

Studying TV narratives and characters

A corpus linguistic approach

Monika Bednarek
Department of Linguistics



THE UNIVERSITY OF
SYDNEY

Sydney 
Corpus Lab

Aims

- › Television dialogue as media language (media linguistics)
 - to develop a new categorisation of the multiple functions of TV dialogue;
- › Television dialogue as a language variety
 - to identify and explain the salient linguistic characteristics of TV dialogue (inductive, corpus-driven);
- › Television dialogue and society (sociolinguistics)
 - to examine noncodified and nonstandard language phenomena in TV dialogue;
- › Television dialogue as a situated practice (production, product, consumption)
 - to provide new insights into production and consumption aspects of TV series, and to connect these to the linguistic analysis.

Outline

- › Data and approach
- › Case study of *ain't*
- › Insights from screenwriters and the audience
- › Conclusion

In brief

- › ‘non-standard’ language (broad definition)
 - covering aspects of language that are marked for divergence from (written) standards that have influenced traditional broadcast speech.
 - regional, vernacular, and stigmatised varieties are ‘non-standard’
- › Characterisation
 - ‘how writers imbue the “people” in their texts with certain characteristics, personalities or identities – how characters are construed in discourse or how readers infer certain characteristics from discourse’ (Bednarek 2010: 98)

Balance and representativeness

- › Not opportunistic ('black box')
- › Not just one or a few series
- › Not just one genre (e.g. sitcom, soap opera)
- › Not just pilot episodes
- › Representing on-screen dialogue (what viewers hear)
- › Accurately capturing interesting linguistic features relevant to characterisation (e.g. differences between *goin'* and *going*; *gonna* and *going to*, etc)
- › Big data?

Issues

- › balance (TV series, pilot episodes)
- › version? (shooting script?)
- › changes during production process
- › performance features (Taylor 2004)
- › The ‘definite script [representing what is actually uttered on screen] ... can only be obtained by transcribing the film’ (Bonsignori 2009: 187)
- › While ‘scripts are a genre in their own right, they are in fact inappropriate for investigations on real movie conversation’ (Forchini 2012: p. 31)

5 INT. KITCHEN. DAY.

Daisy the scullery maid is raking out the clinker. Mrs Patmore, the cook, comes in and ties on her apron.*

MRS PATMORE: Is your fire still in?

DAISY: Yes, Mrs Patmore.

Example: The TV corpus (<https://www.english-corpora.org/tv/>)

- › Subtitles; no speaker names

trying. Sure not going to tell anybody I was there. Well, I **ai n't** ashamed of nothing I did over there. Can't say that I'm
friends on the way with a special package. Oh. This " package " **ai n't** the breathing kind, is it? Invisible ink on sheet music I can
her, she hugged me back? like the sailor said, quote? " **ai n't** that a hole in the boat? " ? I've sunshine enough to
. What, you slow or something? What's up? Nah, I **ai n't** slow, bro. Go get your own ball. - Yo, man
's what you want. Well, yeah, but where's Loretta? She **ai n't** comin'. Yeah, but she said that... She was wrong.
. You here to charter a plane? Cos the place is closed. * **Ai n't** nobody's business if I do... * Looks like you take good care

- › Hey, dipshit, can't you see we need the ball? - This ball? - Yeah, that ball.
What, you slow or something? What's up? Nah, I **ai n't** slow, bro. Go get
your own ball. - Yo, man. - What? What's your problem, son? - What's your
problem? - You my problem. You need to learn how to speak to people.
Oh, you need to have respect, son.

Issues

- › Often exclude names of speakers
- › Variations in quality (Bywood et al 2013: 598)
- › Not identical to on-screen dialogue

```
30
00:01:38,104 --> 00:01:41,806
I don't care anymore. I'm
done! I'm gonna kill him.

31
00:01:43,675 --> 00:01:44,938
You!

32
00:01:45,104 --> 00:01:46,481
(WHISPERS) No.
```

**Inaccuracies
in subtitles**

Actual wording

Subtitles

| | |
|----------------------|-----------------------------------|
| <i>do we</i> | <i>are we gonna</i> |
| <i>clean</i> | <i>clean out</i> |
| <i>do</i> | <i>go through</i> |
| <i>you</i> | <i>do you</i> |
| <i>bring that up</i> | <i>bring Skippy up</i> |
| <i>nothing</i> | <i>nothing happened to Skippy</i> |
| <i>'d</i> | <i>would</i> |
| <i>this goes on</i> | <i>this keeps going on</i> |
| <i>get</i> | <i>bring me</i> |
| ∅ | <i>hi (2x)</i> |
| ∅ | <i>oh (5x)</i> |
| ∅ | <i>hey (2x)</i> |
| ∅ | <i>cause</i> |
| ∅ | <i>so</i> |
| ∅ | <i>would</i> |
| ∅ | <i>bye</i> |
| ∅ | <i>and (2x)</i> |
| ∅ | <i>look</i> |
| ∅ | <i>well</i> |

(From Bednarek 2010,
Gilmore Girls 1.11)

Transcripts produced by fans of specific series

- › Widely used (e.g. Baker 2005, Quaglio 2009, Bednarek 2010)
- › ‘fairly accurate and very detailed, including several features that scripts are not likely to present: hesitations, pauses, repeats, and contractions’ (Quaglio 2008: 191-92)
- › Can be more accurate than subtitles (Bednarek 2010)

| Inaccuracies in fan transcript | Actual wording |
|--------------------------------|----------------|
|--------------------------------|----------------|

cleans out
compare to
towards
bit
so

has to clean out
compared to
toward
bite
and

Issues

- › one or many transcribers? L1 transcriber?
- › even when 99.5% accurate (Bednarek 2012), issues around standardisation (e.g. *gonna*, *wanna*)
- › balance? (TV series that attract fan base)

The Sydney Corpus of Television Dialogue

- › Small, specialized corpus representative of fictional dialogue in US TV series (targeted at adults)
- › Transcripts of one episode each from 66 different series produced since 2000
- › Mostly transcribed from scratch
- › On-screen, audio dialogue (incl. monologue, two-party or multi-party conversation and voice-over, but not screen directions, etc)
- › Award-winning/-nominated and ‘mainstream’
- › Comedy and drama genres
- › Types of episodes
- › ~275,000 words; companion website at www.syd-tv.com

Summary

- › mainly orthographic (more data in less time)
- › marked pronunciation variants, e.g. *gonna* ('going to'), *c'mon* ('come on'), use of the alveolar form /lⁿ/ in words ending in *-ing* (e.g., *somethin'*)
- › contractions (e.g. *should've*), discourse markers (e.g., *oh*), hesitation markers (e.g., *uh*), listening cues (e.g. *mmm*), dis/agreement markers (e.g., *uh-uh*), interjections (e.g., *ugh*)
- › repeats (e.g. *l l l* or *l, l, l*), interruptions
- › voice-over dialogue (V; VOICE); quoted/read speech ("...")
- › audible dialogue from media (e.g. radio, television)
- › punctuation not consistently used to identify aspects such as intonation or speed of delivery

Corpus-based critical sociolinguistics

- › Concepts/theories/research from (critical) sociolinguistics
- › Corpus linguistic techniques

Linguicism

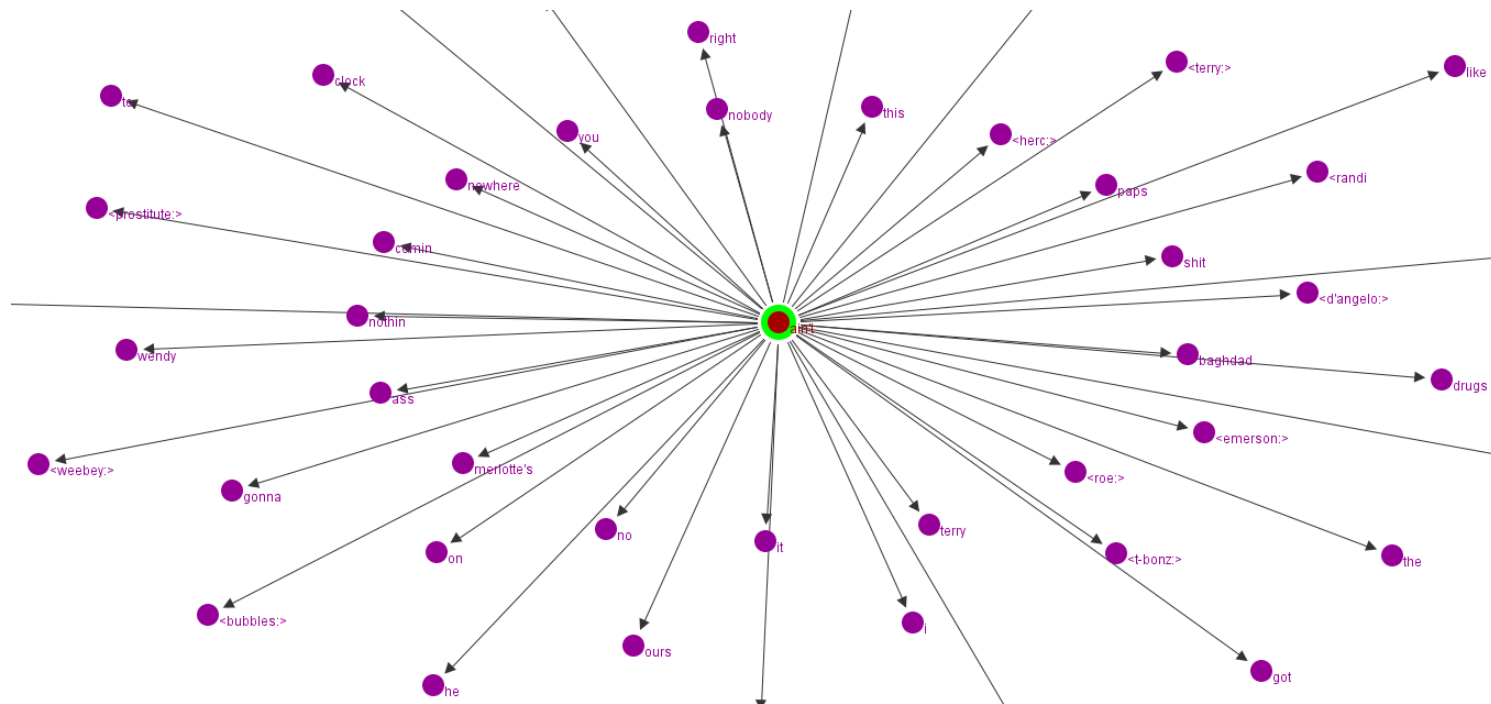
- › Linguistic discrimination, linguistic stereotypes, ‘standard language ideology’ (Lippi-Green 2012)
- › Language variation in animated children’s Disney movies (Lippi-Green 2012)
- › Stereotyped/exaggerated realisations = stylisation (e.g. Androutsopoulos 2012a: 151)
- › Negative representations and stereotypes
 - Non-native speakers (Bleichenbacher 2008, 2012), African American (Green 2002), Native American (Meek 2006, Buscombe 2013), Asian (Lippi-Green 2012: 287, Chung 2013), Latinx (Penfield and Ornstein-Galicia 1985), Irish (Walshe 2011), Southern American (Mitchell 2015), ‘problematic’ identities such as the *wigger* (white hip hop fan; see Bucholtz 2011, Bucholtz & Lopez 2011).

Key points

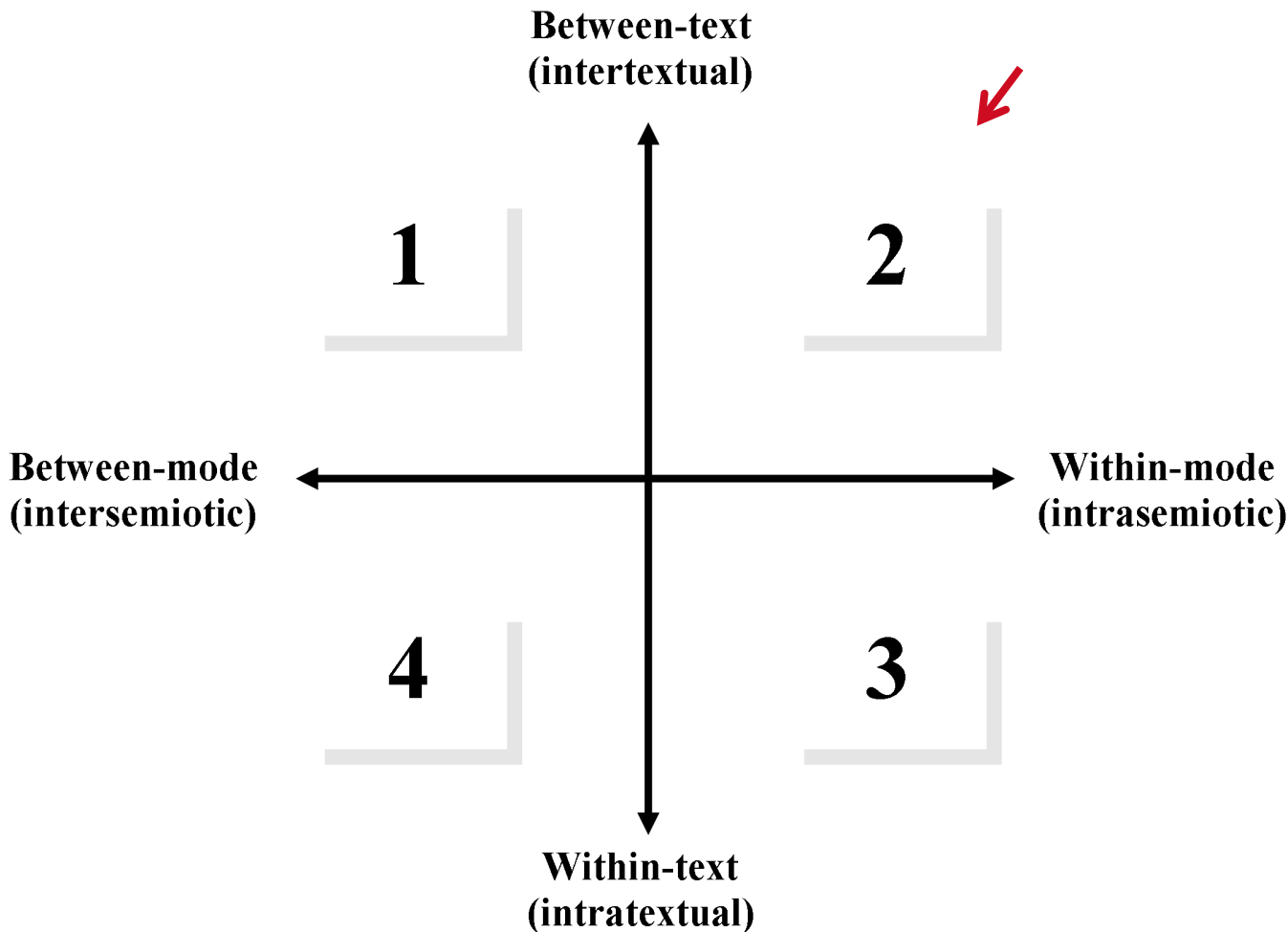
- › ‘Nonstandard’ language use is rare (except for accents);
- › ‘Nonstandard’ language features mark speakers as different, as ‘Other’;
- › ‘Nonstandard’ varieties are represented through linguistic stereotypes or ‘mock’ varieties, which can convey racist and other negative ideologies;
- › The use of ‘nonstandard’ language is associated with negative, minor, humorous, weak characters or characters that represent cultural stereotypes (while ‘standard’ English may be associated with heroes or desirable qualities);
- › ‘Nonstandard’ (non-native English) speakers may be represented as having an inferior language proficiency compared to L1 speakers.

WordSmith (Scott 2017); GraphColl (Brezina et al 2015; now Lancsbox)

- > Frequency
- > Range
- > Concordance
- > Collocation



Topology for discourse analysis
(Bednarek & Caple 2017)



Ain't as stigmatised linguistic feature

- › ‘shibboleth[.] of nonstandard usage’ (Wolfram & Schilling-Estes 2006: 336)
- › ‘Criticism of *ain't* has been so pervasive and effective that despite the word’s widespread use in many nonstandard varieties of American English, as well as in the colloquial speech of many standard American English speakers, many speakers of American English (nonstandard and standard) see the word as improper and the speakers who use it as violating fundamental principles or laws of English.’ (Curzan 2014: 31)
- › In the narrative mass media: Mitchell (2015), Queen (2015)

Variation (f = 92, r = 22; disp = 0.65 [whole corpus])

- › 1x: *Baby Daddy, Bones, Castle, Entourage, Jericho, Lost, My Name is Earl, NCIS, Nurse Jackie, Southland, Tru Calling;*
- › 2x: *Breaking Bad, Dexter, Prison Break;*
- › 3x: *Human Target, Mike and Molly;*
- › 4x: *Weeds;*
- › 6x: *Eastbound and Down, Pushing Daisies, The Shield;*
- › 16x: *True Blood;*
- › 31x: *The Wire*

Collocates

- › Non-standard/colloquial variants: e.g. *til, y'all, gonna, nothin', comin'*;
- › Taboo/curse/swear words: *ass, shit, motherfucker, fucked*;
- › Multiple negation: *no, nobody, nowhere, nothin'*;
- › Particular social groups (?): *drugs, <prostitute>* ;
- › Particular characters: e.g. *<D'angelo>, <herc>*

<SAVINO:> Like the man said, it **ain't** ours.

<HERC:> It **ain't** yours? So, you don't mind if we just take it off your hands?

(*The Wire*)

Character diffusion (third type of distribution)

- › Used by 53 different characters in total
- › 24/53 = African American (less than 50%)
- › 42/53 = male (almost 80%)
- › Used by only one character in episode (e.g. *Lost*: Sawyer, *Weeds*: Heylia)
- › Used by two or more characters (e.g. *Mike & Molly*: Carl, Grandma)
- › Most variation in *True Blood*, *The Wire*:
 - *True Blood*: four African Americans, five non-African Americans, five female, five male, indexing Southern setting/identity (Louisiana);
 - *The Wire*: 11 African Americans, four non-African Americans, one female, both drug trade (7 + 2 drug addicts) and police force (4 + 1 drug counsellor); may not be highly stigmatised

Usage

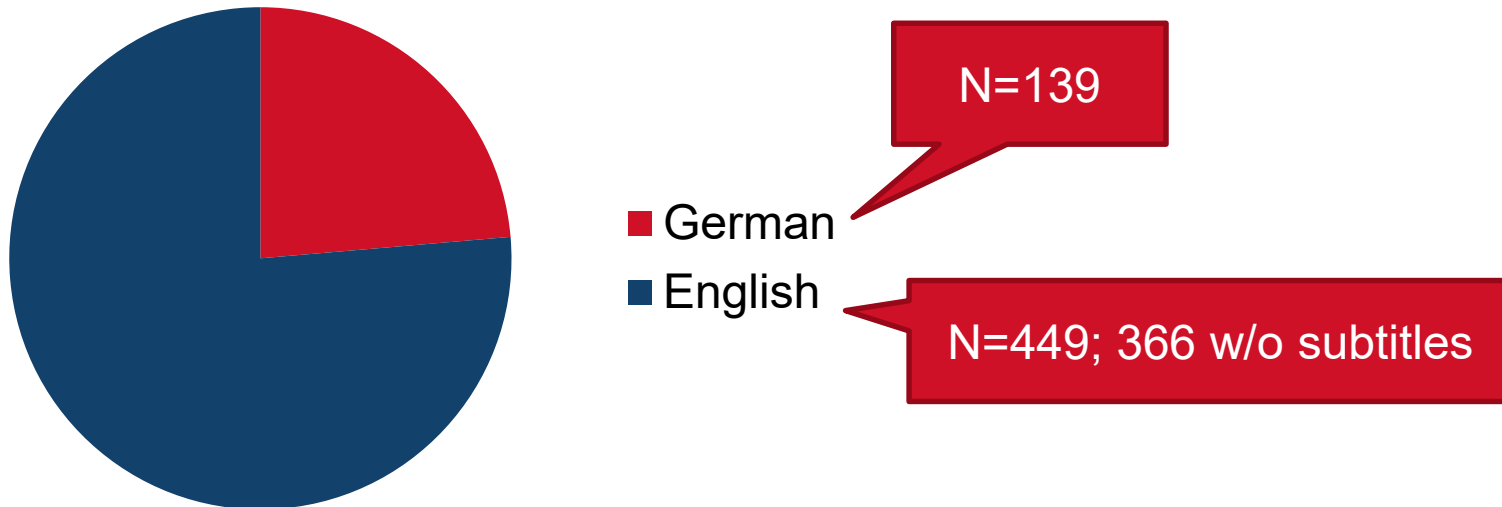
- › Vast majority: Variant of BE (*'m not, isn't, aren't*) or HAVE (*haven't, hasn't*)
 - attested in most/all vernacular varieties in North America (Wolfram 2008: 524; Anderwald 2012: 313)
 - 'extremely common' in colloquial American English (Murray & Simon 2008: 405)
 - contrasts with *ain't* as variant of *didn't* ('distinctive' urban African American Vernacular English, Wolfram 2008: 524)
- › Largely familiar usage → audience design
- › Mirrors findings from Hollywood films (Bucholtz & Lopez 2011: 689)

5 (white, male/female) Hollywood screenwriters

- › What type of character would use *ain't*?
 - 'Blue-collar', 'working class', 'less-educated', 'lower class', 'stupid/dumb'
 - Associations with education/class rather than ethnicity
- › What is your view on representing dialect?
 - Desire not to offend or stereotype
 - Fear of errors
 - Avoid being too specific
 - Dialect 'flavour' (casting)
 - Leaving it up to the actors

Viewers

- › Survey (~600 students, 7 German universities)



NB Broadcast in **dubbed, German** version only

Linguistic awareness

- › Q10: In my view there are linguistic differences between the language spoken by TV characters in such series and the language spoken by 'real' speakers
 - No, there are no differences
 - I don't know/am unsure
 - Yes, the following differences exist: [space for answers]

Dialects/accents

- › Q10 ('yes, the following differences exist'):
 - fewer dialects/accents, less dialect variation, dialects/accents less strong (n=27)
 - dialects/accents exaggerated, stereotypical, unrealistic, inaccurate (n=10)
 - 'dialects/accents'

Non-standard language in contemporary television

› Variation

- ... each series will have its own linguistic profile, being a cultural artefact in its own right ... (Bednarek 2012: 59-60).

› Clustering

› Mirroring increased diversity

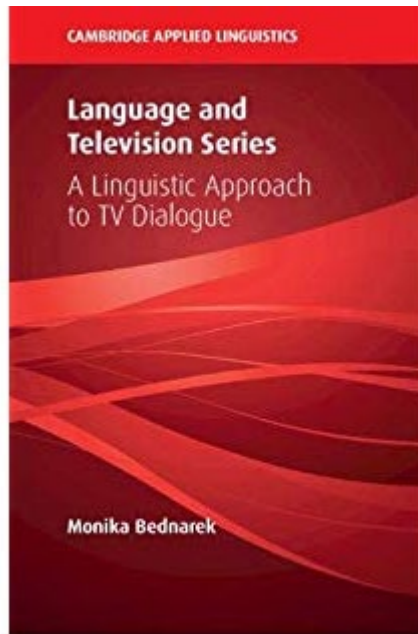
› Potential for linguistic stereotyping

› Points for consideration (Bednarek 2010: 218):

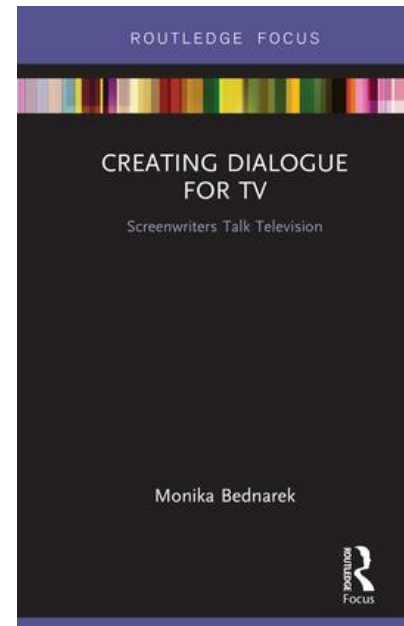
- Visibility (presence, how many...)
- Portrayal (positive, negative, stereotypical...)
- Evaluations by other characters (positive, negative)
- Positioning of the audience (invited to share portrayal/evaluation?)

- › Some linguistic awareness of the danger of stereotyping
- › Viewers' ability to recognize (and contest?) these

› Linguistic study (2018)



› Interviews (2019)



www.syd-tv.com; www.sydneycorpuslab.com