MANCHESTER 1824
The University of Manchester

Conducting Corpus-based Research on Spoken Language

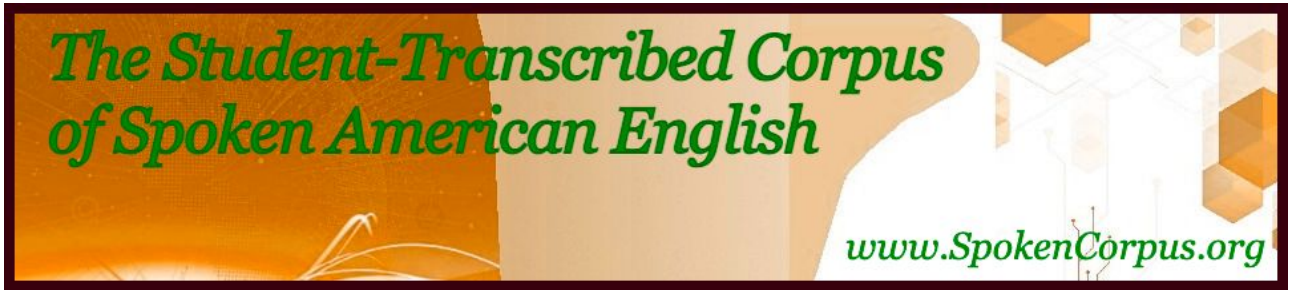
Undergraduate Scholars Programme 2021

MANCHESTER 1824
The University of Manchester

About the Project

- Learn how to transcribe speech professionally
- Use spoken corpus data for a linguistic research question

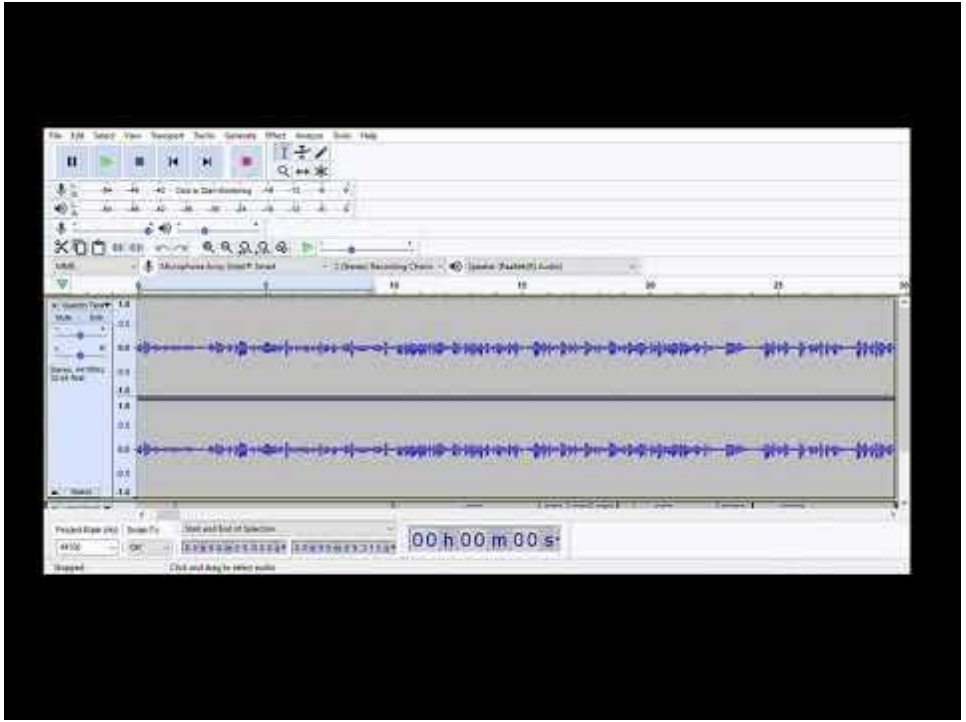
About the Corpus



- Focus on Spoken American English
- Native speaker of American English
- A total of 97,605 word tokens (including disfluencies and punctuation)

Transcription

www.spokencorpus.org



Our Research Question

What are determinants of the variation between the hedges 'sort of' and 'kind of' in spoken American English?

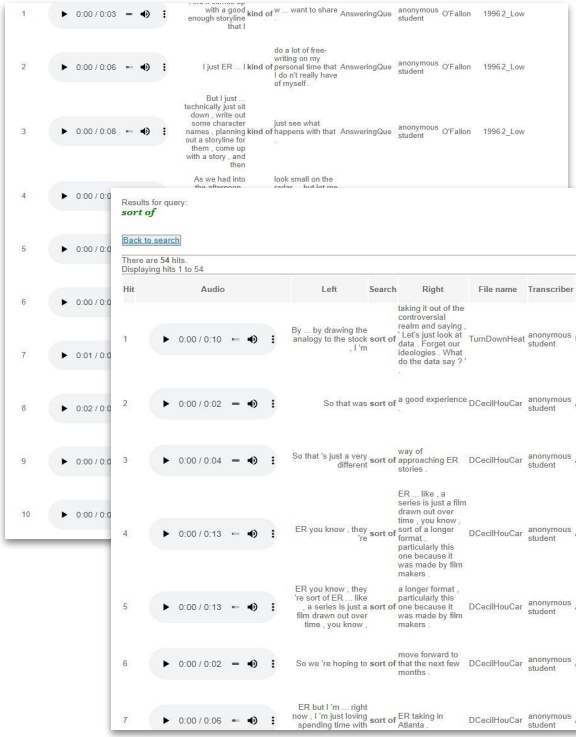
(1) So that was **sort of** a good experience
(Student-Transcribed Corpus, DCecilHouCar,
Speaker: from Texas, born 1973, education middle)

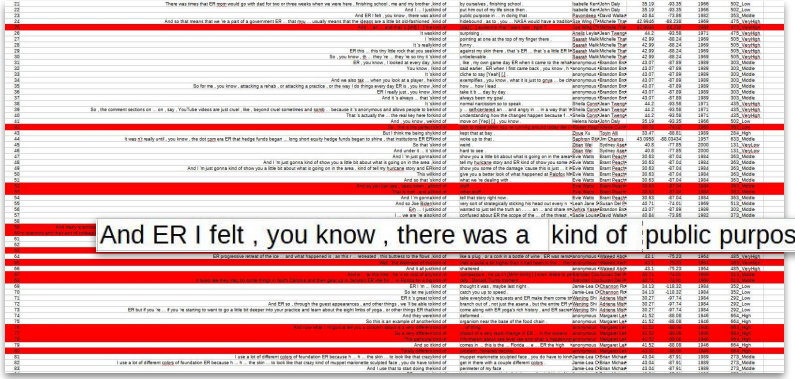
(2) Two thousand eighteen was **kind of** a depressing year.
(Student-Transcribed Corpus, MyBoatIsGone,
Speaker: from Florida, born 1984, education middle)

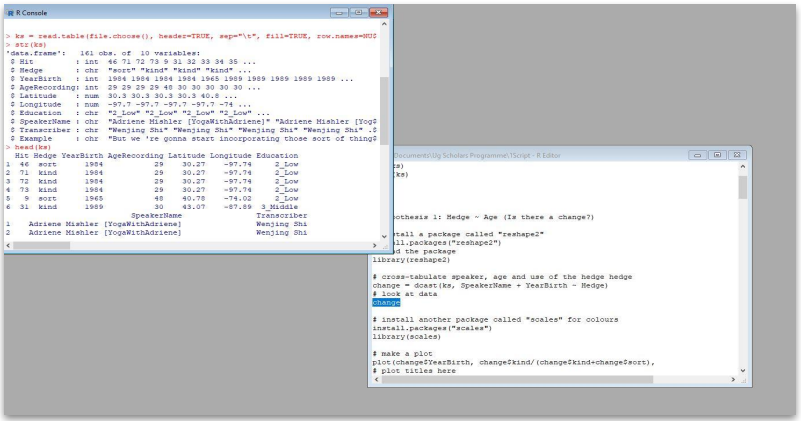
Linguistic hedge: a word or phrase used to express a lack of certainty, commitment, or accuracy, of an utterance.

MANCHESTER 1824
The University of Manchester

Our Research Method







MANCHESTER 1824
The University of Manchester

Effect of age on speakers' use of

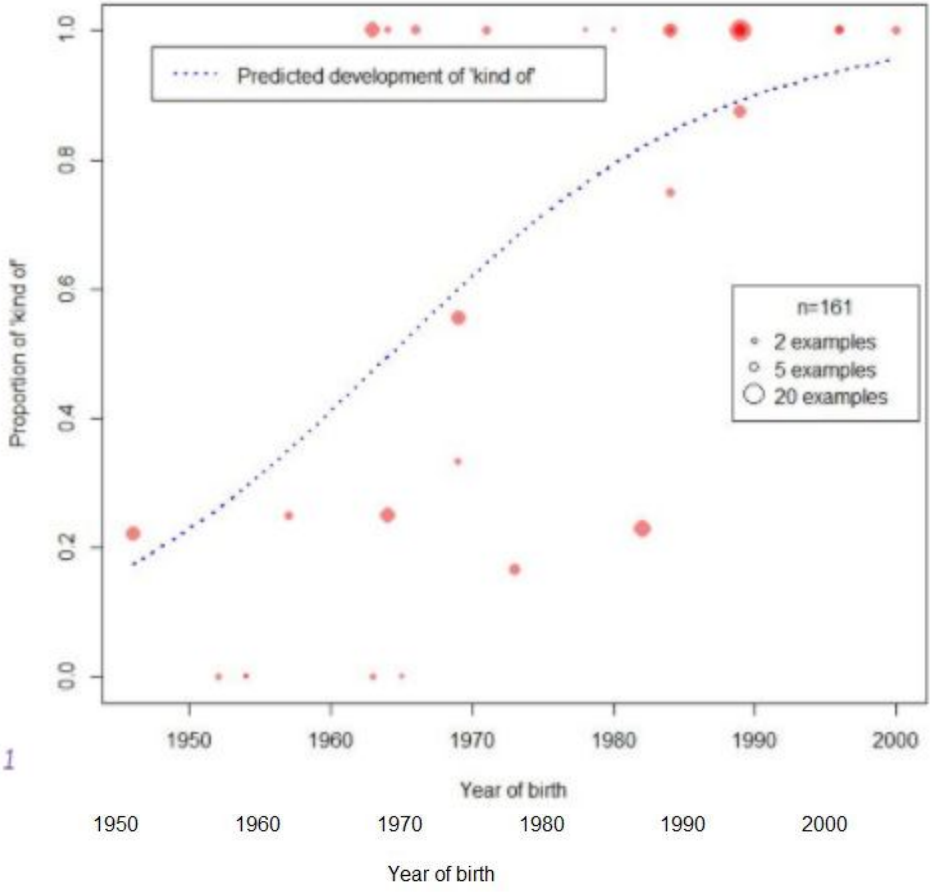


Figure 1

Figure 1

1950 1960 1970 1980 1990 2000
Year of birth

Effect of dialectal area on speakers' use of "kind of" and "sort of"

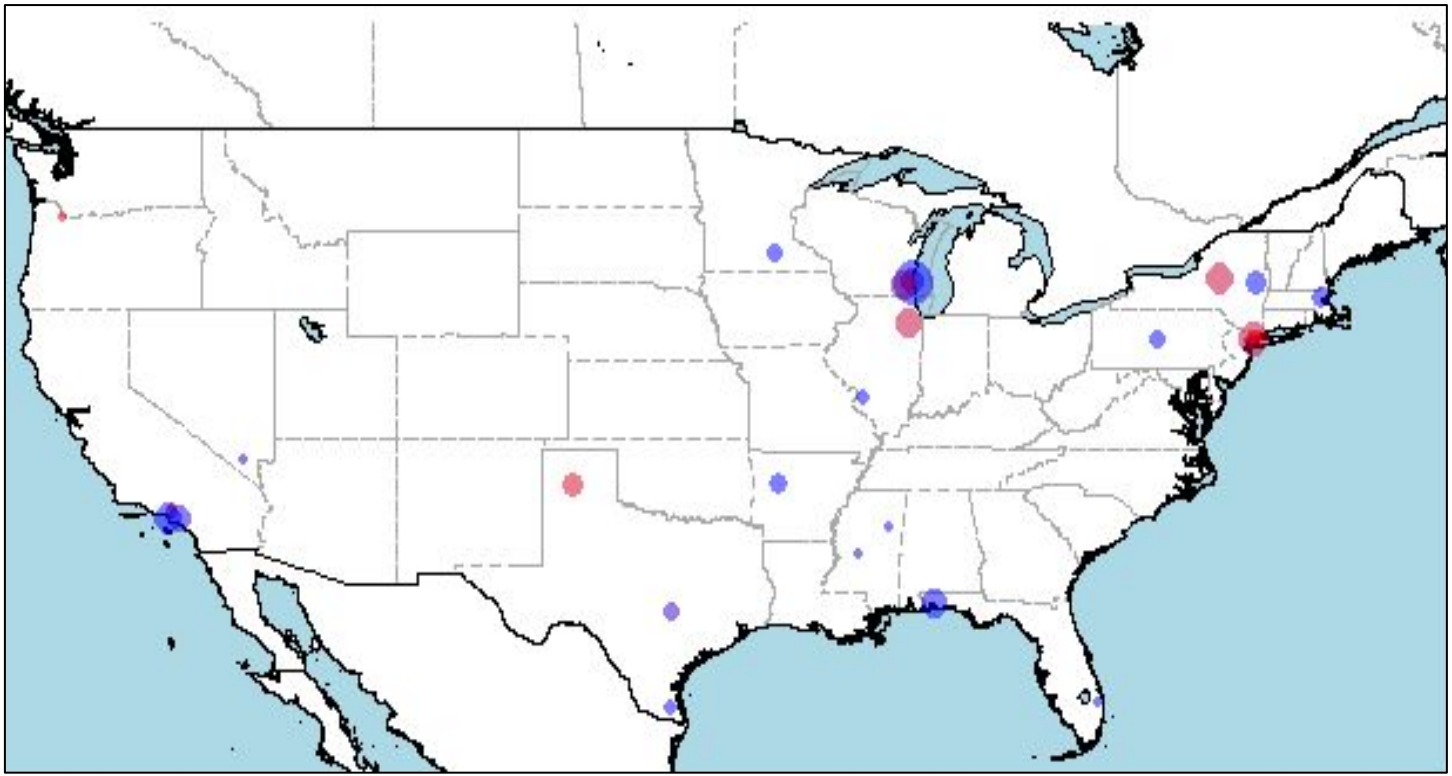


Figure 2

Effect of education on speakers' use of "kind of" and "sort of"

(χ^2=15.1, df=1, *p*<0.001***)

(odds ratio: 0.1, 95%CI: 0.02-0.36)

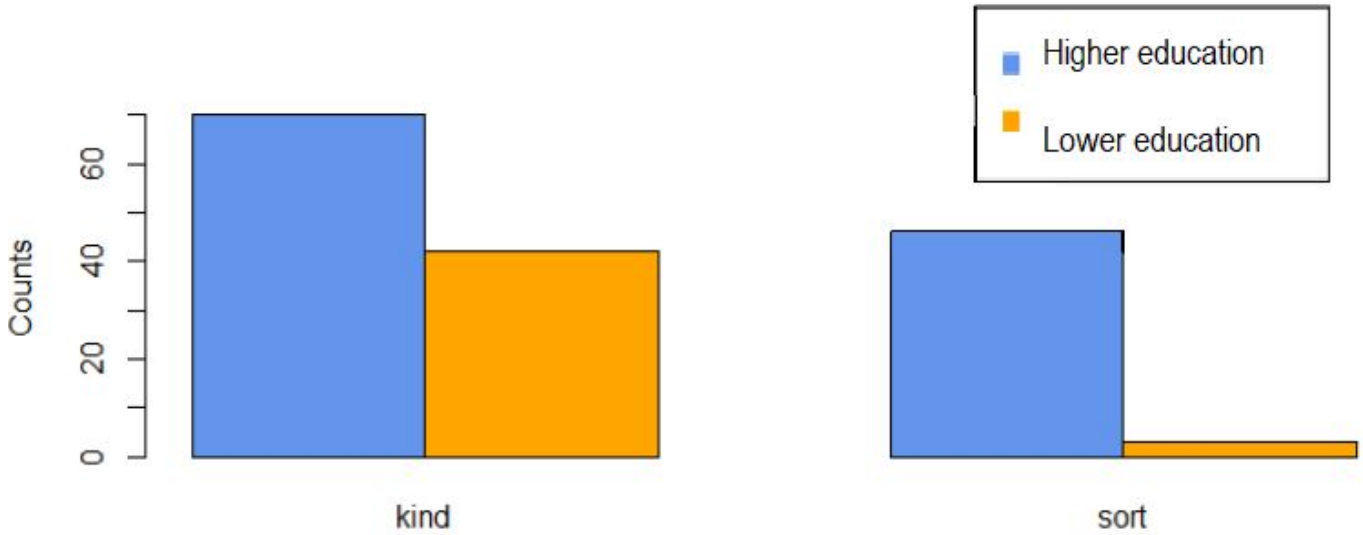


Figure 3

What We Observed

“sort of”

- More conservative
- Possibly used more in northern states
- Mostly used by people with a high education

“kind of”

- Becoming more popular
- Dispersed more inconsistently
- Used by people of both lower education and higher education, but more by lower educated people

MANCHESTER 1824
The University of Manchester

Conclusion

- Our findings are still not too clear because the corpus is small
- We would be interested in revisiting our research question again in the future when the corpus is larger
- Useful skills we've learned:
  - How to do professional transcription
  - Using audio editing software
  - Using R to analyse data