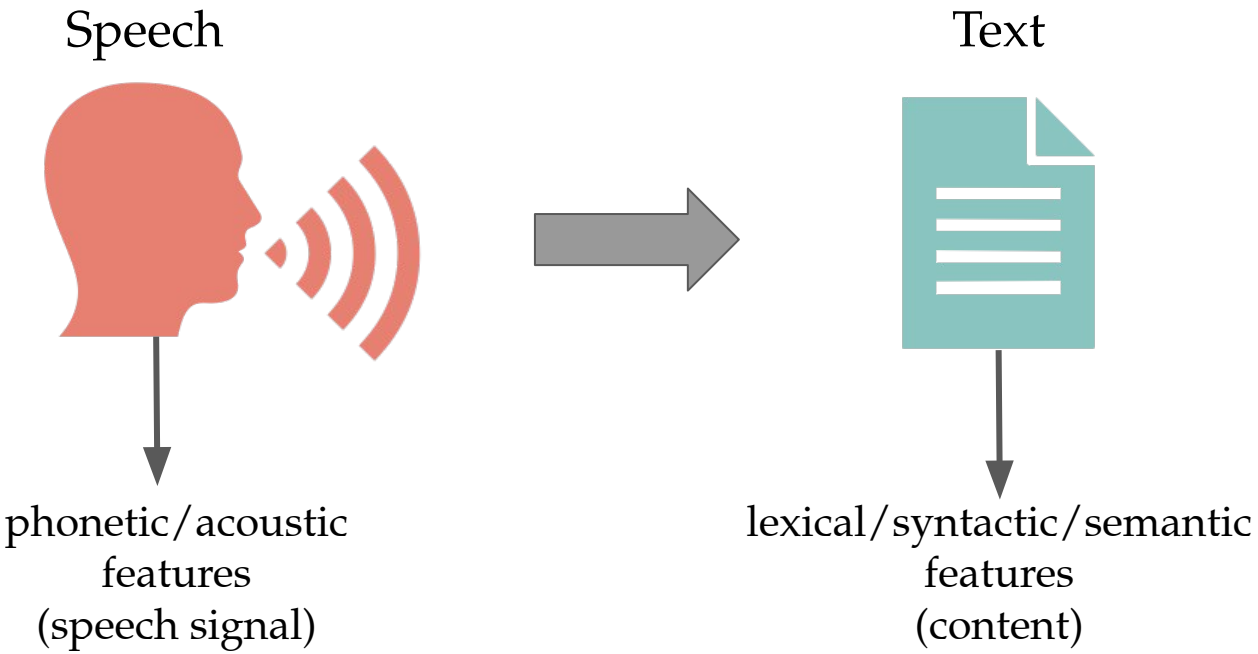
Testing authorship analysis on spoken language transcripts: Establishing a first benchmark

Cristina Aggazzotti, Ph.D.
Elizabeth Allyn Smith, Ph.D.
Nicholas Andrews, Ph.D.

UQÀM

JOHNS HOPKINS UNIVERSITY

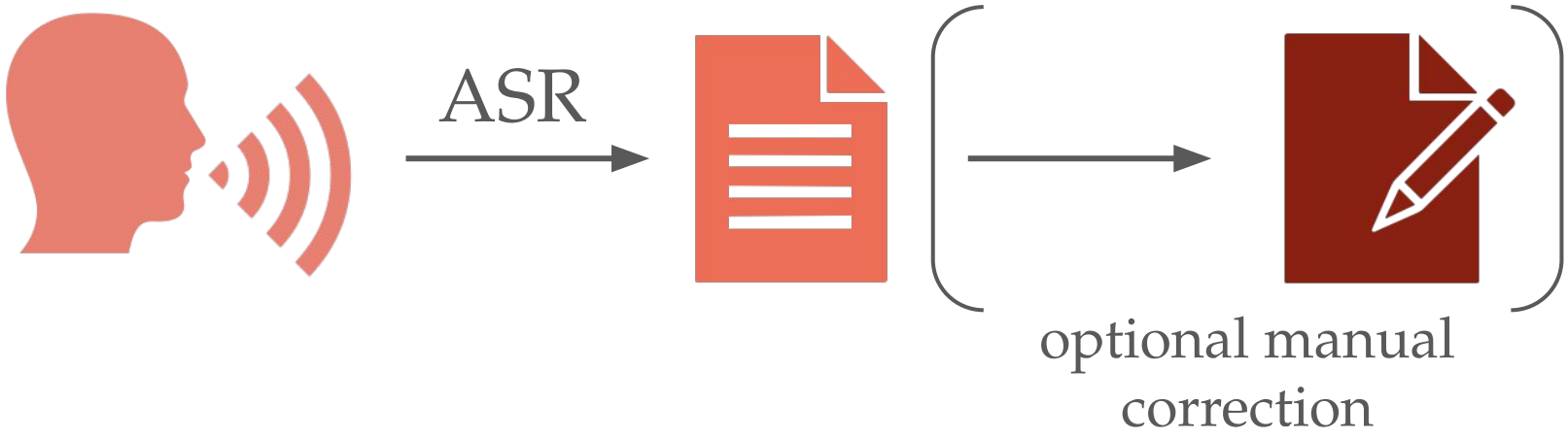
Why speech transcripts?



Advantages of combining approaches

- Provide a more comprehensive speaker profile
- Potentially improve accuracy of speaker identification
- Make speaker ID possible when speech signal degraded or unreliable
- Working with text often requires less computational power than audio files
- Transcript quality can motivate using deeper linguistic features.
- Expose a weakness of current speaker anonymization methods

Speech transcripts



Speech transcript features

A: hi
B: hey how's it going
A: pretty good
B: nice to meet you
A: you too
B: so we're supposed to talk about food huh
A: i guess the what was the topic um if we'd r- rather eat out or
B: right
B: uh it was would you rather eat out or in and uh
A: why
B: why i guess yeah all right
A: okay
B: um
A: there's like advantages to both [laughter]
B: yeah absolutely absolutely

Speech transcript features

A: hi
B: hey how's it going
A: pretty good
B: nice to meet you
A: you too
B: so we're supposed to talk about food huh
A: i guess the what was the topic um if we'd r- rather eat out or
B: right
B: uh it was would you rather eat out or in and uh
A: why
B: why i guess yeah all right
A: okay
B: um
A: there's like advantages to both [laughter]
B: yeah absolutely absolutely

Speech transcript features

A: hi
B: hey how's it going
A: pretty good
B: nice to meet you
A: you too
B: so we're supposed to talk about food huh
A: i guess the what was the topic um if we'd r- rather eat out or
B: right
B: uh it was would you rather eat out or in and uh
A: why
B: why i guess yeah all right
A: okay
B: um
A: there's like advantages to both [laughter]
B: yeah absolutely absolutely

Speech transcript features

A: hi
B: hey how's it going
A: pretty good
B: nice to meet you
A: you too
B: so we're supposed to talk about food huh
A: i guess the what was the topic um if we'd r- rather eat out or
B: right
B: uh it was would you rather eat out or in and uh
A: why
B: why i guess yeah all right
A: okay
B: um
A: there's like advantages to both [laughter]
B: yeah absolutely absolutely

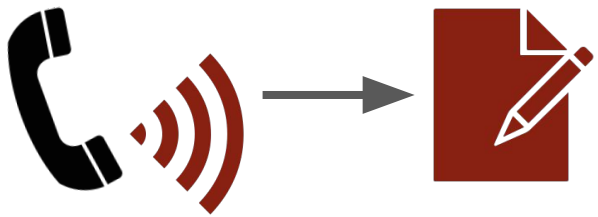
Speech transcript features

A: hi
B: hey how's it going
A: pretty good
B: nice to meet you
A: you too
B: so we're supposed to talk about food huh
A: i guess the what was the topic um if we'd r- rather eat out or
B: right
B: uh it was would you rather eat out or in and uh
A: why
B: why i guess yeah all right
A: okay
B: um
A: there's like advantages to both [laughter]
B: yeah absolutely absolutely

Challenges of using speech transcripts

- Speech data differs from written data.
- Authorship models were developed for written data.
- Speech must be transcribed.
- Transcription adds a layer of potential errors/noise.

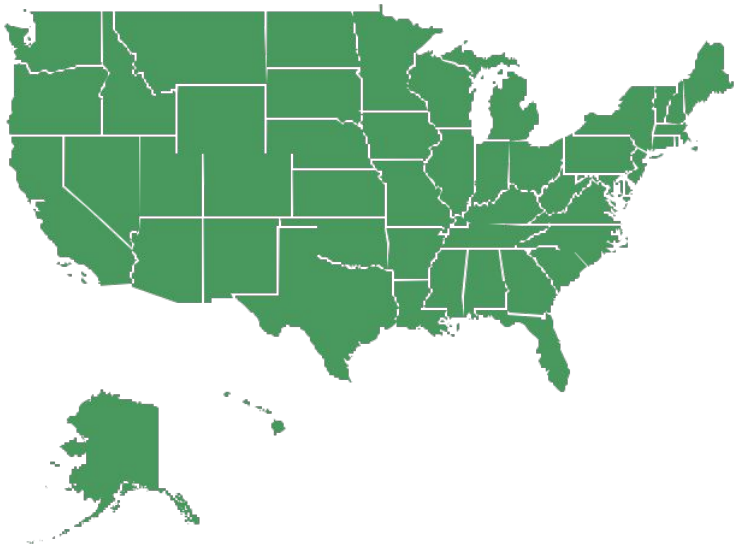
Fisher English Training Speech Transcripts Dataset



- 11,699 calls = 1,960 hours
  - Calls lasted ~10 minutes
  - Speakers on multiple calls

Cieri et al. (2004); dataset made available by the Linguistic Data Consortium

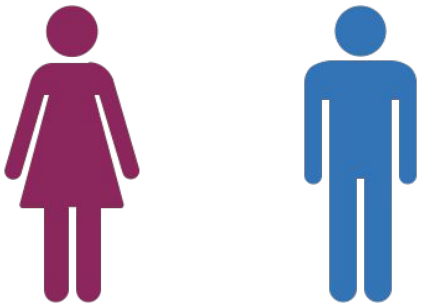
Fisher English Training Speech Transcripts Dataset



- 11,699 calls = 1,960 hours
  - Calls lasted ~10 minutes
  - Speakers on multiple calls
- 11,917 speakers

Cieri et al. (2004); dataset made available by the Linguistic Data Consortium

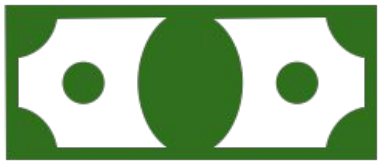
Fisher English Training Speech Transcripts Dataset



- 11,699 calls = 1,960 hours
  - Calls lasted ~10 minutes
  - Speakers on multiple calls
- 11,917 speakers
- 53% female, 47% male

Cieri et al. (2004); dataset made available by the Linguistic Data Consortium

Fisher English Training Speech Transcripts Dataset



- 11,699 calls = 1,960 hours
  - Calls lasted ~10 minutes
  - Speakers on multiple calls
- 11,917 speakers
- 53% female, 47% male
- 40 topics (part 1)
  - avg 400 calls per topic

Cieri et al. (2004); dataset made available by the Linguistic Data Consortium

Two transcription styles

Text-like

L: Yeah, they crawl.
R: They don't technically swim. All right. They crawl.
L: Mhm.
R: But they do exist under water. But if I wanted to extend this further, you know, I would -- I would expand the band to all things that I- [LAUGH] --
L: [LAUGH]
R: That live and [MN] r- reproduce underwater, I guess [MN].

Normalized

A: yeah they crawl
B: they don't technically swim all right they crawl
A: mhm
B: but they do exist under water but if i wanted to extend this further you know i would i would expand the band to all things that I- [laughter]
A: [laughter]
B: that live and [mn] r- reproduce underwater i guess [mn]

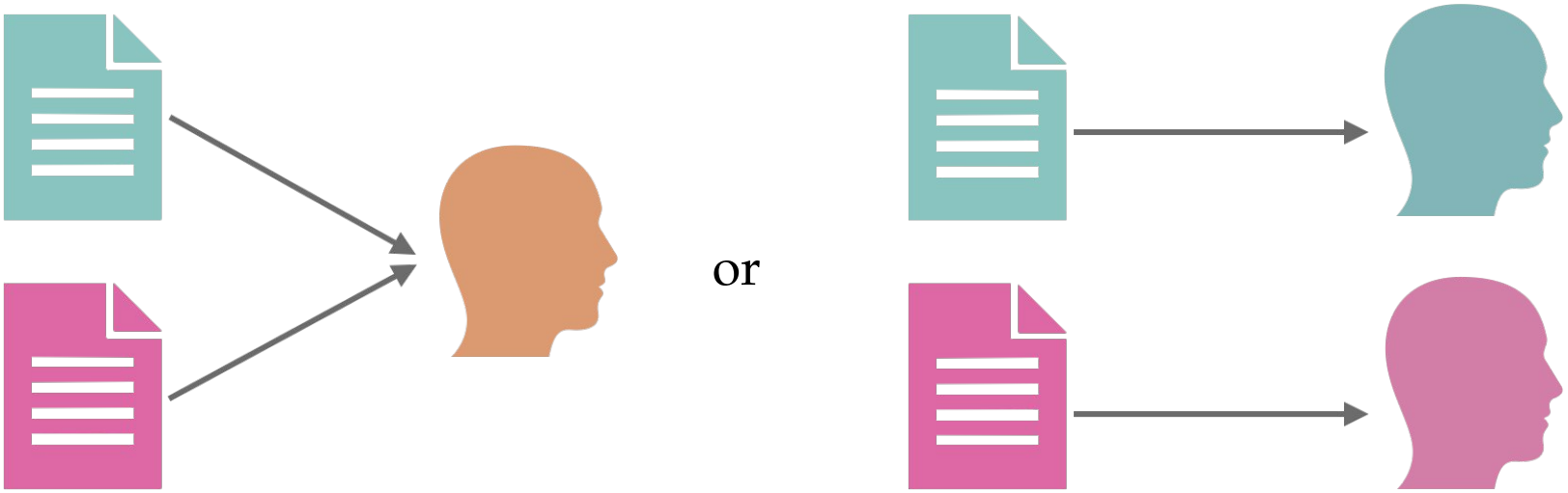
Research questions

1. How does textual authorship analysis perform on speech transcripts?

2. How does transcription style affect model performance?

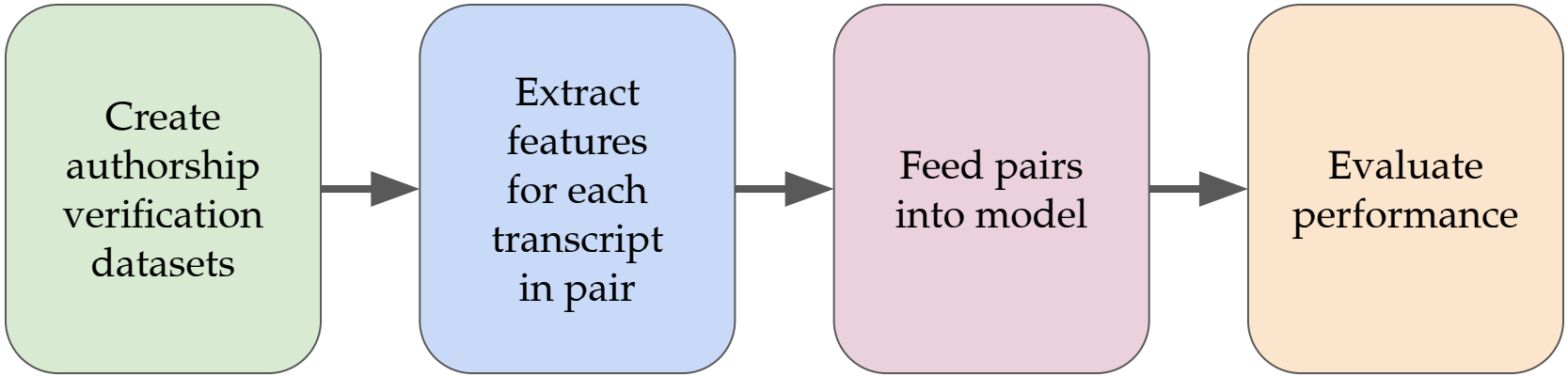
3. How does topic affect model performance?

Authorship verification

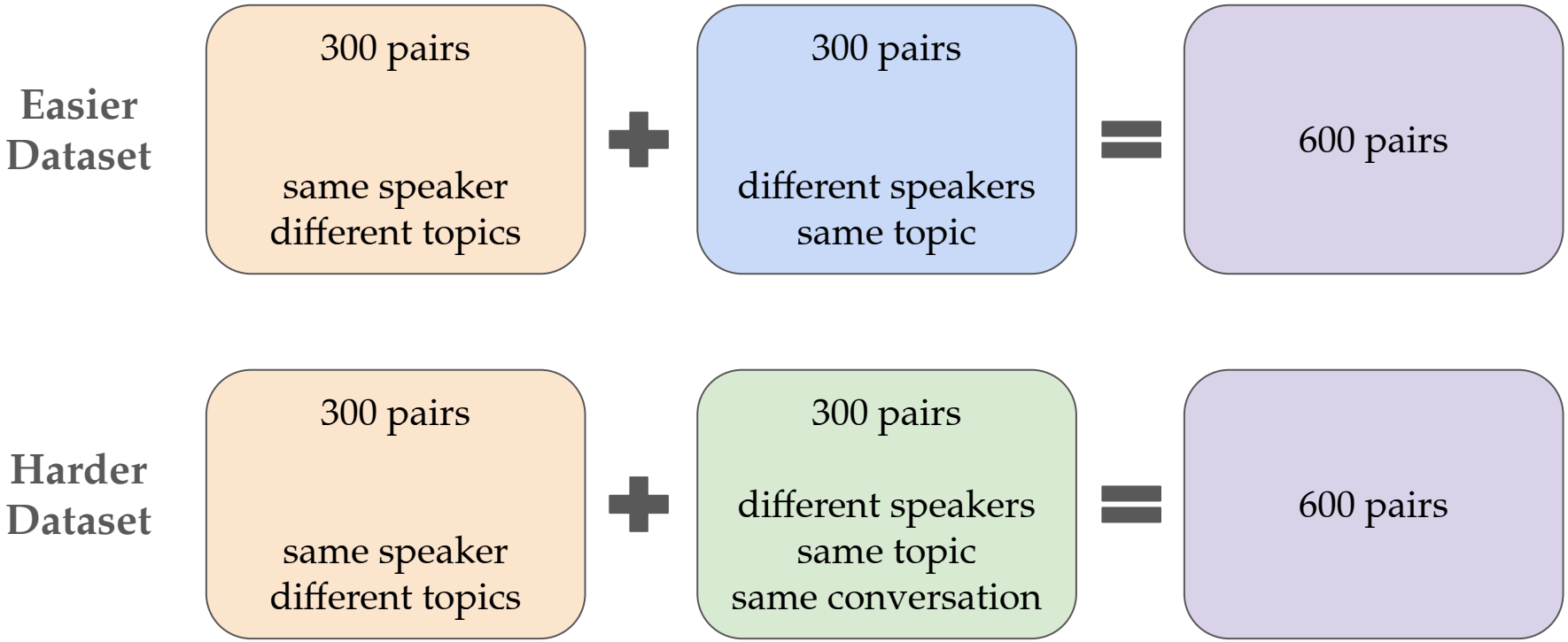


same author? or different authors?

Method


Create authorship verification datasets
Extract features for each transcript in pair
Feed pairs into model
Evaluate performance

1. Create authorship verification datasets



Same speaker, different topics

What -- what -- what should the United States do?
Well, my thought is that we have a very evil person running Iraq [MN] and he pretty clearly has either developed or is about to develop some very bad weapons that would have awful consequences for the world if they were used.
[MN] And I'm of the opinion that he needs to be removed.
[LIPSMACK] How does that compare with your thoughts on the matter?

No that -- that's all right. They, th- -- the question was, uh, what's our favorite TV sport, and how often do we watch it and things like that?
Oh really?
So, you're -- you're a pretty heavy baseball fan?
And let's see. Your favorite team is -- would be, I suppose the Yankees or the Mets?
The New York Yankees?

Different speakers, same topic

I'm awfully -- only watch professional football.
Yeah, when the Olympics are on I like to watch -- I guess that's not professional sports though.
Yeah.
I grew up with season tickets to the forty niners.
Yes. Where are you from?
So do you watch the eagles? Or --

Um, I def- -- I watch most all sports but my favorite sport's baseball.
Uh, I watch, uh, the Phillys, actually I'm watching them right now.
Um, I live in New Jersey but I, uh --
-- but I'm so close to -- I'm like twenty minutes away out of Philly then I watch Phillys.
Uh, no. I mean I watch all -- like if there's a gam- a good game on I'll watch all games but --

Different speakers, same topic, same conversation (*harder*)

So we're supposed to talk about the minimum wage increase?

Yeah, I guess so. Um, you think it's enough?

Yeah, ah, truth, I wasn't even aware it had gone up.

[LAUGH] I wasn't either.

[LAUGH]

I actually -- I thought it had already gone up to that a couple of years ago. I guess -- not. [MN]

Yeah.

[NOISE] Yeah.

That's actually what I thought. I'm like, I didn't know -- I don't think there's too many minimum wage jobs out there anymore, truthfully. [NOISE]

Really?

2. Extract stylometric features

Document-level

- Readability metrics
- Vocabulary richness
- Top character n-grams (n = 3, 4, 5)
- Top token n-grams (n = 3, 4, 5)
- Top POS n-grams (n = 2, 3, 4, 5)

Sentence-level

- Number of sentences
- Average sentence length

Mosteller & Wallace 1964, Stamatatos 2009, Sapkota et al. 2015, Pokou et al. 2016, Neal et al. 2017, Altakrori et al. 2021, Juola 2021, Strom 2021, Weerasinghe 2022, ...

2. Extract stylometric features

Token-based

- Number of total tokens (tokens)
- Number of unique tokens (types)
- Average word length
- POS tag counts
- Top most frequent tokens
- Unusual words

Token-based ratios

- type : token
- word lengths : num of characters
- short words : num of words
- long words : num of words
- words in all caps : num of words
- capitalized words : num of words

Mosteller & Wallace 1964, Stamatatos 2009, Sapkota et al. 2015, Neal et al. 2017, Altakrori et al. 2021, Juola 2021, Strom 2021, Weerasinghe 2022, among others

2. Extract stylometric features

Expression-based

- Function words and phrases
- Contractions vs. spelled out*
- Acronyms vs. spelled out*

Character-level

- Number of characters
- Letter counts
- Punctuation mark counts
- letters : num of characters
- uppercase letters : num of chars

*Taken from pre-compiled lists available on Wikipedia

Mosteller & Wallace 1964, Stamatatos 2009, Sapkota et al. 2015, Neal et al. 2017, Altakrori et al. 2021, Juola 2021, Strom 2021, Weerasinghe 2022, ...

3. Models

- Multinomial Naive Bayes
  - Assumes features are independent
  - Correlated features can affect performance.
- Logistic Regression
  - Can work even if some of the features are correlated (our case)

All models were implemented using Python's scikit-learn library.

4. Evaluation

- Cross-validation
  - 5 fold (i.e. average over 5 iterations of the experiment)
- Metrics
  - Accuracy
  - Receiver Operating Characteristic (ROC) curve
  - Area Under the ROC Curve (AUC)
  - Equal Error Rate (EER)

Implemented using Python's scikit-learn library

Experiment 1: Pilot Study

- Data
  - 600 pairs (easier); 600 pairs (harder)
- Features
  - Did not include n-grams
- Method
  - Used spaCy's tokenizer and POS tagger
- Model
  - Multinomial Naive Bayes
- Evaluation
  - No cross-validation
  - Accuracy and EER metrics only

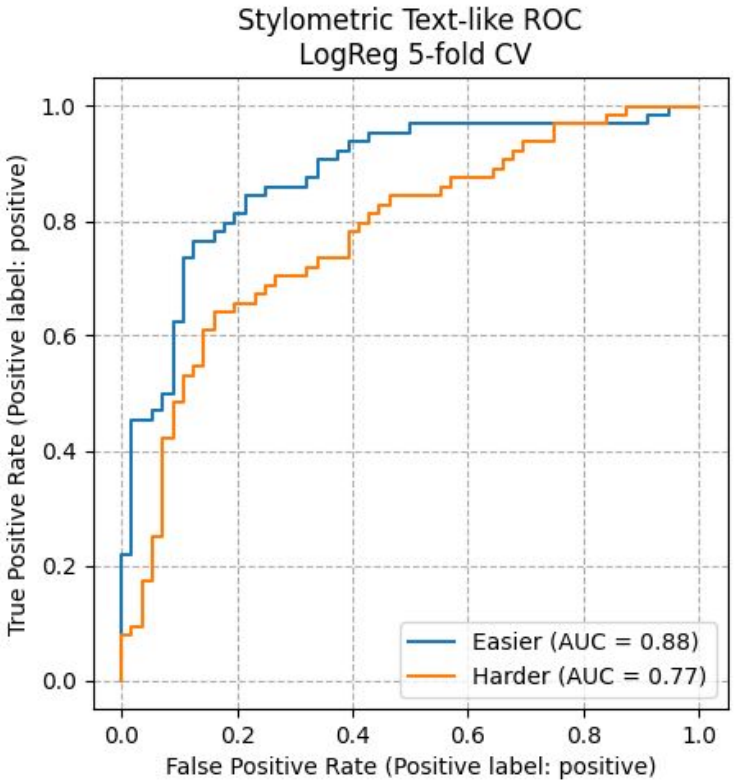
Experiment 2: Expanded

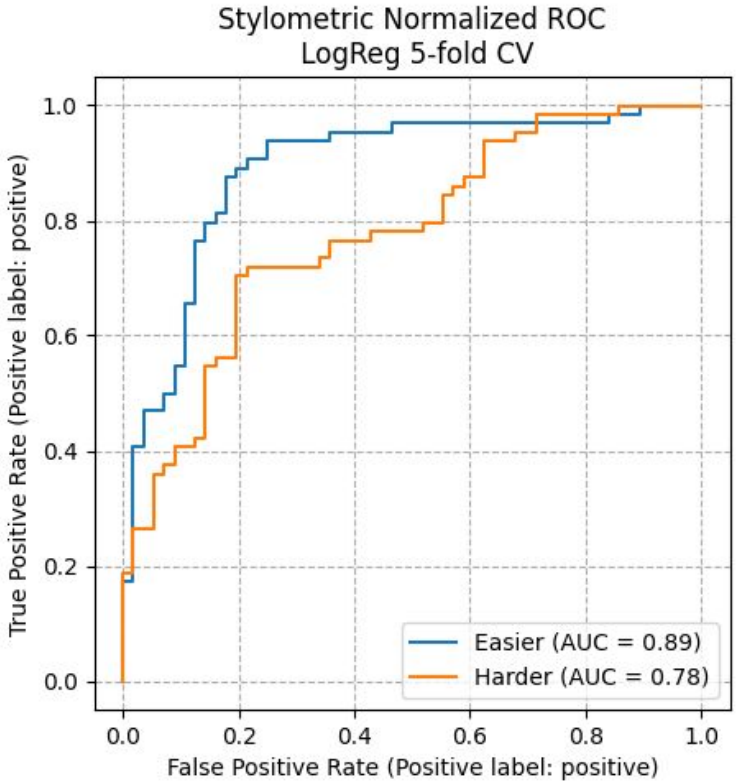
- Data
  - 600 pairs (easier); 600 pairs (harder)
- Features
  - Added more features, e.g. n-grams
  - Stress tested each feature and improved as needed
- Method
  - Used Stanza's tokenizer and POS tagger (Qi et al. 2020)
  - Parameter testing
- Models
  - Multinomial Naive Bayes
  - Logistic regression
- Evaluation
  - Cross-validation
  - Accuracy, ROC curve, AUC, and EER metrics

Results

Logistic Regression	*Text-like*		*Normalized*	
	Accuracy	**AUC**	**Accuracy**	**AUC**
Easier Dataset	0.808	0.88	0.783	0.89
Harder Dataset	0.708	0.77	0.717	0.78

Results cont.





What does this tell us?

1. How does textual authorship analysis perform on speech transcripts?
    - Comparable to authorship analysis of written language
2. How does transcription style affect model performance?
    - Does not significantly affect a stylometric model's performance
3. How does topic affect model performance?
    - Speakers in the same conversation are indeed harder to distinguish.

Summary

- ☒ Establish baseline performance of authorship analysis on speech transcripts
- ☒ Compare the effect of transcription style on performance
- ☒ Assess the effect of topic on performance (preliminary)

Summary and Future Work

- ☒ Establish baseline performance of authorship analysis on speech transcripts
- ☒ Compare the effect of transcription style on performance
- ☒ Assess the effect of topic on performance (preliminary)

- ❑ Compare stylometric method to other methods on the same data
- ❑ Combine with acoustic/phonetic-based speaker recognition methods
- ❑ Test feature subsets for which are most effective at distinguishing speakers
- ❑ Examine transcription errors and see how they affect performance
- ❑ …and various others!

Thank you / Merci / Dziękuję !