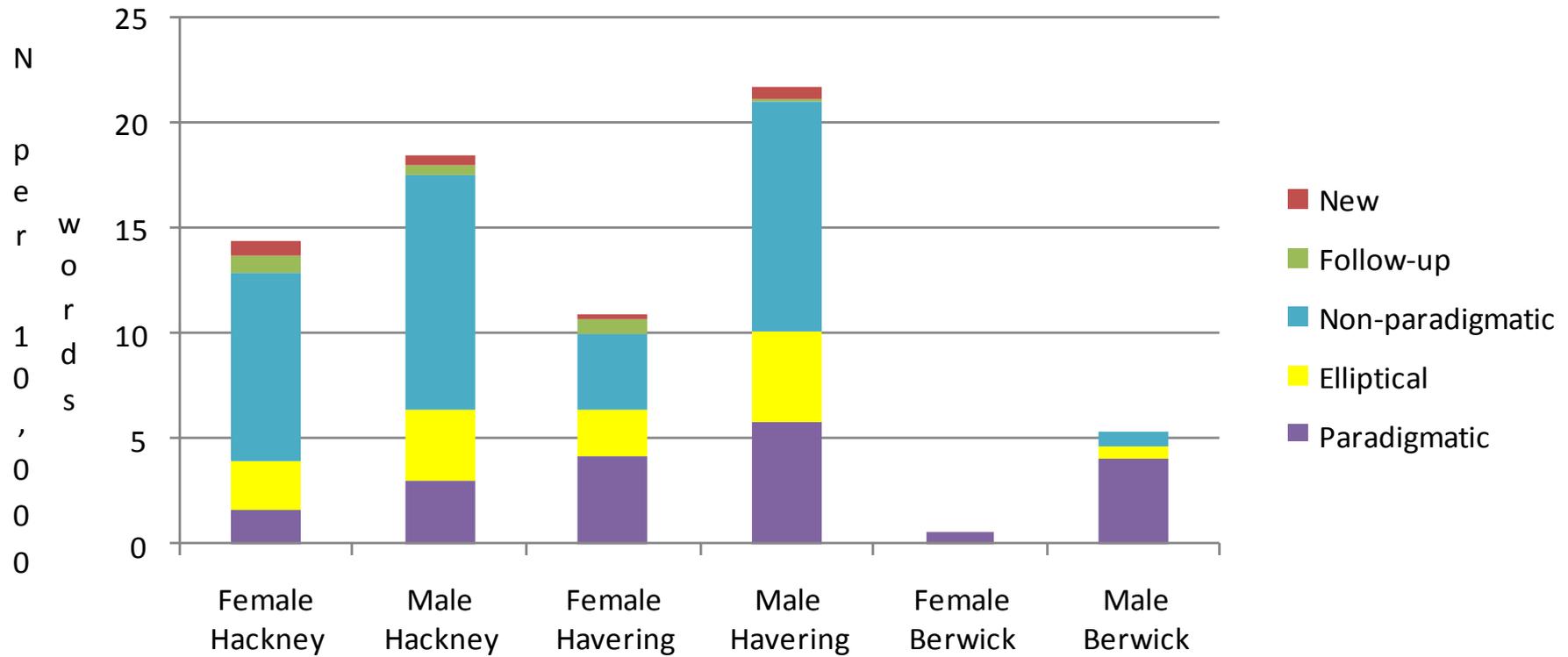
- After subordinate clauses

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

## *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 *innit* is used.

# References

- Andersen, G. (2001). *Pragmatic Markers and Sociolinguistic Variation*. Amsterdam: Benjamins.
- Cheshire, J. (1982). *Variation in an English Dialect: A sociolinguistic study*. Cambridge: Cambridge University Press.
- Gabrielatos, C., Torgersen, E., Hoffmann, S. & Fox, S. (in press). A corpus-based sociolinguistic study of indefinite article forms in London English. *Journal of English Linguistics*. Prepublished 18th February 2010, doi:10.1177/0075424209352729.
- Pichler, H. (2008). *A qualitative-quantitative analysis of negative auxiliaries in a northern English dialect: I DON'T KNOW and I DON'T THINK, innit?* Unpublished PhD Dissertation, University of Aberdeen.
- Pichler H. & Torgersen E. (2009). The sociolinguistic and functional distribution of *innit* in Berwick English and Multicultural London English. Paper presented at UKLVC7, University of Newcastle.
- Stenström, A-B., Andersen, G. & Hasund, I.K. (2002). *Trends in Teenage Talk*. Amsterdam: Benjamins.
- Torgersen, E., Gabrielatos, C., Hoffmann, S. & Fox, S. (under review). A corpus-based study of pragmatic markers in London English. *Corpus Linguistics and Linguistic Theory*.