

Linguistic Variation

Seminar 'Corpus Linguistics'

Quirin Würschinger, LMU Munich

July 10, 2025

Session Overview

- Research projects group work
- Linguistic variation in the use of tag questions
 - theory: Tottie and Hoffmann (2006)
 - ⇒ structure
 - ⇒ variation between BrE and AmE
 - ⇒ social variation (gender, age, social class)
 - practice: BNC 2014 Spoken corpus on SkE

Workshop Projects

- group task: draft or refine a mini study plan and add one slide to the shared deck
- shared deck:
<https://1drv.ms/p/c/9a2ec97d593520f9/EUu0cQtCO1pMnlyH0FZKnRMBF>
- template fields to complete:
 - research question and hypothesis
 - data and methods
 - expected results
- plenary: each group gets ~60 s to pitch with short Q&A

Tag Questions

What are tag questions?

- consist of an anchor clause + a tag
- subject in tag must be pronoun, *there*, or *one*
- verb in tag must be auxiliary or modal

	Anchor	Question Tag	Polarity
(1)	Makes you really think,	doesn't it. (LSAC) ²	Positive–Negative (+/–)
(2)	Oh it's not very valuable	is it? (BNC-S)	Negative–Positive (–/+)
(3)	So this is the letter he sent you	is it? (LSAC)	Positive–Positive (+/+)
(4)	Yes, they don't come cheap	don't they? (BNC-S)	Negative–Negative (–/–)

Polarity Patterns

- reversed polarity most common
 - + -: *It's hot today, isn't it?*
 - - +: *It isn't working, is it?*
- constant positive and constant negative possible but rare
 - + +: *She can drive, can she?*
 - - -: *He can't swim, can't he?*

Study background

Tottie, Gunnel, and Sebastian Hoffmann. 2006. 'Tag Questions in British and American English'. *Journal of English Linguistics* 34 (4): 283–311.

<https://doi.org/10.1177/0075424206294369>.

- compares tag questions in British (BNC-S) and American (LSAC) spoken corpora
- examines form, polarity and pragmatic functions
- explores sociolinguistic factors: gender, age, social class
- finds tags about three times more frequent in British English, reversed polarity dominant ($\approx 75\%$)
- women use tags more frequently and for facilitative purposes

Pragmatic Functions

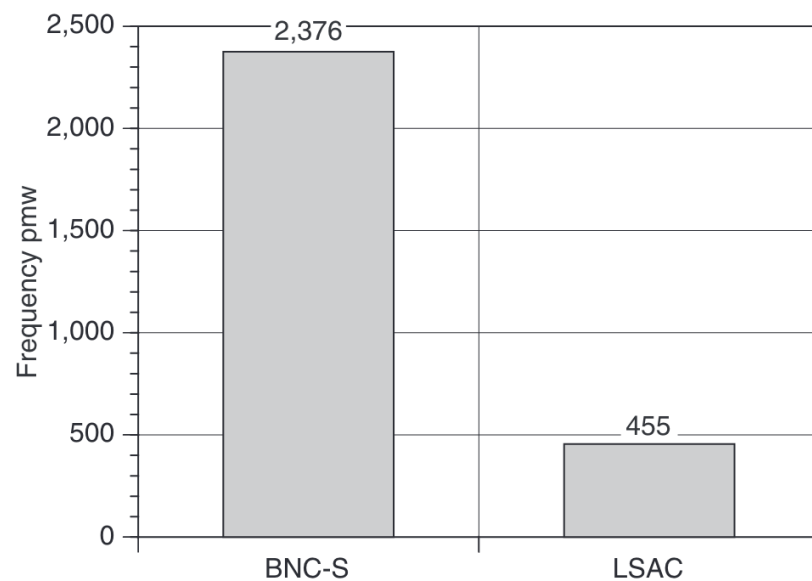
Table 6
Classifications of Pragmatic Functions of Tags

Macro Category	Holmes	Algeo	The Present Work
Epistemic modal	Modal	Informational Confirmatory	Informational Confirmatory
Affective	Facilitative Softening	Confirmatory	Facilitating
	Challenging	Punctuational Peremptory Aggressive, antagonistic	Attitudinal Peremptory Aggressive

- Informational
 - *You live in London, don't you?*
- Confirmatory
 - *This report was sent yesterday, wasn't it?*
- Facilitating
 - *Let's move on, shall we?*
- Attitudinal
 - *That was a fantastic game, wasn't it?*
- Peremptory
 - *Close the window, will you?*
- Aggressive
 - *You're going to mess this up again, aren't you?*

Variation Across British and American English

Figure 1
The Frequency of Tag Questions in British and American English, Calculated Per Million Words



Note: BNC-S = the spoken component of the British National Corpus; LSAC = the Longman Spoken American Corpus.

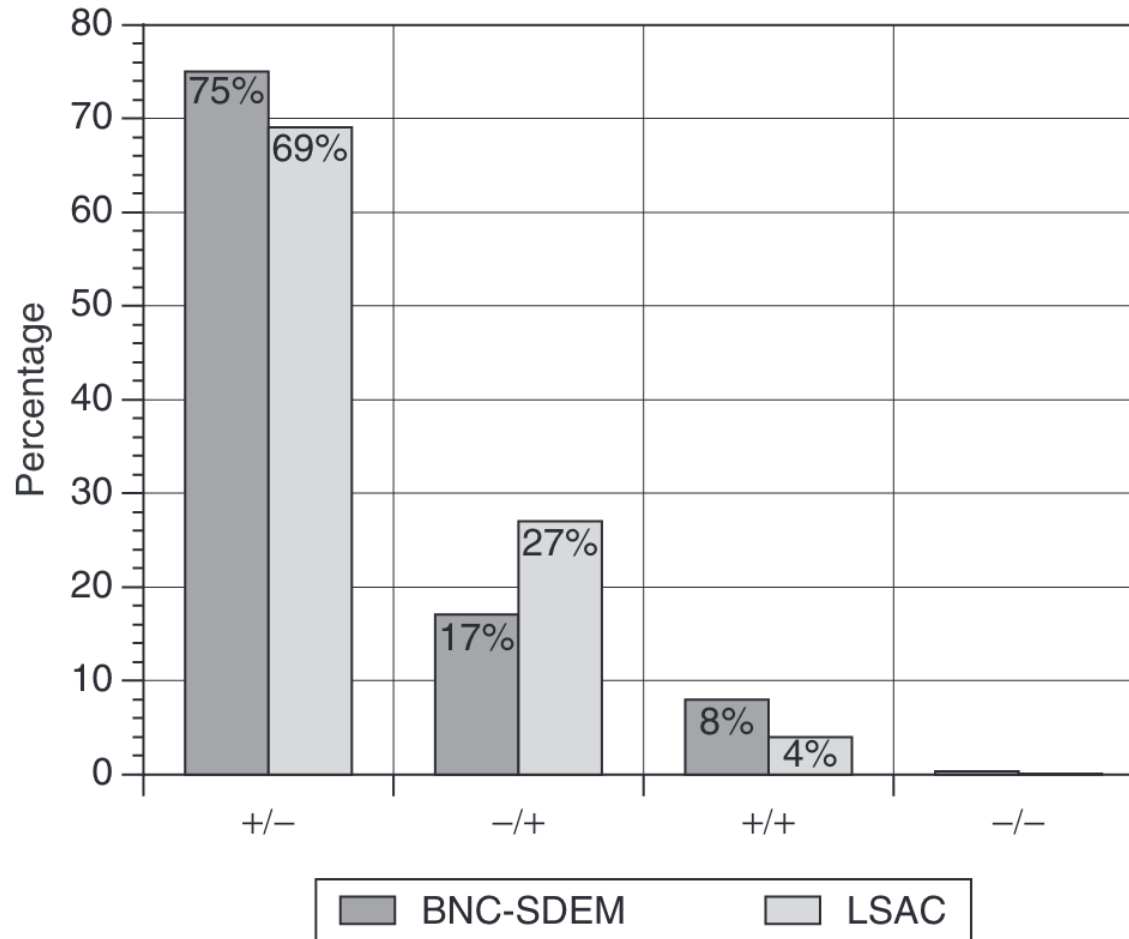
BNC-S

the spoken demographic subpart of the British National Corpus

LSAC

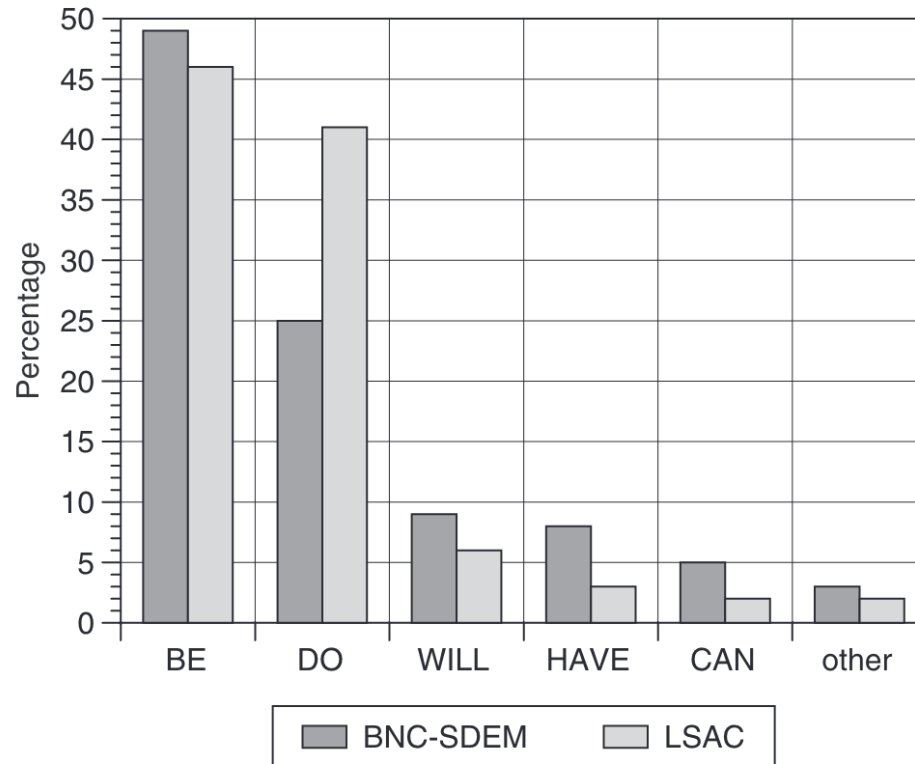
the Longman Spoken American Corpus

Figure 3
Proportional Distribution of Polarity in Two Subsets of
1,000 Words Each from BNC-SDEM and LSAC



Auxiliary Choice

Figure 4
Proportions of Auxiliary Use in Question Tags in BNC-SDEM and LSAC



- modal auxiliaries dominate (especially *will*, *would*)
- regional differences in choice of *be* vs *do*

Pronouns and Phrases

Figure 5
Proportions of Pronouns in Question Tags in BNC-SDEM and LSAC

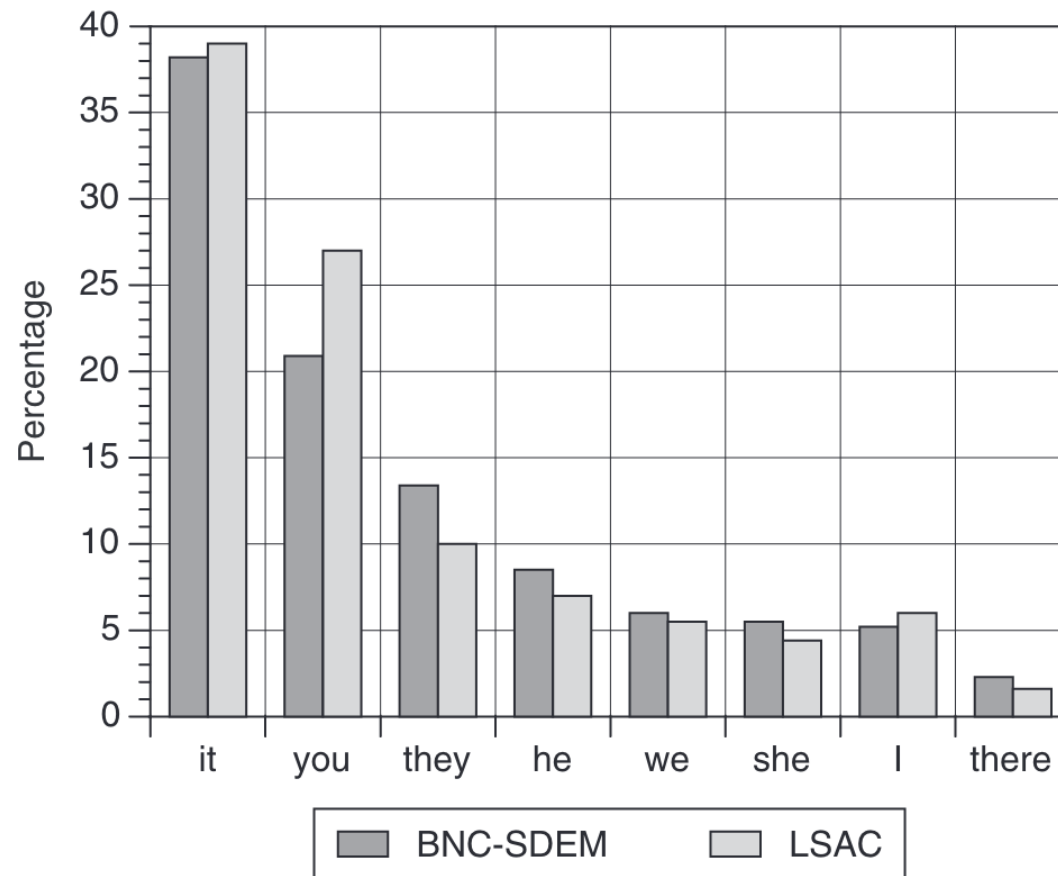


Table 5¹¹
The Fifteen Most Frequent Question Tags in BNC-SDEM and LSAC

Tag	BNC-SDEM			LSAC		
	<i>N</i> = 3,724	%	Rank	<i>N</i> = 2,311	%	Rank
<i>isn't it?</i>	760 ^a	20.4	1	429	18.6	1
<i>is it?</i>	227	6.1	2	115	5	4
<i>aren't they?</i>	133	3.6	3	65	2.8	9
<i>don't you?</i>	99	2.7	4	124	5.4	2
<i>do you?</i>	89	2.4	5	123	5.3	3
<i>don't they?</i>	88	2.4	6	55	2.4	11
<i>aren't you?</i>	82	2.2	7	56	2.4	10
<i>wasn't it?</i>	76	2	8	76	3.3	6
<i>haven't you?</i>	69	1.9	9	21	0.9	25
<i>are you?</i>	63	1.7	10	52	2.3	12
<i>weren't it</i>	62	1.7	11	—	—	—
<i>didn't you?</i>	61	1.6	12	70	3	7
<i>isn't he?</i>	57	1.5	13	43	1.9	13
<i>didn't he?</i>	52	1.4	14	24	1	21
<i>doesn't it?</i>	52	1.4	15	101	4.4	5

Note: BNC-SDEM = the spoken demographic subpart of the British National Corpus; LSAC = the Longman Spoken American Corpus.

a. This figure includes 319 instances of *innit* that are not clearly invariant.

Sociolinguistic Factors

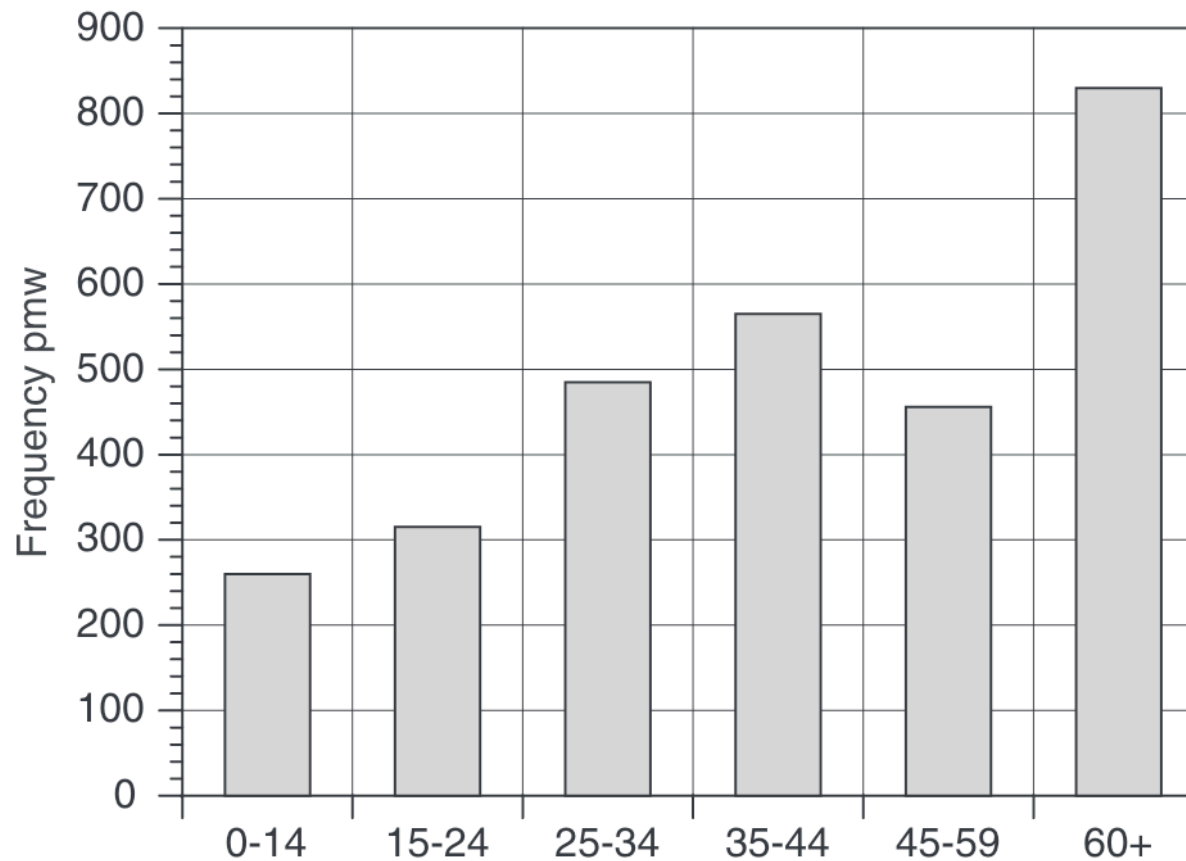
Gender Comparison (Women vs Men)

Table 8
The Use of Tag Questions By Men and Women in BNC-SDEM and LSAC

	BNC-SDEM		LSAC	
	<i>N</i>	pmw	<i>N</i>	pmw
Male	6,081	4,209	680	453
Female	10,454	4,647	1,067	489

Note: BNC-SDEM = the spoken demographic subpart of the British National Corpus; LSAC = the Longman Spoken American Corpus.

Figure 8
The Use of Tag Questions in Different Age Groups in LSAC



Interim Summary

- reversed polarity tags most frequent
- significant BE/AE contrasts
- gender & age influence usage
- form–function mapping varies with context

Practice: Corpus Study with Sketch Engine

Corpus: BNC 2014 Spoken

Objective: replicate and extend Tottie's analysis on present-day data

Study Questions

About the structure of tag questions:

1. How many false positives did your query return?
2. Which question tags are most frequent (e.g. *is n't it*, *are n't you*, etc.)?
3. Which pronouns are most frequent?
4. Which verbs are most frequent?
5. Are negated or positive tags more frequent?

About their use in different contexts:

6. How do tag questions differ in different age groups?
7. How do tag questions differ in different genders?
8. How do tag questions differ in different social classes?
9. How do tag questions differ in different educational levels?

Further study:

10. Which polarity patterns are most frequent? (e.g. + - vs - +)


Overview of steps

- Retrieve tag question attestations
- Evaluate false positives
- Frequency of tag forms, pronouns & verbs
- Distribution across metadata





1: Retrieve Attestations

CQL query:

1: [lemma="be|do|have" | tag="MD"] 2: [lemma="not"]? 3: [tag="PP.?"] [word="\?"]

 SKETCH ENGINE

CQL builder

CQL: 1: [lemma="be|do|have" | tag="MD"] 2: [lemma="not"] ? 3: [tag="PP.?"] [word="\?"]

normal token

+

1: [lemma="be|do|have" | tag="MD"]

+

normal token

2: [lemma="not"] ?

+

normal token

3: [tag="PP.?"]


+


normal token


[word="\?"]

+

|



result example 

USE THIS CQL 

oh how d'you manage that? lid in the oh it's alright it's like sealed </unclea>
really want to yeah we could I brought you some wine yeah think we drank it all though
rich truffles mm nice that is nice wine as well thank you very much it's quite nice
g corks giant bung things well I'm sure there'll be more corks though the the cork's like the vulnerable thing
eah yeah I can have a shower it's just fine but a bath's just so much more it's a bit more of a project as well
ce you've got it like you've gotta be s stay in there for a while and that's a bit weird round someone's house
' acceptable yeah and the way to do it culturally in Japan yeah and like hot tubs are just baths with bubbles
then it's just a bath with your friends and yeah but that's normally their swimming costume and stuff though
and the sort of tepid water outdoors it they're normally like you'd be in a erm like a swimsuit for that
you do depends </unclea> I wouldn't mind if I would but like they wouldn't be watching us anyway

is it ?
did n't we ?
is n't it ?
is n't it ?
is n't it ?
is n't it ?
are n't they ?
is n't it ?
would n't you ?
would they ?

yeah the lid like it's vacuum sealed on so you have to ping it off I think I don't really know I did the vacuum s
mm think we anyway we made our way through I think that was part of the problem the
mm mm mm </unclea> Chile chile so what was the conclusion with the wine thing should you pour it?
that's like the thing mm I suppose that's the bit yeah that so the smaller that is the better in a way mm mm r
and it uses lots of water and you probably wanna s yeah once you've got it like you've gotta be s stay in the
it is yeah mm yeah though if they're not there it's not so weird but also but then that's also a bit weird too lik
mm mm be more than happy to go in those I think though like if you're ever in a hot tub with other people if
yeah gets weird really quickly I would imagine and the sort of tepid water outdoors it they're
usually yeah yeah whereas like an outdoor bath thing I would go in naked not if your mates are there to be
I hope not that would be quite weird if they were I hope not they'd be standing at the

Hints

- You can exclude utterance-initial hits with `<u> []{1,} ... within <u/>`.
- You can exclude preceding *wh*-words with `[lemma!="where|who|which|when"]`.
- You can exclude cases where a verb follows the pronoun immediately by adding `[tag!="V.*"]` after the pronoun in your pattern.
- You can exclude cases where an adjective follows the pronoun by adding `[tag!="J.*"]` after the pronoun.

Step 2: Evaluate False Positives

- download a random sample of 50 hits
- annotate in Excel (**Label** column: 0 = false positive, 1 = true positive)
- refine the CQL query or use Excel filters to reduce noise.
- Model sheet (template):

<https://1drv.ms/x/s!AvkgNVI9yS6aokmqDbTz5BmfbU6C>

Step 3: Frequency of Tag Forms

- most frequent complete tag phrases (*is n't it, are n't you, etc.*)
- most frequent pronouns
- most frequent verbs
- export frequency tables to Excel for charts

Step 4: Distribution across Metadata

cross-tabulate frequencies by

- `Age range`
- `Gender`
- `Class: Social grade`
- `Highest qualification`

Summary & Discussion

- compare BNC 2014 findings with Tottie and Hoffmann (2006)
- reflect on methodological challenges (false positives, metadata)
- discuss potential research extensions

References

Tottie, Gunnel, and Sebastian Hoffmann. 2006. "Tag Questions in British and American English." *Journal of English Linguistics* 34 (4): 283–311. <https://doi.org/10.1177/0075424206294369>.