Text Analysis for Social Scientists and Leaders

Class 4: Sentence & Dialogue Structure

Prof. Michael Yeomans

Communicating Warmth in Distributive Negotiations is Surprisingly Counter-Productive

(Jeong, Minson, Yeomans & Gino, 2018)

Communicating Warmth in …

(Jeong, Minson, Yeomans & Gino, 2018)

Communicating Warmth in …

(Jeong, Minson, Yeomans & Gino, 2018)

Warm offers garner worse counteroffers

Communicating Warmth in …

(Jeong, Minson, Yeomans & Gino, 2018)

Warm offers garner worse counteroffers

- natural field experiment audit study
- 775 iPhone sellers from 15 US cities in May/June 2017

Communicating Warmth in …

(Jeong, Minson, Yeomans & Gino, 2018)

Warm offers garner worse counteroffers

- natural field experiment audit study
- 775 iPhone sellers from 15 US cities in May/June 2017

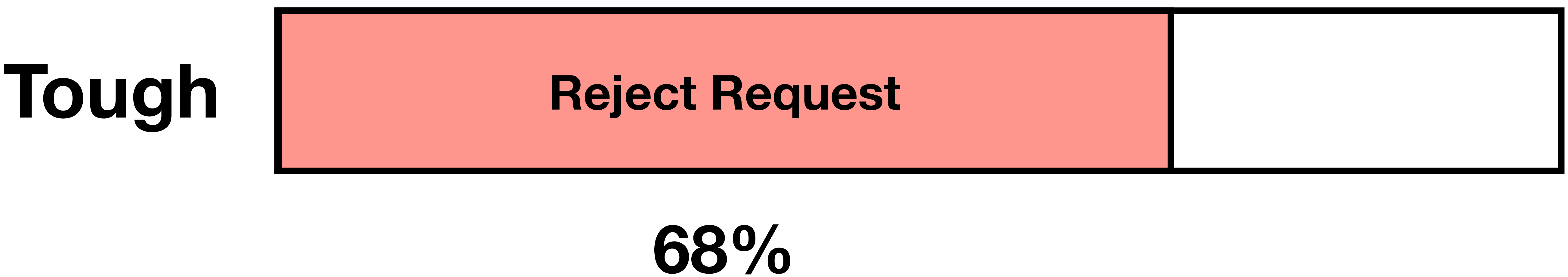
Tough

Communicating Warmth in …

(Jeong, Minson, Yeomans & Gino, 2018)

Warm offers garner worse counteroffers

- natural field experiment audit study
- 775 iPhone sellers from 15 US cities in May/June 2017

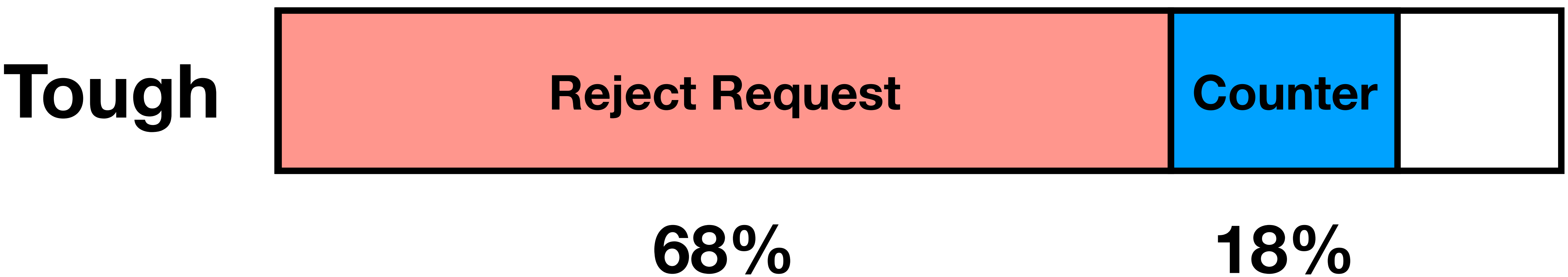


Communicating Warmth in …

(Jeong, Minson, Yeomans & Gino, 2018)

Warm offers garner worse counteroffers

- natural field experiment audit study
- 775 iPhone sellers from 15 US cities in May/June 2017

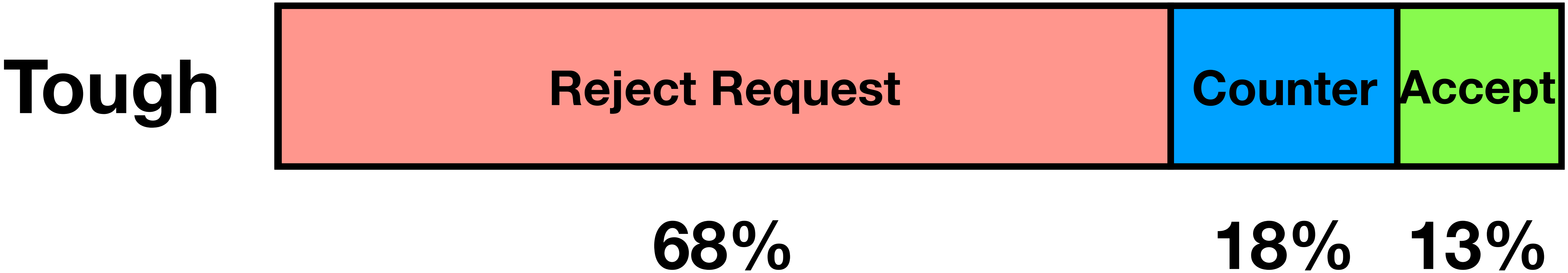


Communicating Warmth in …

(Jeong, Minson, Yeomans & Gino, 2018)

Warm offers garner worse counteroffers

- natural field experiment audit study
- 775 iPhone sellers from 15 US cities in May/June 2017

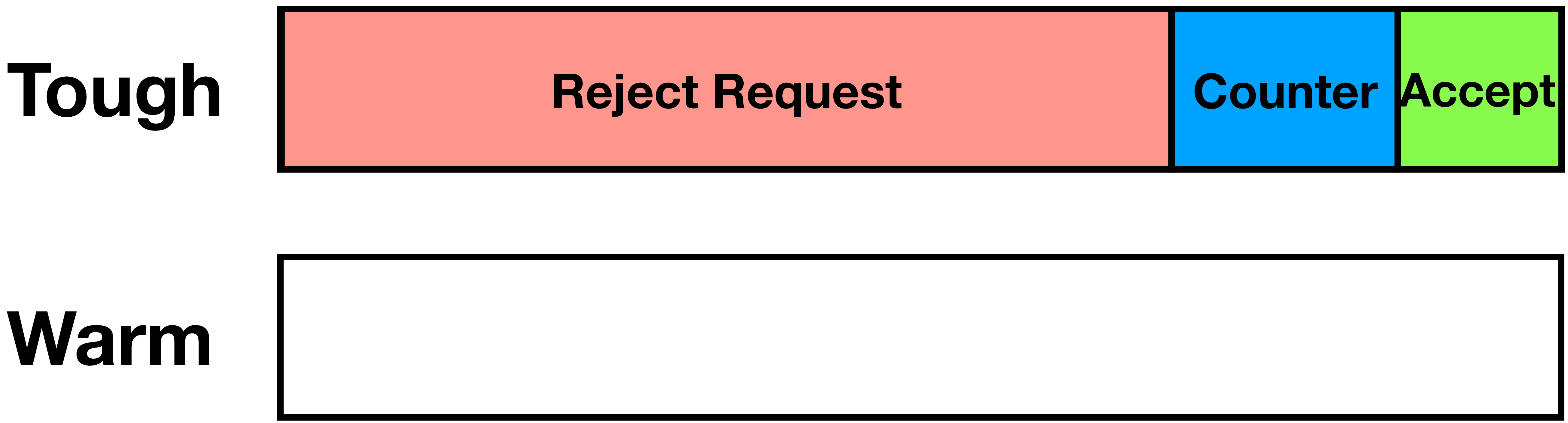


Communicating Warmth in …

(Jeong, Minson, Yeomans & Gino, 2018)

Warm offers garner worse counteroffers

- natural field experiment audit study
- 775 iPhone sellers from 15 US cities in May/June 2017

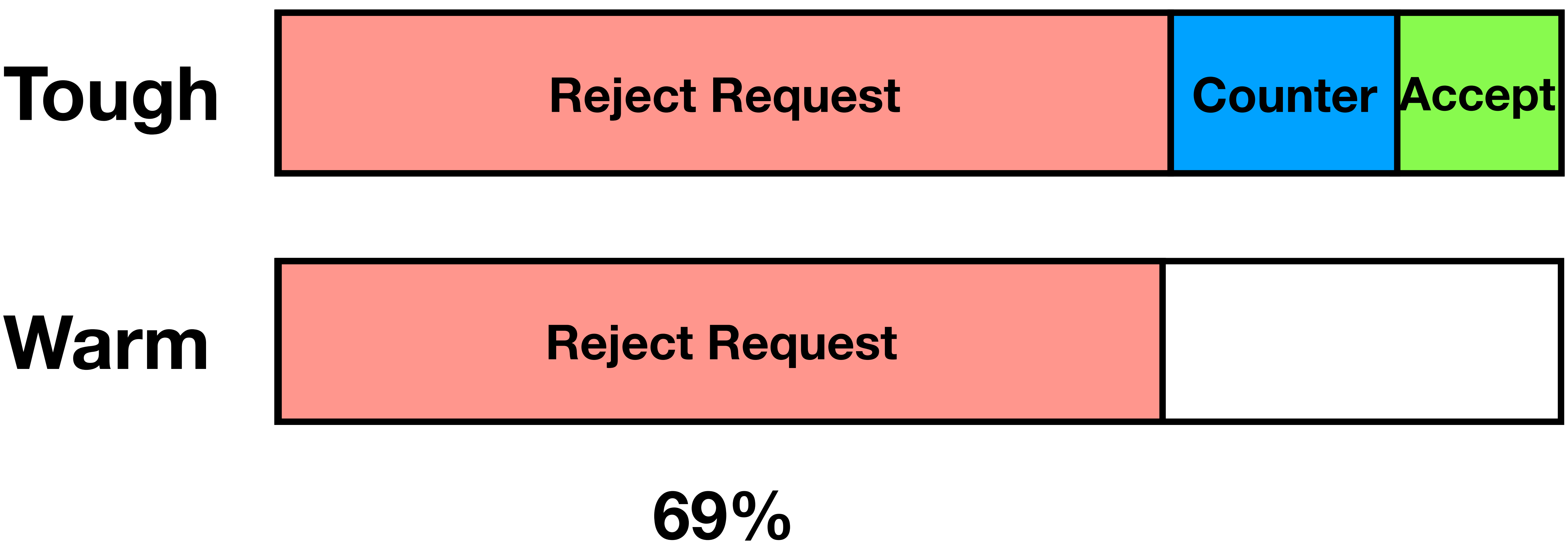


Communicating Warmth in …

(Jeong, Minson, Yeomans & Gino, 2018)

Warm offers garner worse counteroffers

- natural field experiment audit study
- 775 iPhone sellers from 15 US cities in May/June 2017

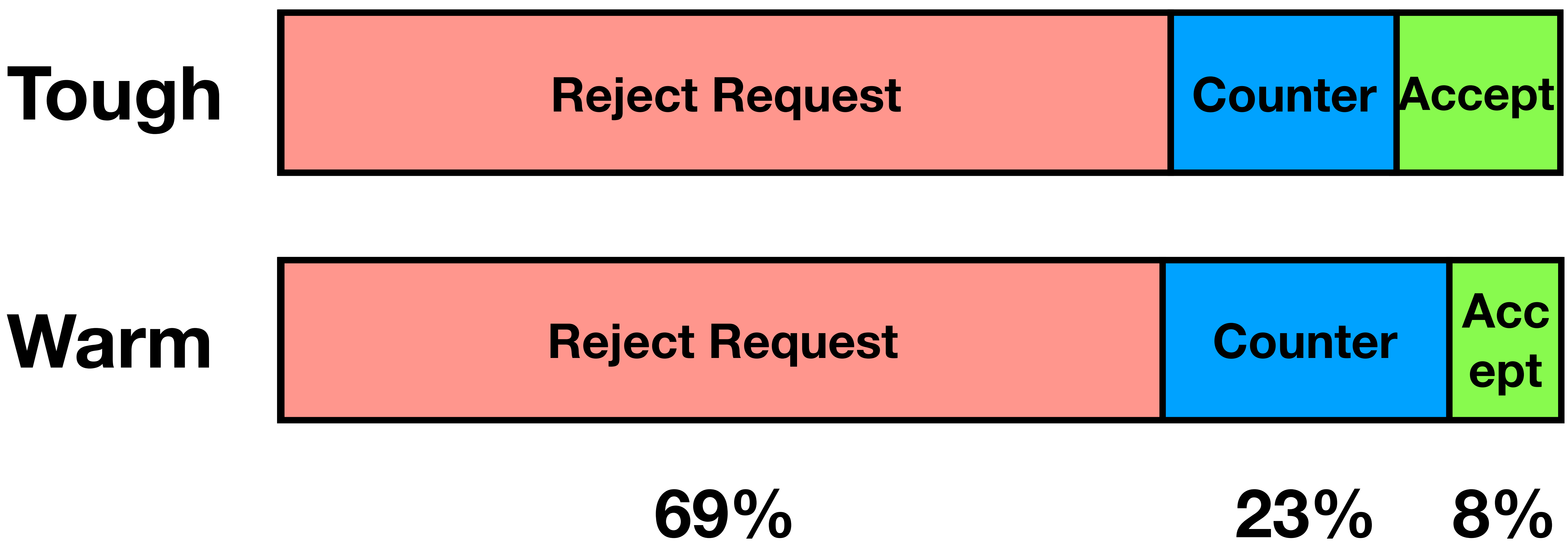


Communicating Warmth in …

(Jeong, Minson, Yeomans & Gino, 2018)

Warm offers garner worse counteroffers

- natural field experiment audit study
- 775 iPhone sellers from 15 US cities in May/June 2017



Communicating Warmth in …

(Jeong, Minson, Yeomans & Gino, 2018)

Warm offers earn worse final deals

Communicating Warmth in …

(Jeong, Minson, Yeomans & Gino, 2018)

Warm offers earn worse final deals

- incentivized lab experiment with dyads

Communicating Warmth in …

(Jeong, Minson, Yeomans & Gino, 2018)

Warm offers earn worse final deals

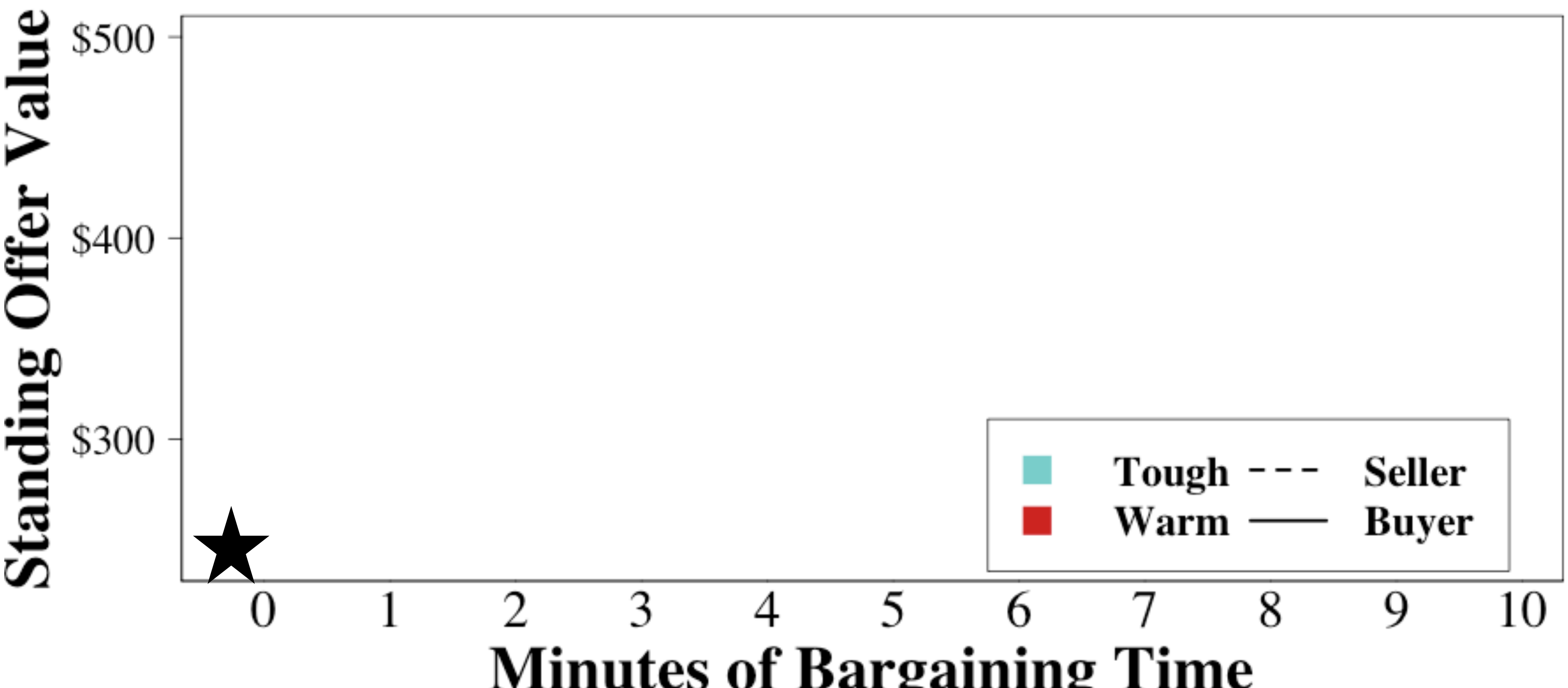
- incentivized lab experiment with dyads
- driven by early seller concessions

Communicating Warmth in …

(Jeong, Minson, Yeomans & Gino, 2018)

Warm offers earn worse final deals

- incentivized lab experiment with dyads
- driven by early seller concessions

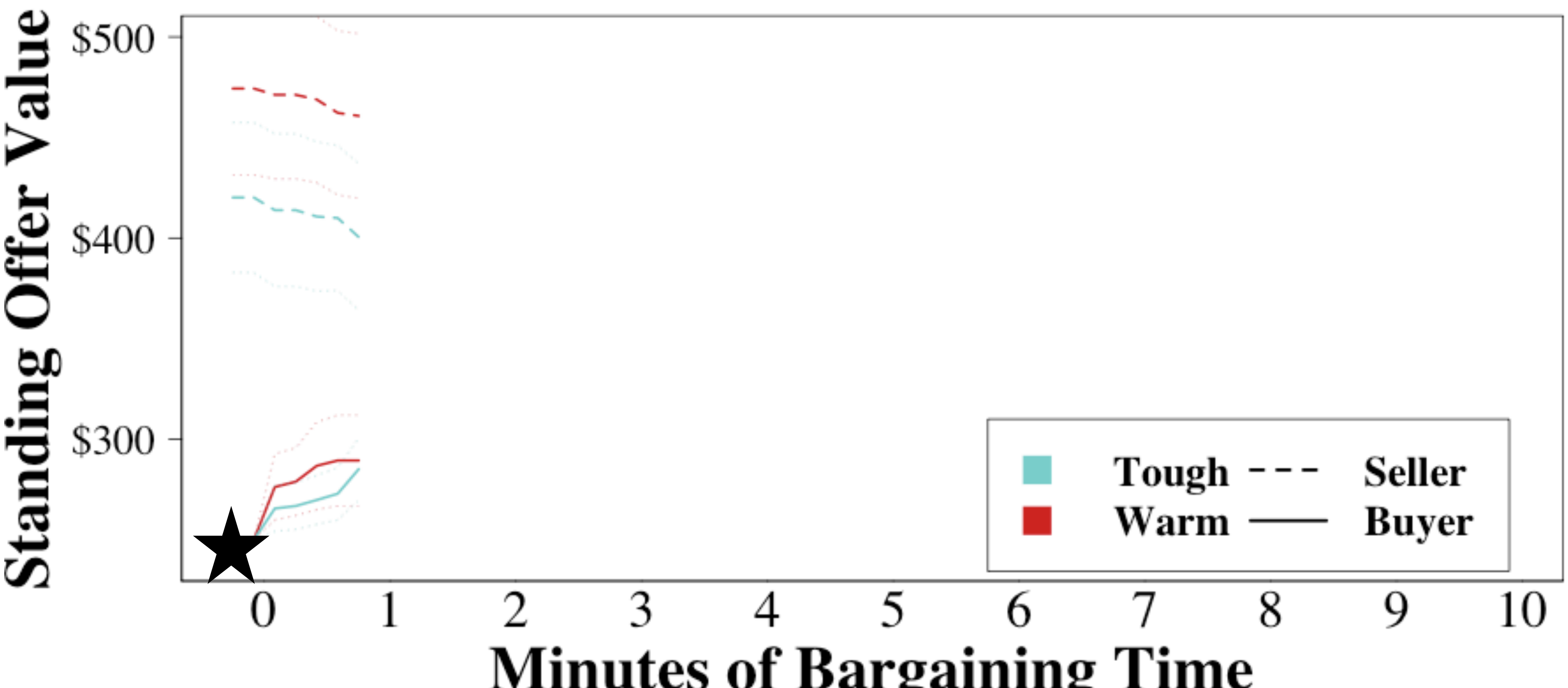


Communicating Warmth in …

(Jeong, Minson, Yeomans & Gino, 2018)

Warm offers earn worse final deals

- incentivized lab experiment with dyads
- driven by early seller concessions

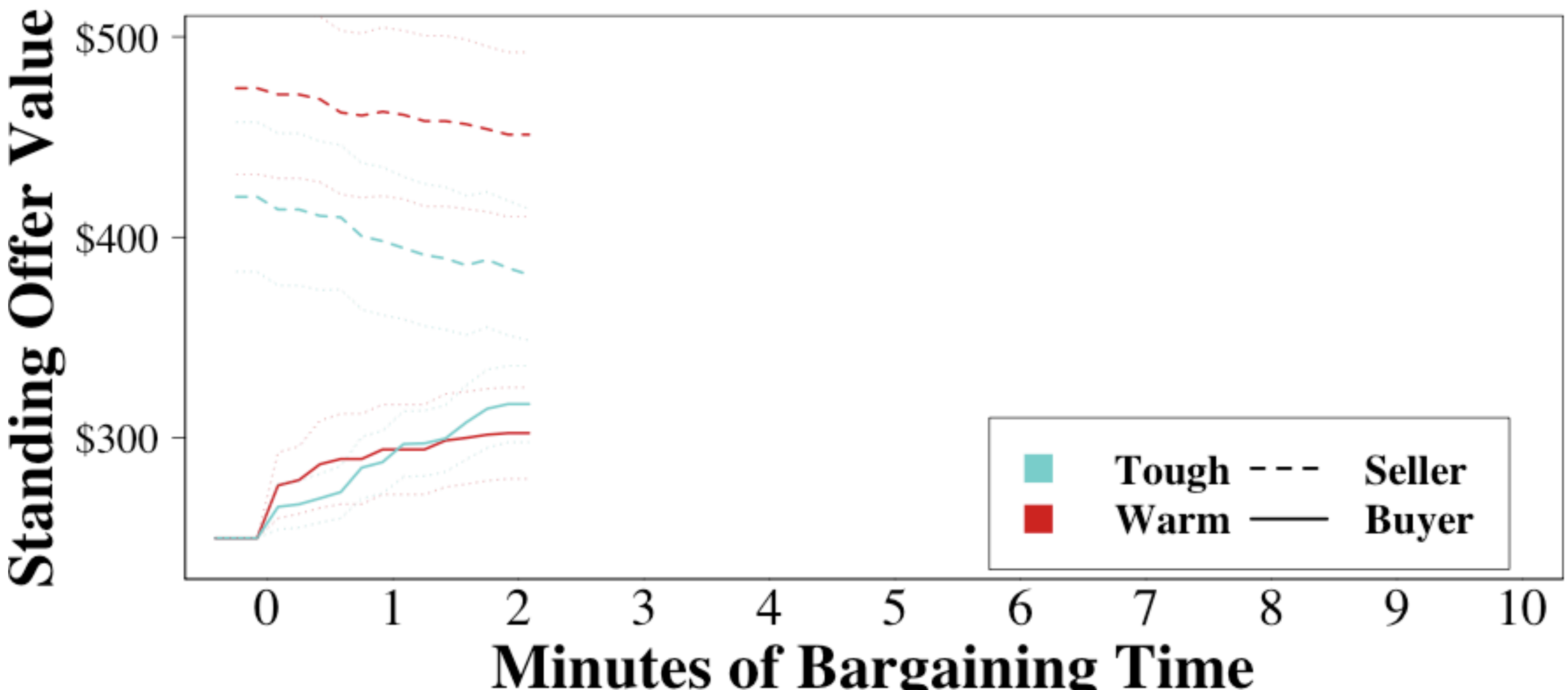


Communicating Warmth in …

(Jeong, Minson, Yeomans & Gino, 2018)

Warm offers earn worse final deals

- incentivized lab experiment with dyads
- driven by early seller concessions

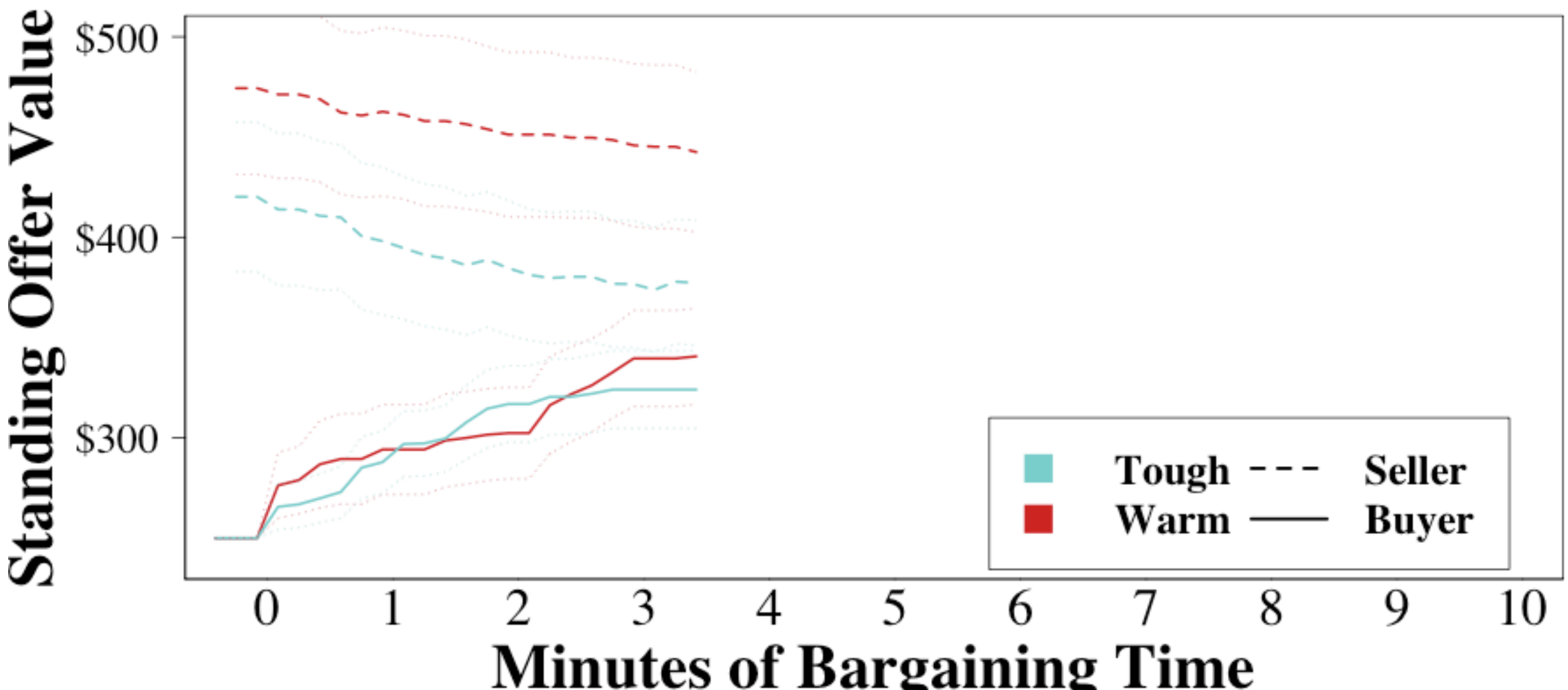


Communicating Warmth in …

(Jeong, Minson, Yeomans & Gino, 2018)

Warm offers earn worse final deals

- incentivized lab experiment with dyads
- driven by early seller concessions

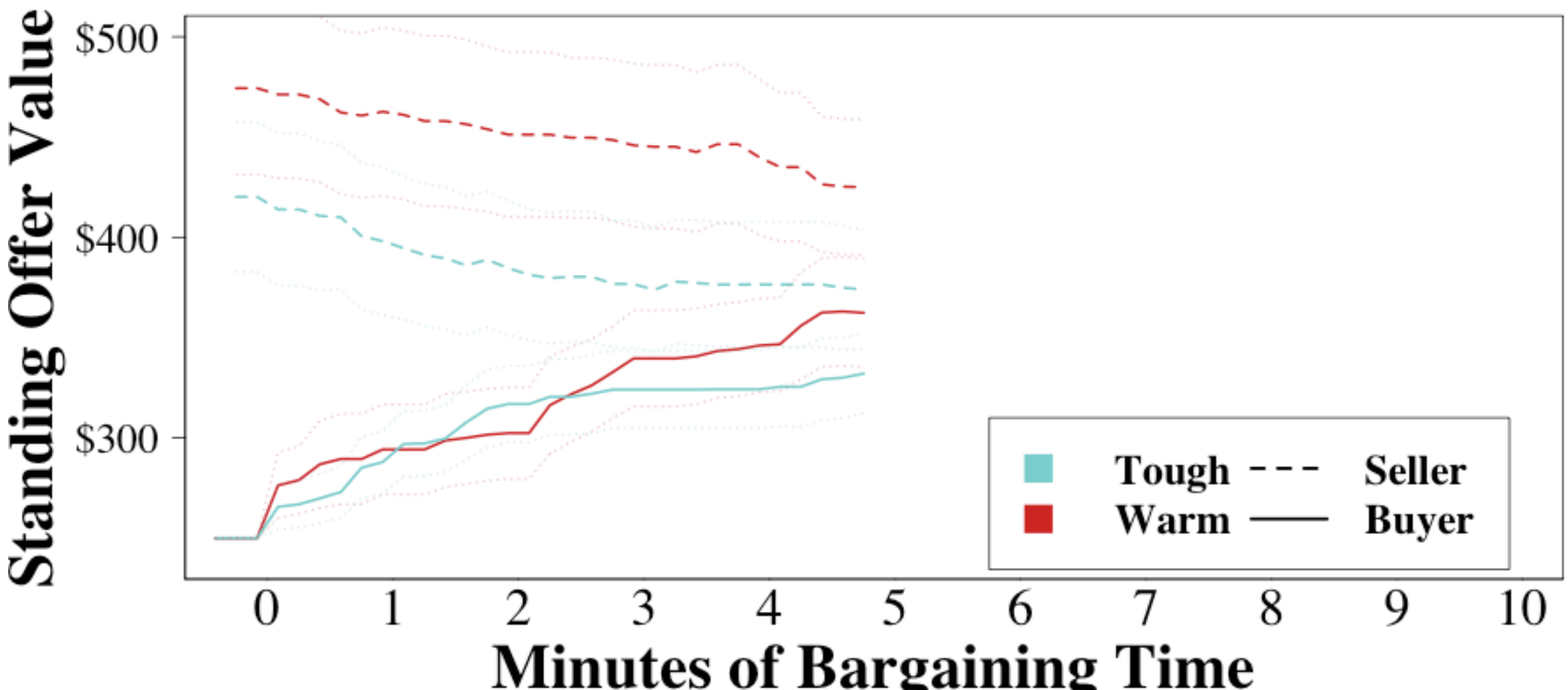


Communicating Warmth in …

(Jeong, Minson, Yeomans & Gino, 2018)

Warm offers earn worse final deals

- incentivized lab experiment with dyads
- driven by early seller concessions

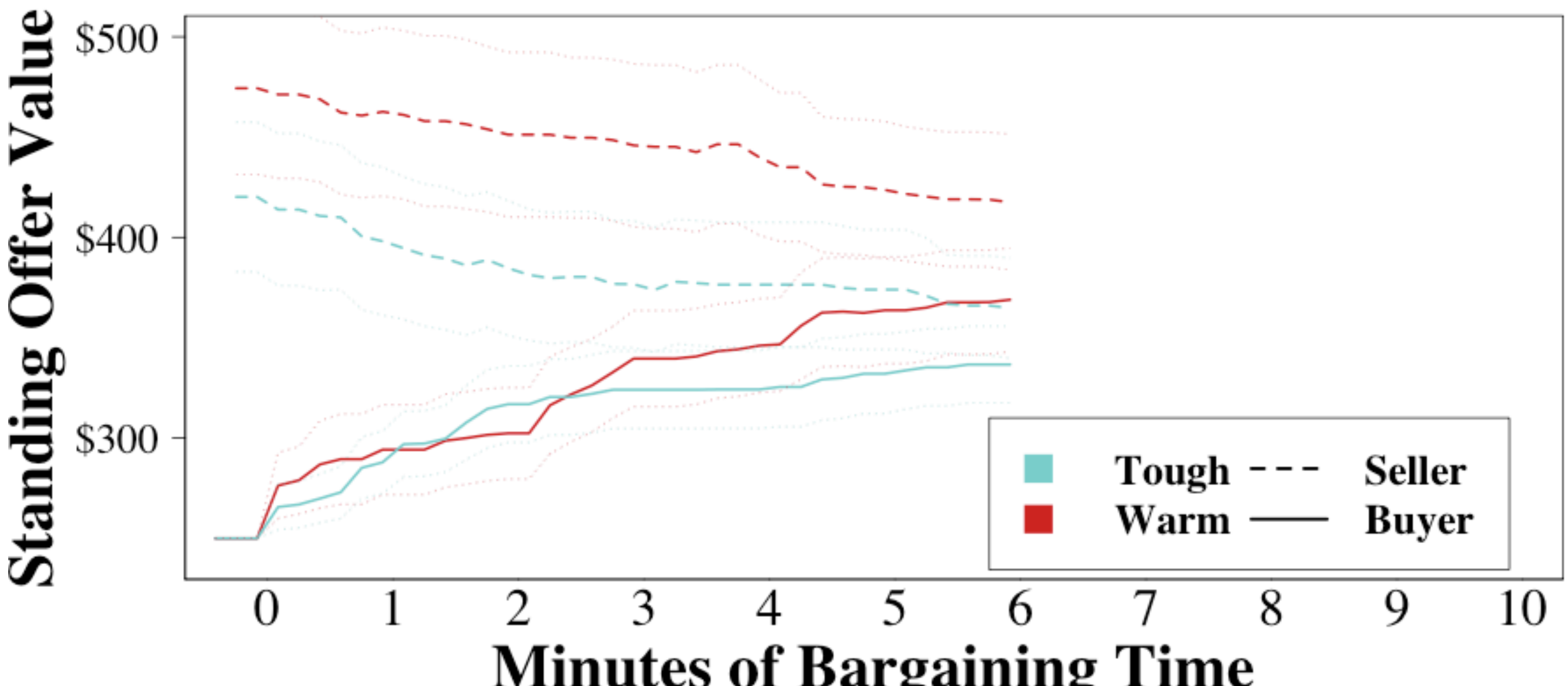


<u>Communicating Warmth in …</u>

(Jeong, Minson, Yeomans & Gino, 2018)

Warm offers earn worse final deals

- incentivized lab experiment with dyads
- driven by early seller concessions

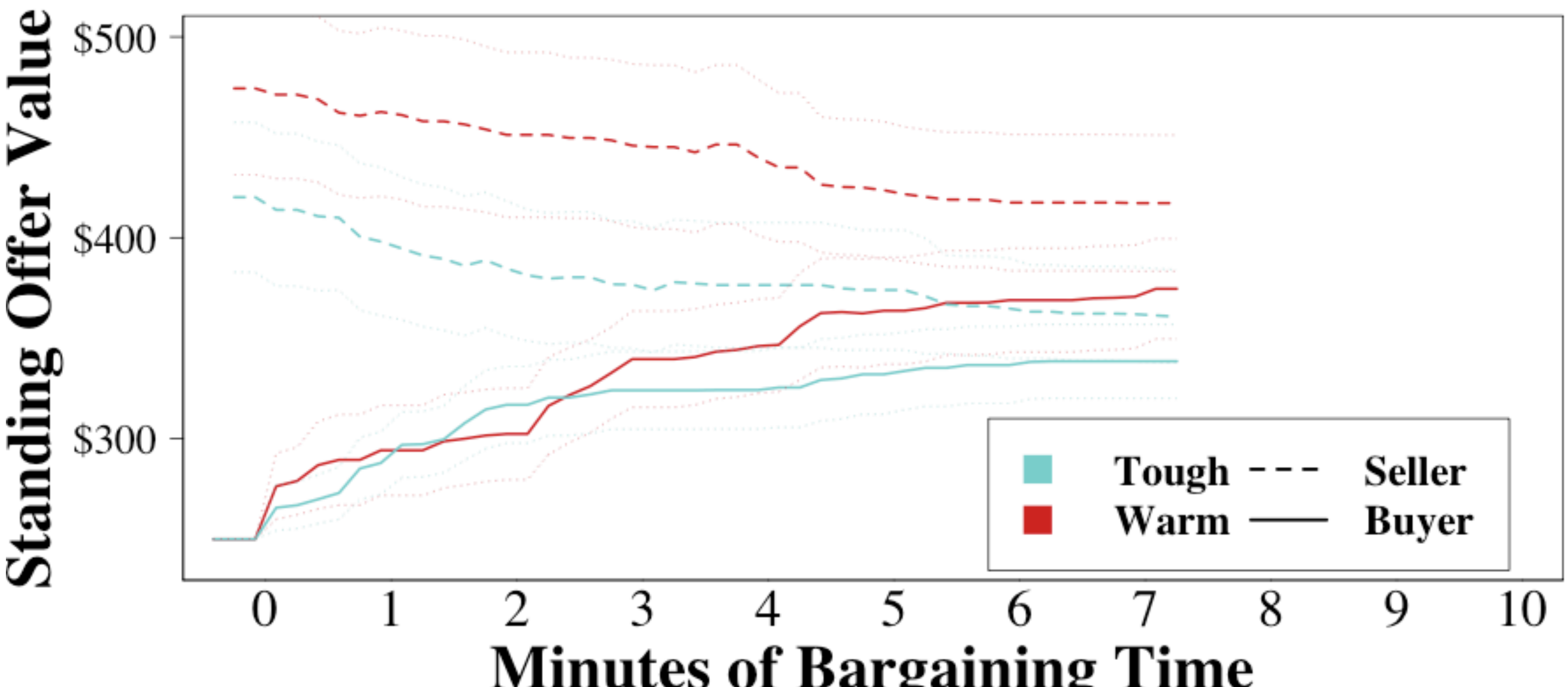


Communicating Warmth in …

(Jeong, Minson, Yeomans & Gino, 2018)

Warm offers earn worse final deals

- incentivized lab experiment with dyads
- driven by early seller concessions

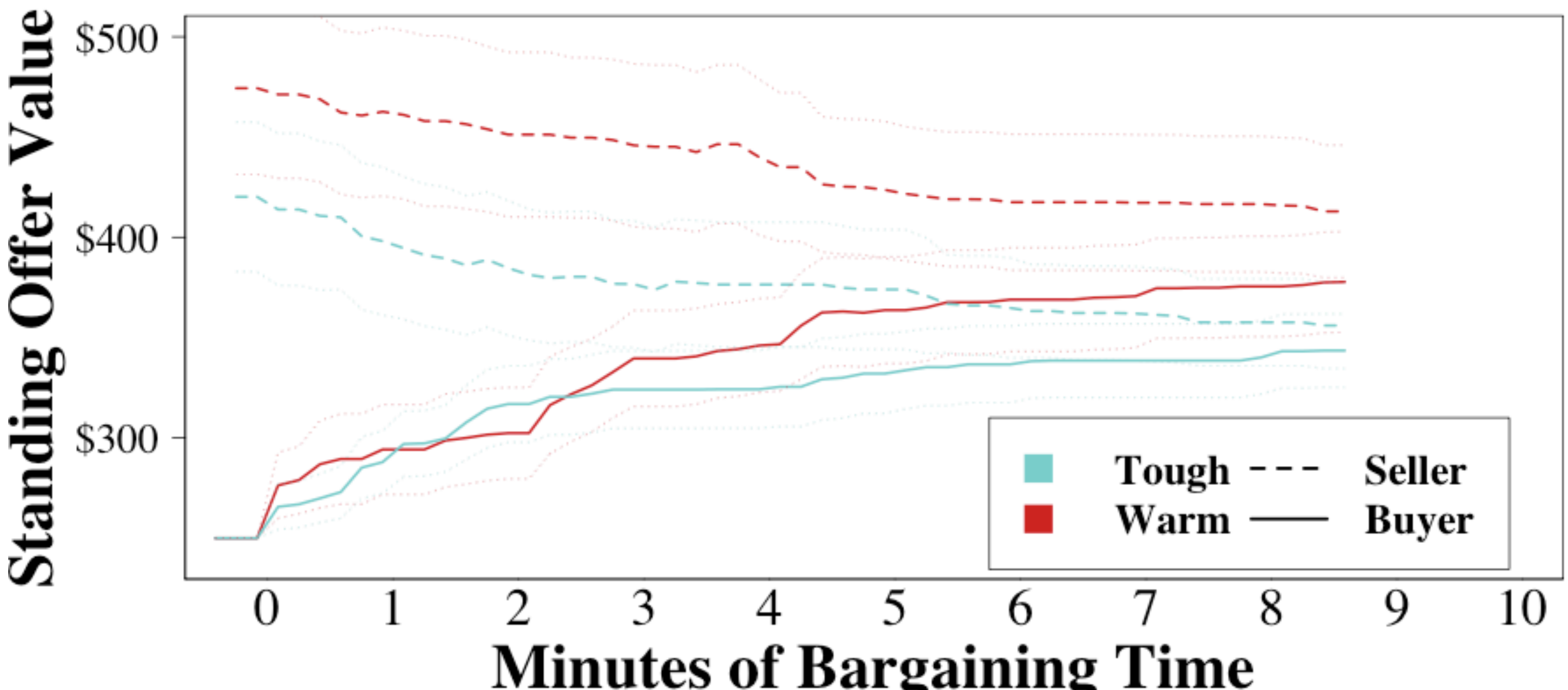


Communicating Warmth in …

(Jeong, Minson, Yeomans & Gino, 2018)

Warm offers earn worse final deals

- incentivized lab experiment with dyads
- driven by early seller concessions

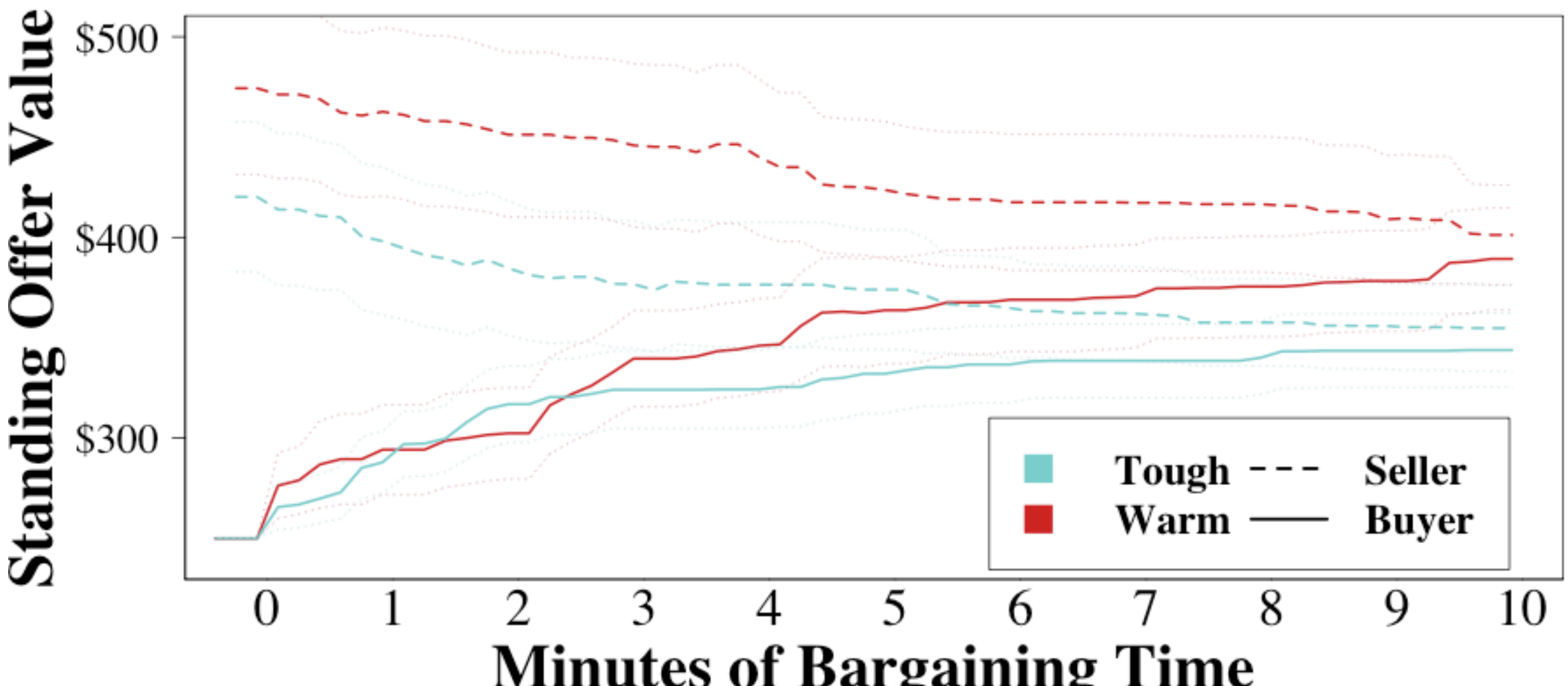


Communicating Warmth in …

(Jeong, Minson, Yeomans & Gino, 2018)

Warm offers earn worse final deals

- incentivized lab experiment with dyads
- driven by early seller concessions



How is Warmth Communicated?

How is Warmth Communicated?

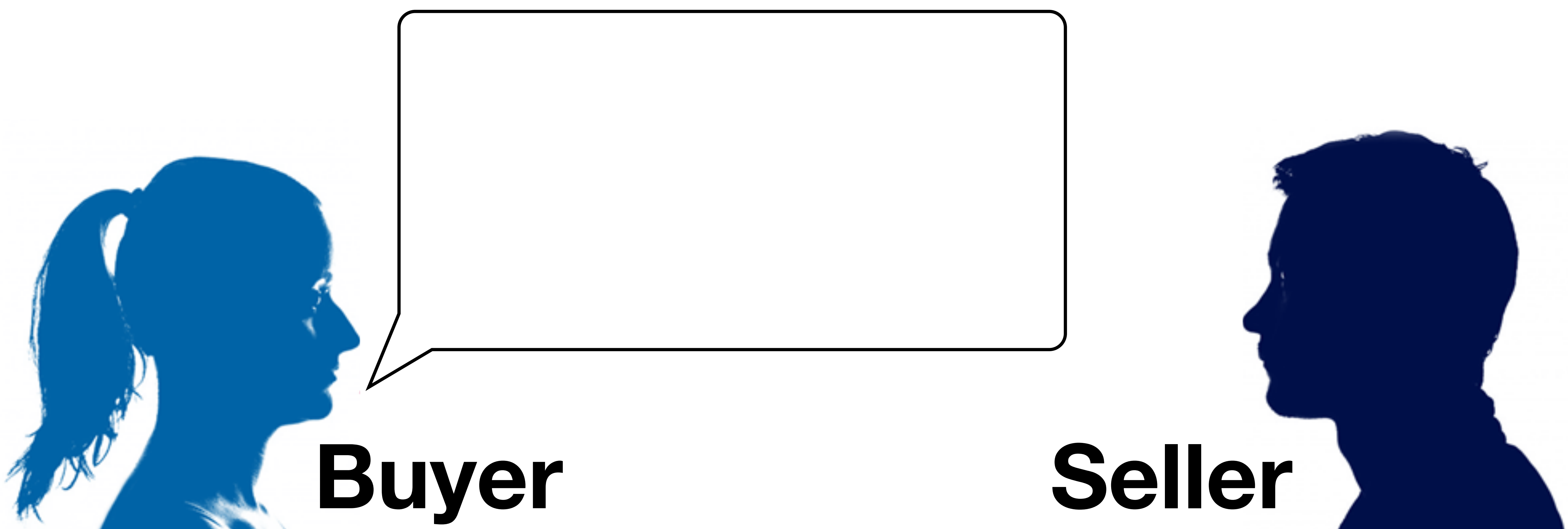
Buyer

How is Warmth Communicated?

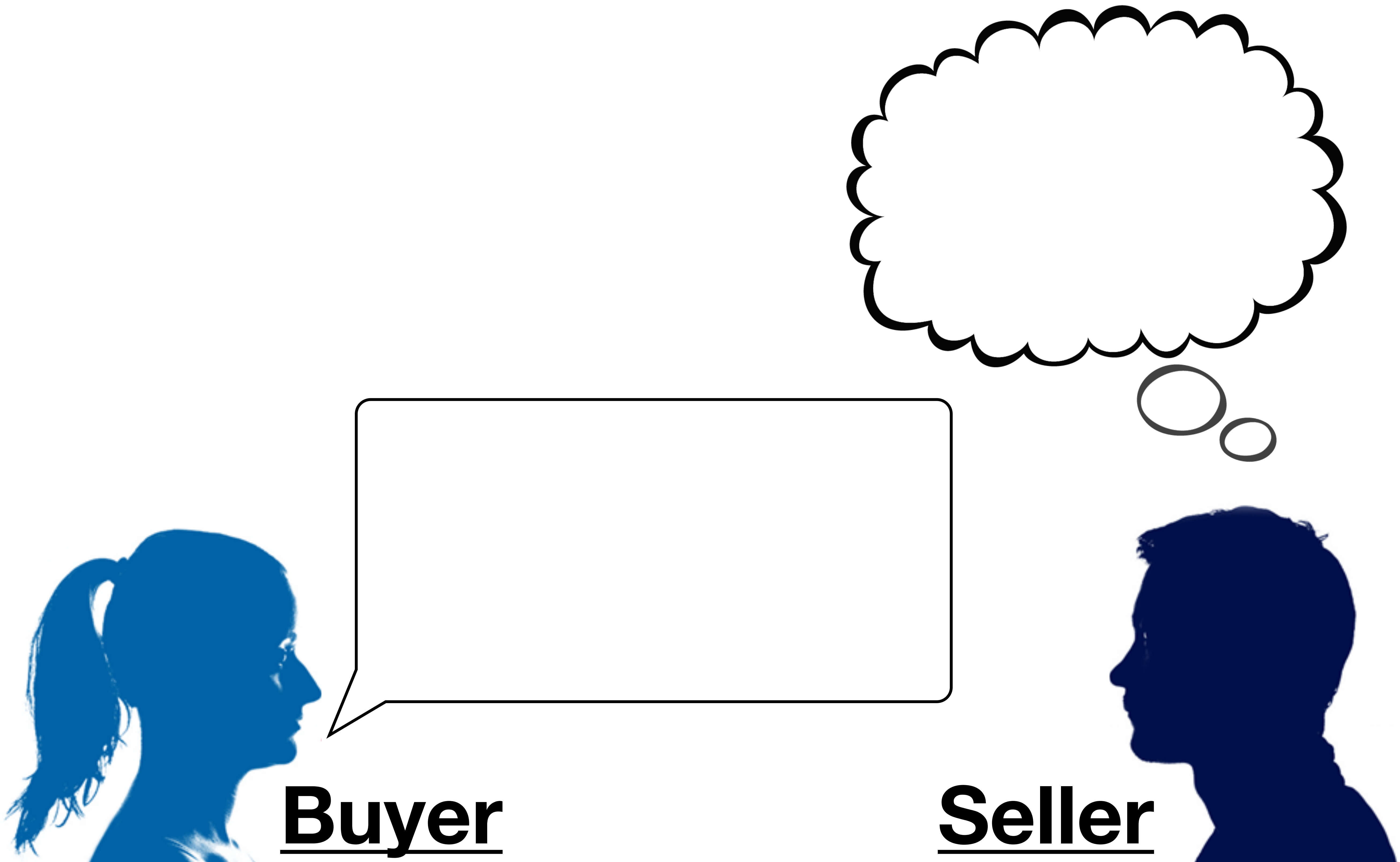
Buyer

Seller

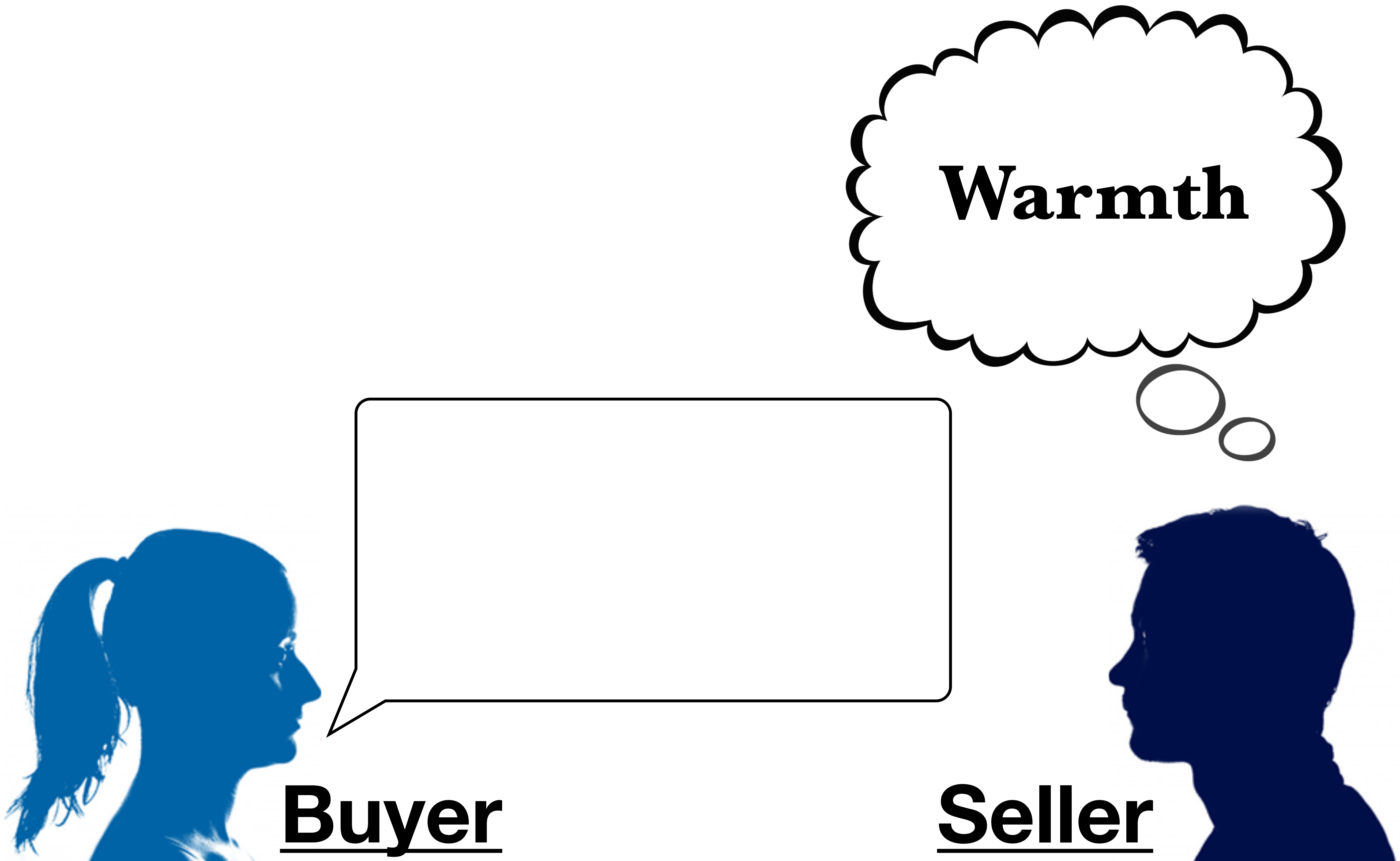
How is Warmth Communicated?



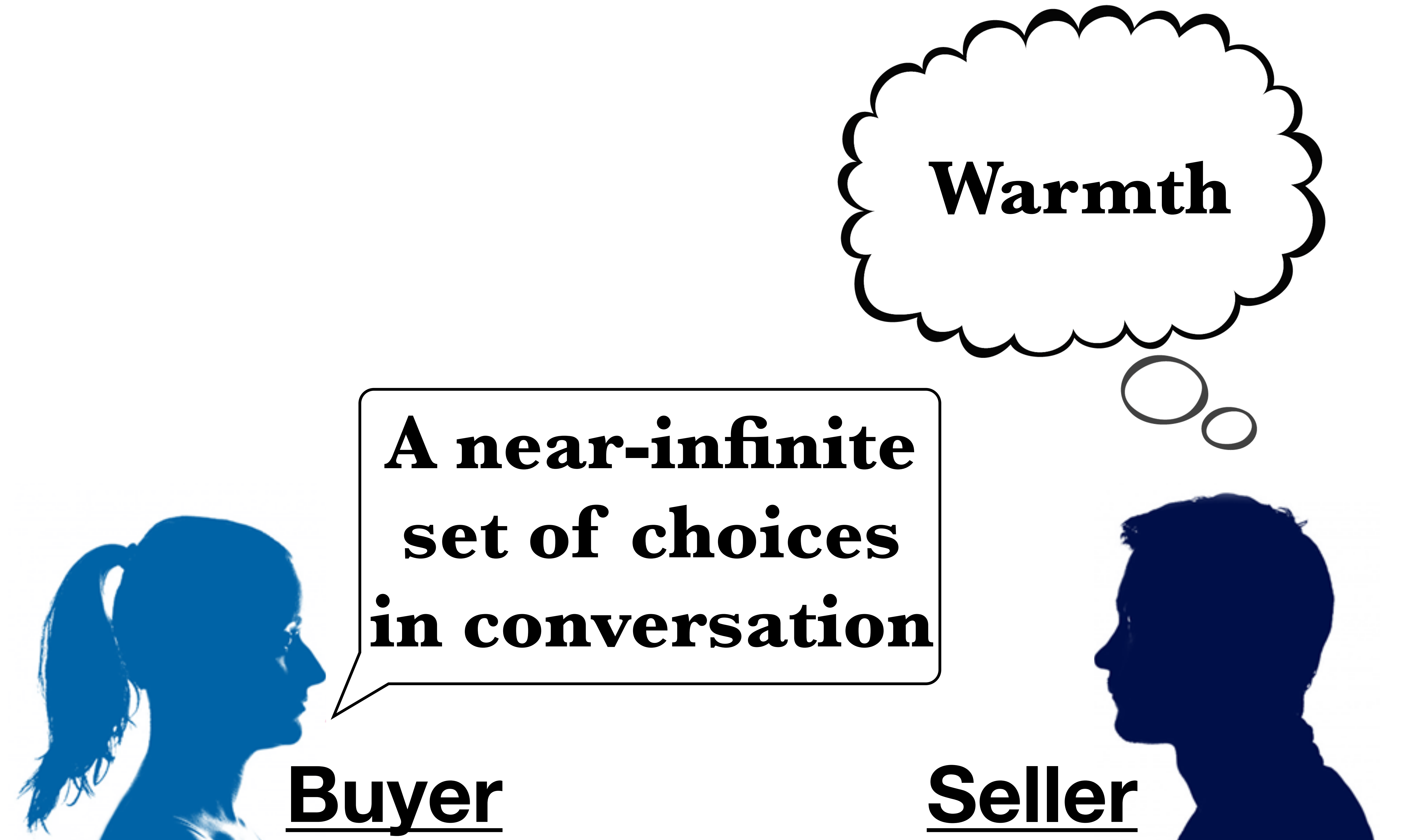
How is Warmth Communicated?



How is Warmth Communicated?



How is Warmth Communicated?



A Linguistic Model of Warmth

1. Randomly assign negotiation strategies

A Linguistic Model of Warmth

1. Randomly assign negotiation strategies

2. Transcribe those negotiations

A Linguistic Model of Warmth

1. Randomly assign negotiation strategies

2. Transcribe those negotiations

3. Train an algorithm to detect the strategy automatically

A Linguistic Model of Warmth

1. Randomly assign negotiation strategies

2. Transcribe those negotiations

3. Train an algorithm to detect the strategy automatically

We want to build an algorithm that is:

- scaleable

A Linguistic Model of Warmth

1. Randomly assign negotiation strategies

2. Transcribe those negotiations

3. Train an algorithm to detect the strategy automatically

We want to build an algorithm that is:

- scaleable
- interpretable

A Linguistic Model of Warmth

1. Randomly assign negotiation strategies

2. Transcribe those negotiations

3. Train an algorithm to detect the strategy automatically

We want to build an algorithm that is:

- scaleable
- interpretable
- valid

A Linguistic Model of Warmth

$$\hat{y} = a_0 \qquad\qquad\qquad\qquad + e$$

A Linguistic Model of Warmth

$$\hat{y} = a_0 + \quad x_1 + \quad x_2 + \quad x_3 + \ldots + e$$

Theory-Driven Feature Curation

Select set of observables ***x*** from literature

A Linguistic Model of Warmth

$$\hat{y} = a_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + e$$

Theory-Driven Feature Curation

Select set of observables $\boldsymbol{x}$ from literature

Empirical Feature Estimation

Determine $\boldsymbol{\beta}$ weights from ground-truth data

A Linguistic Model of Warmth

$$\hat{y} = a_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + e$$

Theory-Driven Feature Curation

Select set of observables ***x*** from literature

Empirical Feature Estimation

Determine **β** weights from ground-truth data

Actuarial Model of Politeness

(Meehl et al., 1954; Dawes, 1979; Grove et al., 1989)

Communicating Warmth in …

mTurkers can write phone offers! (n = 355)

- encouragement design - tough vs. warm instructions
- “ground truth” is treatment effect

Message	Condition
Hi, since this is a...	0
Hi, I saw your ad...	1
Hey, I hope you are...	1
Hi, I am interested...	0
Hi, I am interested...	0
I saw your add on...	0

Communicating Warmth in …

mTurkers can write phone offers! (n = 355)

- encouragement design - tough vs. warm instructions
- "ground truth" is treatment effect

Message	Condition
Hi, since this is a...	0
Hi, I saw your ad...	1
Hey, I hope you are...	1
Hi, I am interested...	0
Hi, I am interested...	0
I saw your add on...	0

We want to build an algorithm that is:

- interpretable
- valid
- scaleable

Communicating Warmth in …

mTurkers can write phone offers! (n = 355)

- encouragement design - tough vs. warm instructions
- "ground truth" is treatment effect

Message	Condition
Hi, since this is a...	0
Hi, I saw your ad...	1
Hey, I hope you are...	1
Hi, I am interested...	0
Hi, I am interested...	0
I saw your add on...	0

We want to build an algorithm that is:

- interpretable

- valid

- scaleable

Communicating Warmth in …

mTurkers can write phone offers! (n = 355)

- encouragement design - tough vs. warm instructions
- “ground truth” is treatment effect

Hello, I am interested in purchasing this phone for work purposes but unfortunately have a spending limit of $115 per my company's budget. Is there any way that you would be willing to work with me on this? I really appreciate your time and consideration.   Thank You, Ashleigh.

Communicating Warmth in …

mTurkers can write phone offers! (n = 355)

- encouragement design - tough vs. warm instructions
- “ground truth” is treatment effect

Hello, Hope you are doing well. I recently saw your posting for the iPhone 6 Plus for sale and I am interested. I would love to come by and buy it from you today. I can pay you in cash. Would you be willing to sell it to me for $115 paid in cash today? Thanks so much. I really appreciate it.

Communicating Warmth in …

mTurkers can write phone offers! (n = 355)

- encouragement design - tough vs. warm instructions
- “ground truth” is treatment effect

Come on. The price you are offering on a product that ISN'T NEW is unreasonable. Now, I for one am very interested in getting this item. BUT, I will only pay $115. I am not paying a penny more.

Communicating Warmth in …

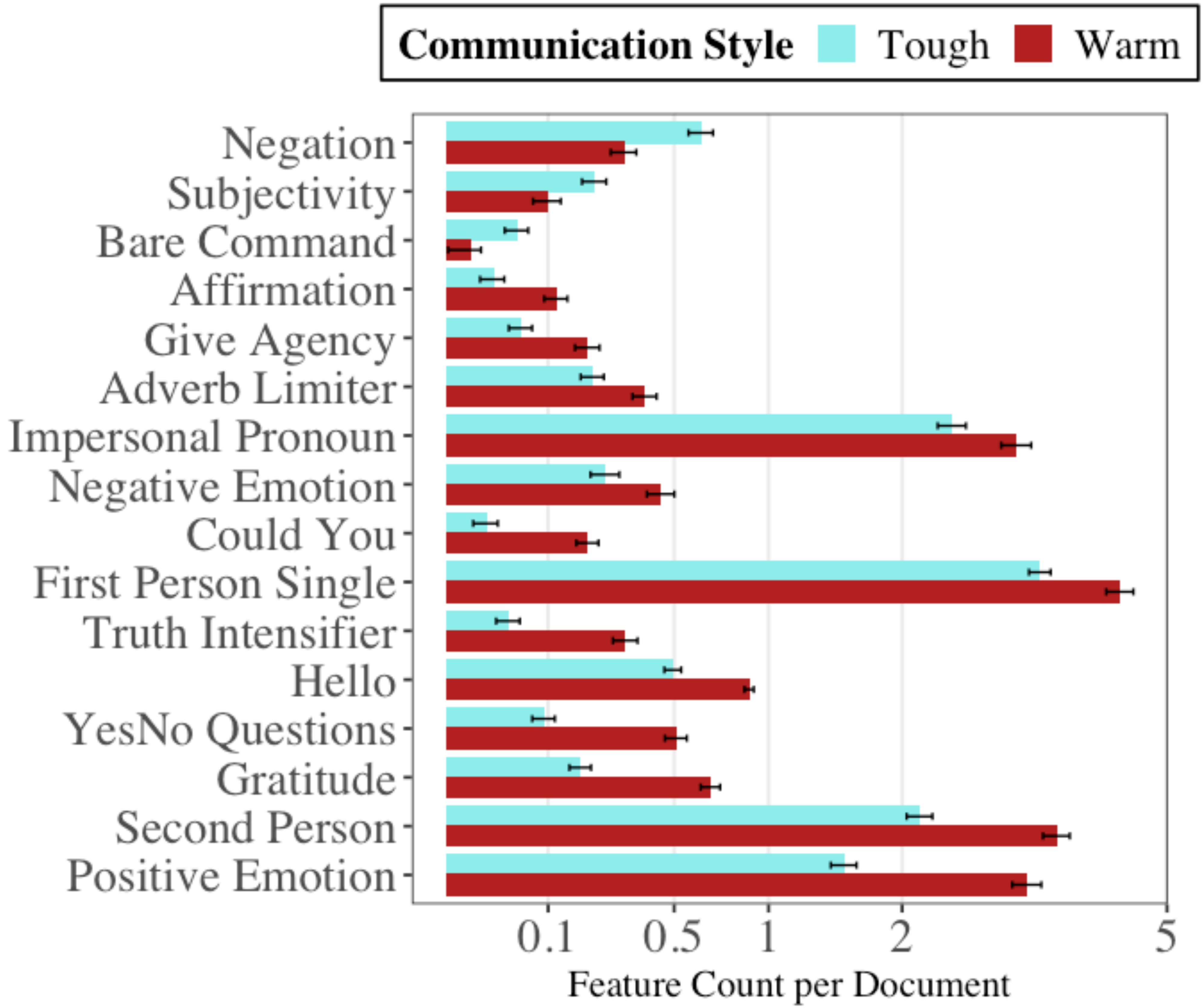
mTurkers can write phone offers! (n = 355)

- encouragement design - tough vs. warm instructions
- “ground truth” is treatment effect

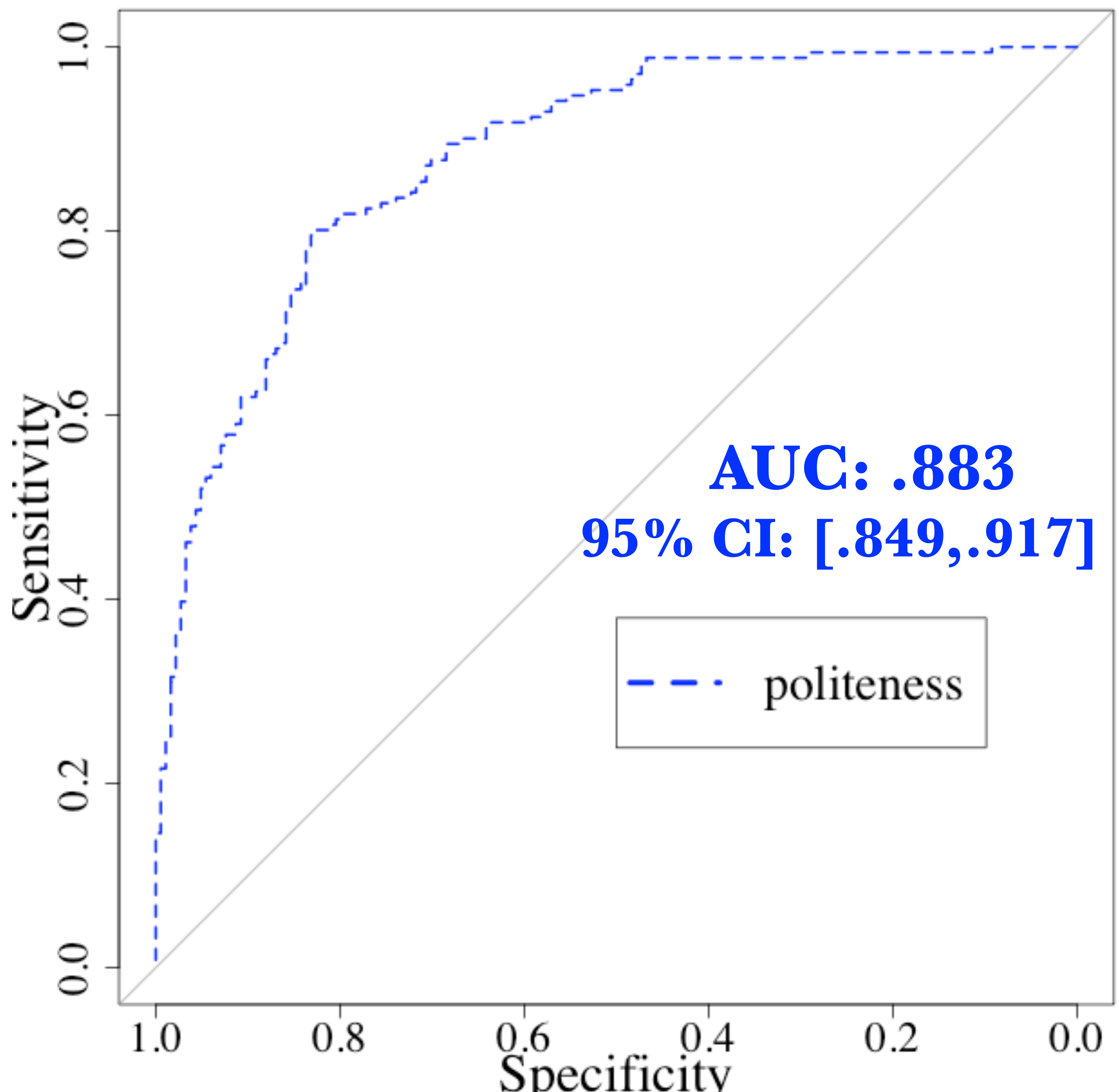
I will buy the phone as is for $115. I don't want to pay more than the amount that I stated. If you accept my price please contact me within 24 hours. If I don't hear from you in the next 24 hours I will take it that you will accept my price.

Communicating Warmth in …

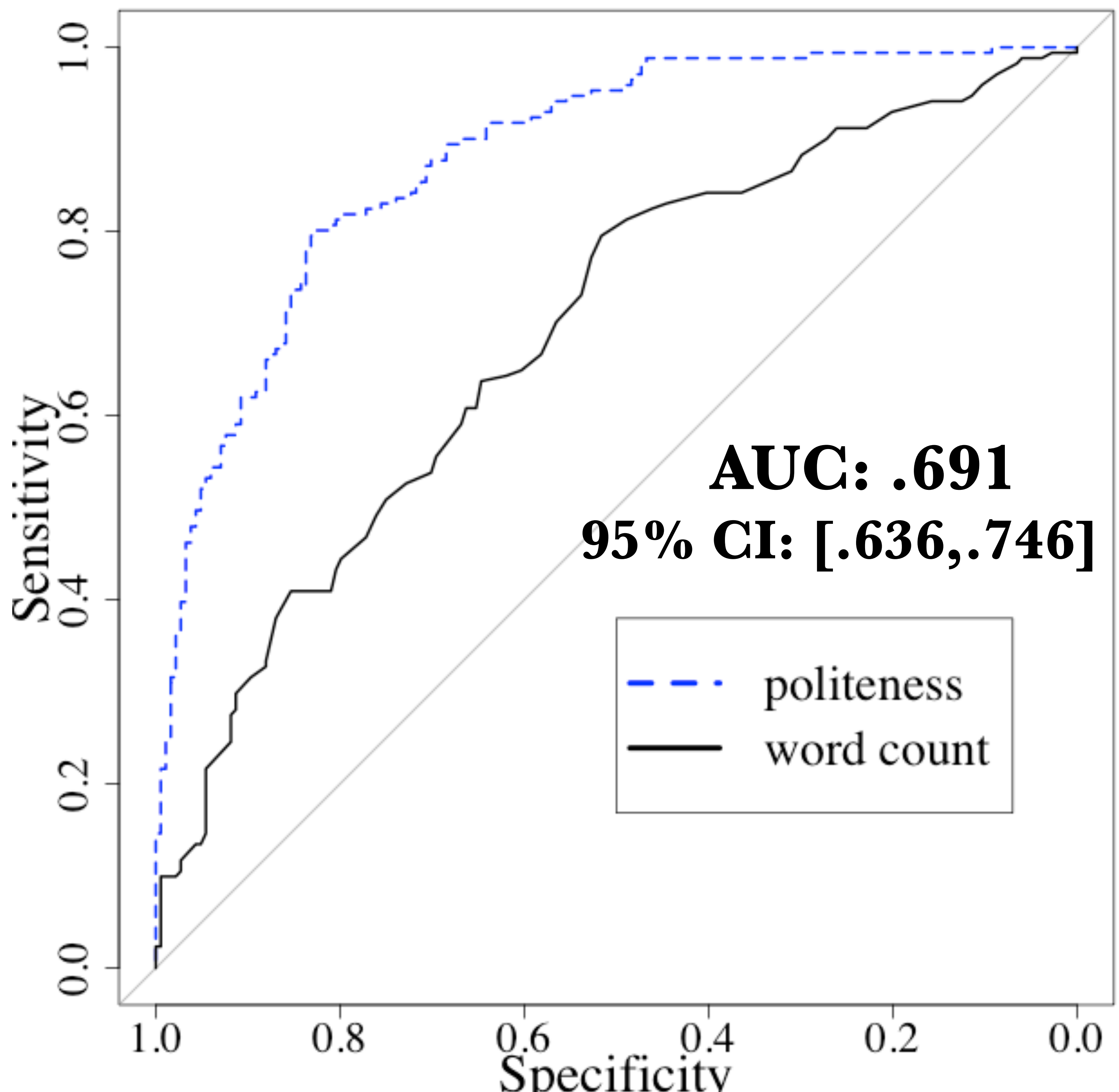
Communicating Warmth in …



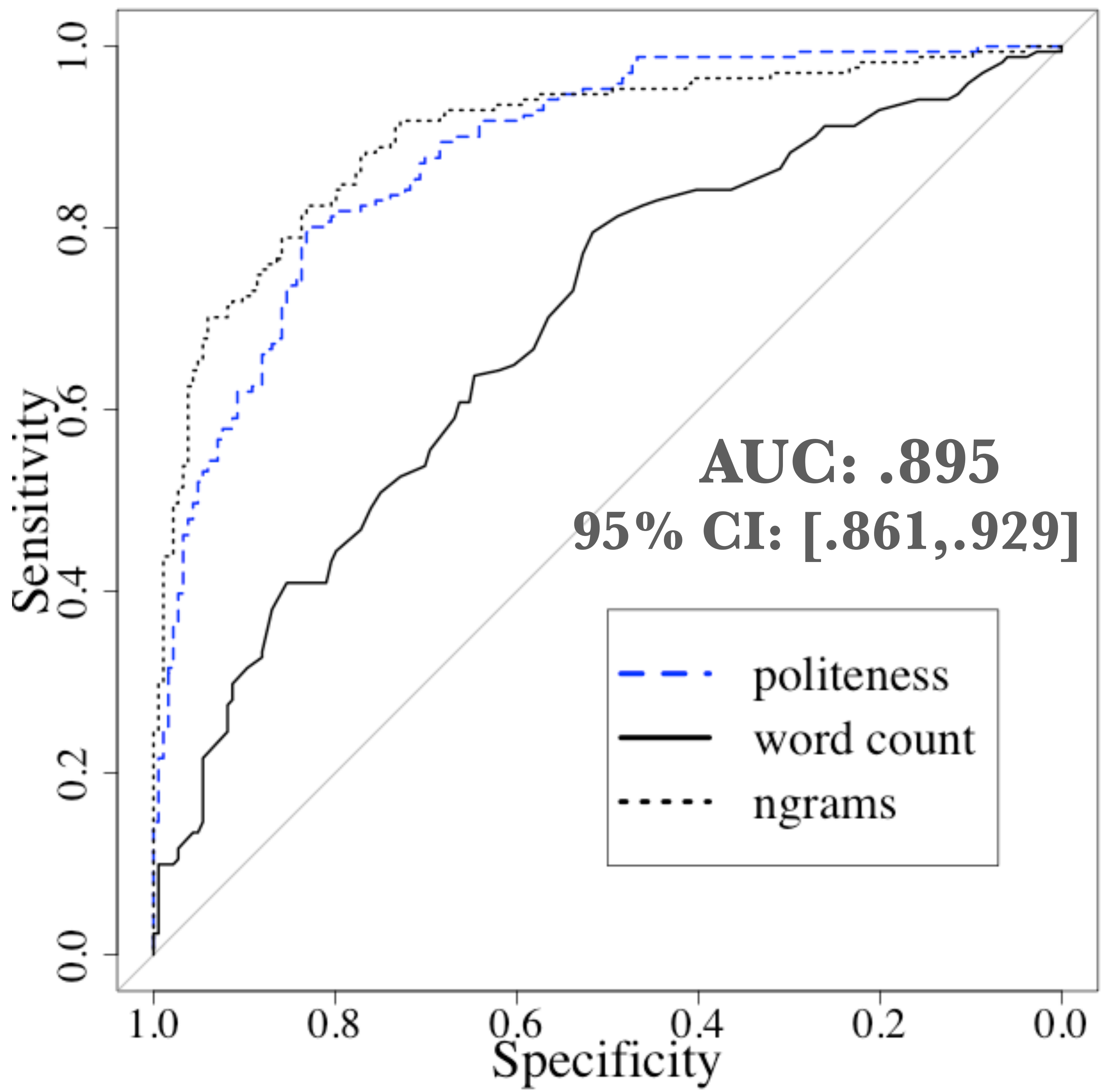
Communicating Warmth in …



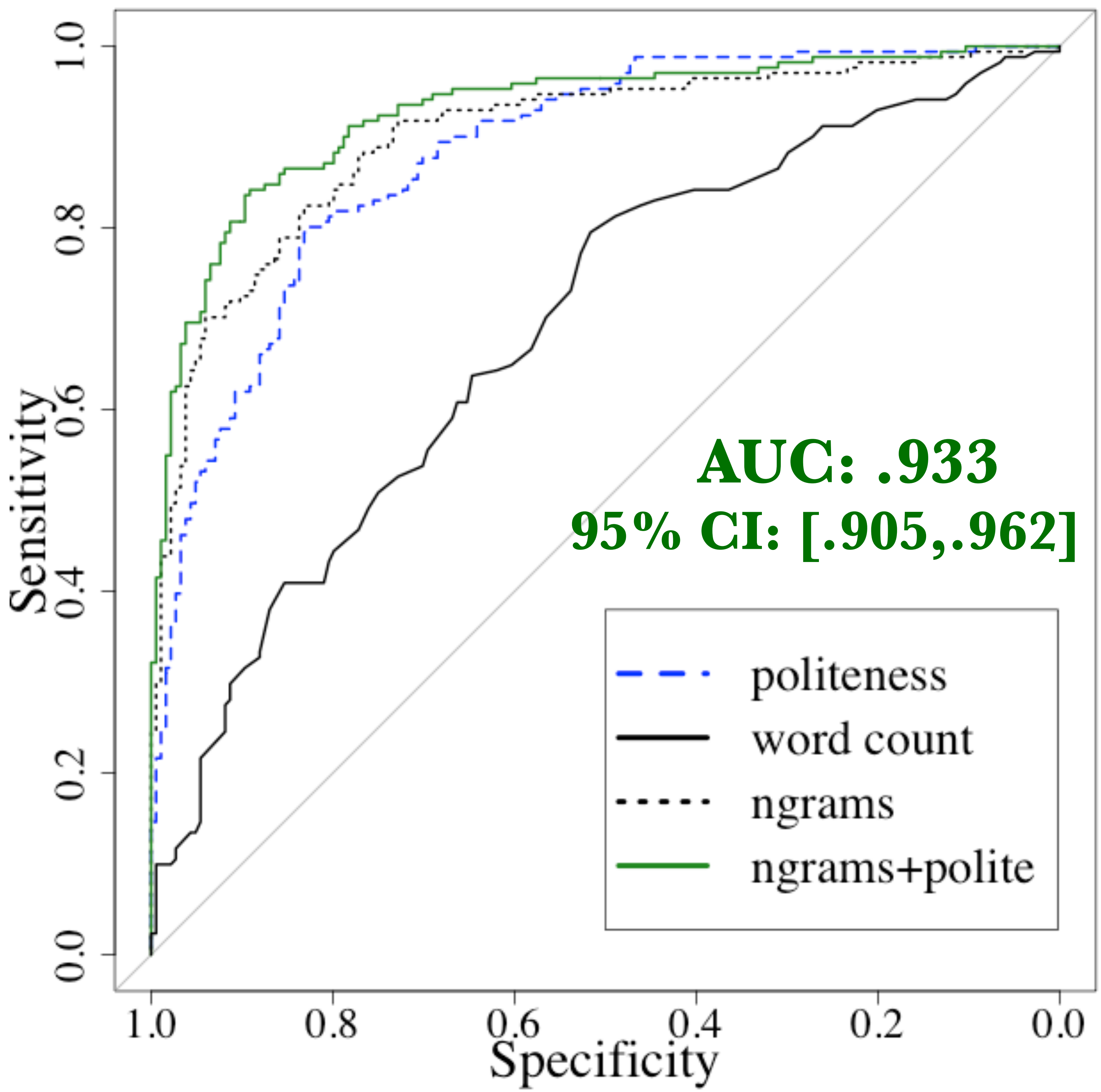
Communicating Warmth in …



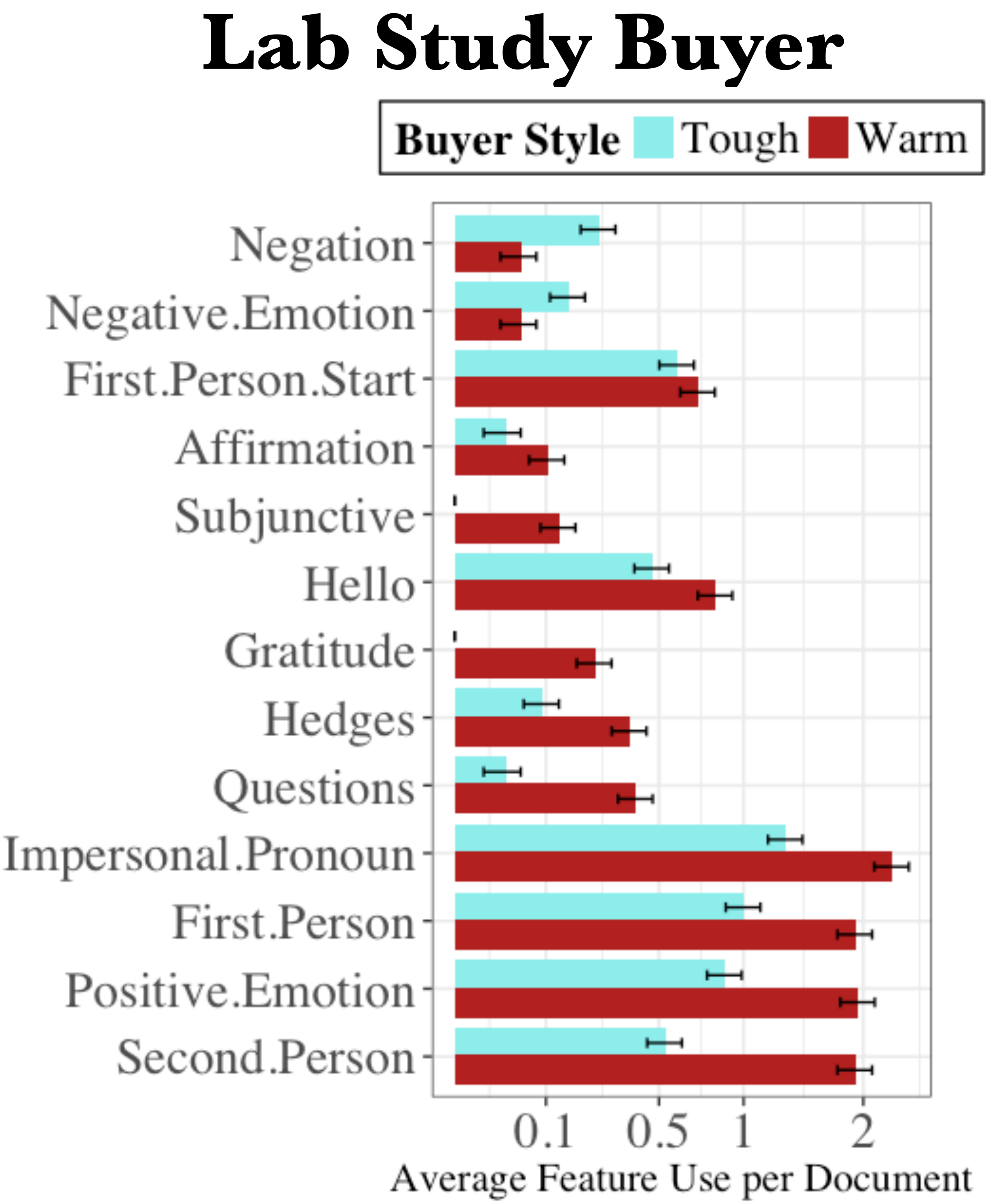
Communicating Warmth in …



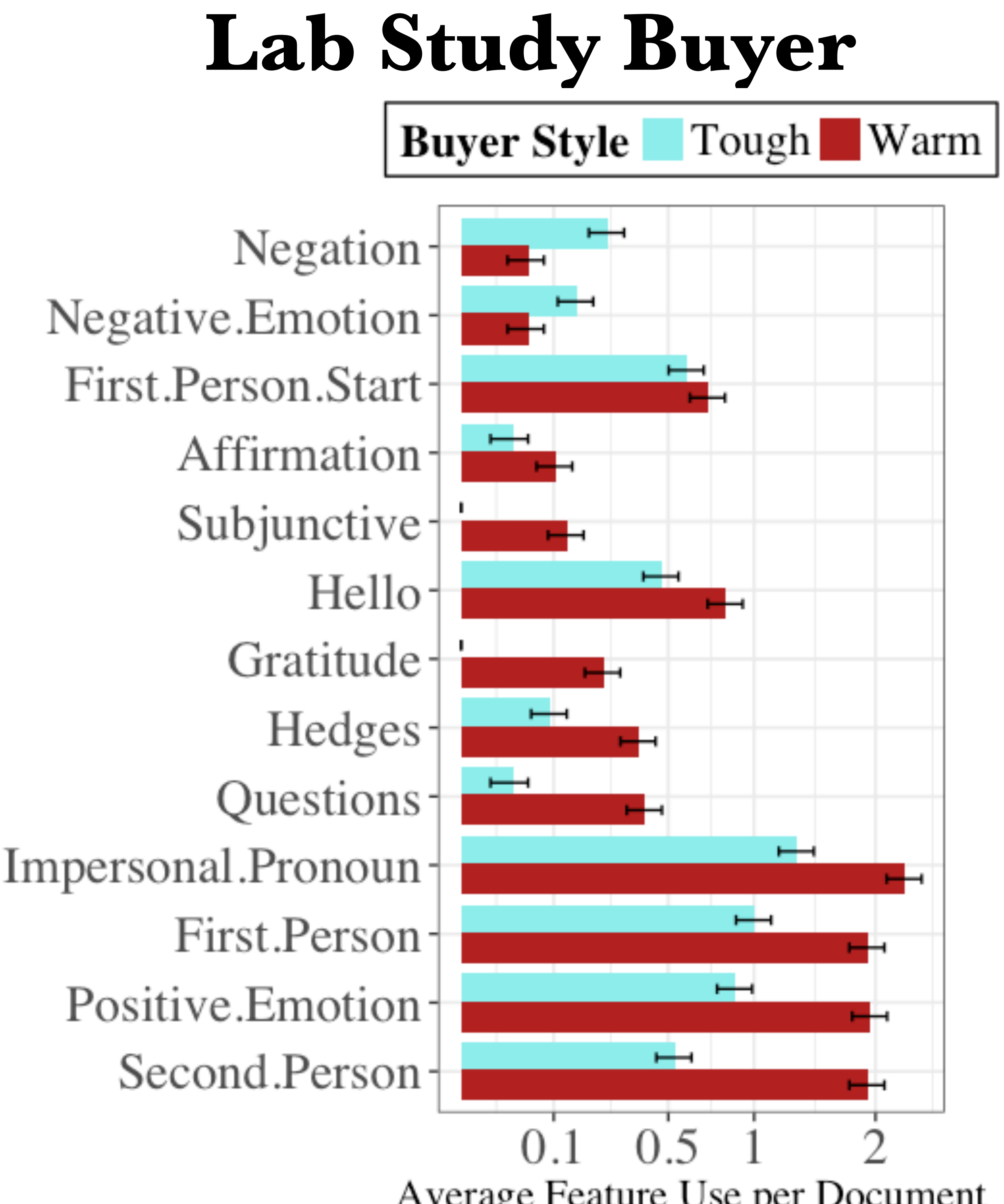
Communicating Warmth in …



Communicating Warmth in …

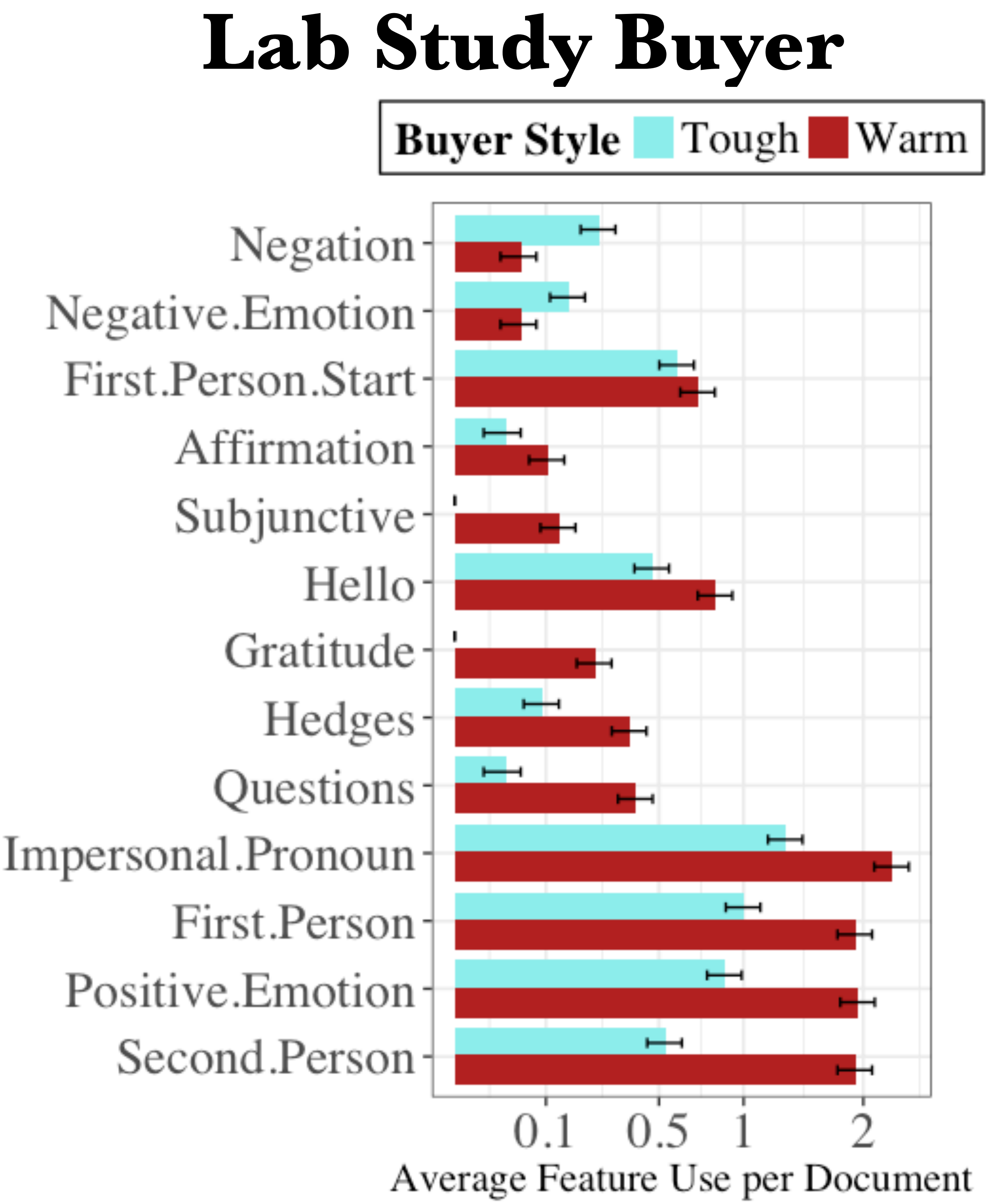


Communicating Warmth in …

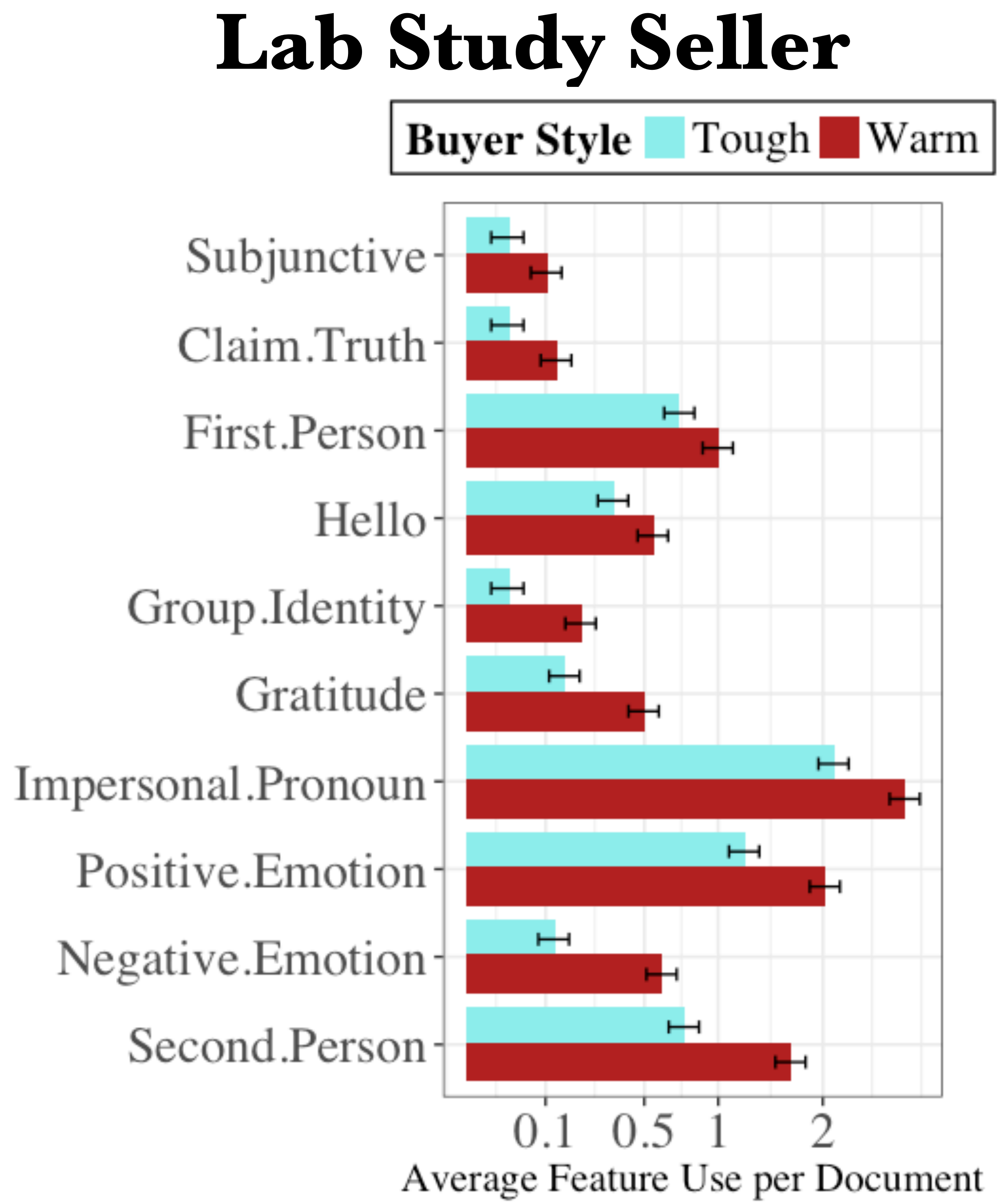


AUC = .887 [.810, .963]

Communicating Warmth in …



AUC = .887 [.810, .963]



AUC = .764 [.651, .877]

Markers of Politeness

Goffman, 1967; Lakoff, 1973; Brown & Levinson, 1987

Markers of Politeness

Goffman, 1967; Lakoff, 1973; Brown & Levinson, 1987

“universal dimension in human communication”

Markers of Politeness

Goffman, 1967; Lakoff, 1973; Brown & Levinson, 1987

Positive Politeness

Markers of Politeness

Goffman, 1967; Lakoff, 1973; Brown & Levinson, 1987

Positive Politeness

Bolstering listener's self-image

Markers of Politeness

Goffman, 1967; Lakoff, 1973; Brown & Levinson, 1987

Positive Politeness

Bolstering listener's self-image

+ Gratitude
+ Complements
+ Open-ended Questions

+ In-Group Identity
+ Formal Titles
- Informal Titles

Markers of Politeness

Goffman, 1967; Lakoff, 1973; Brown & Levinson, 1987

Positive Politeness

Bolstering listener's self-image

+ Gratitude
+ Complements
+ Open-ended Questions
+ In-Group Identity
+ Formal Titles
- Informal Titles

Negative Politeness

Upholding listener's autonomy

Markers of Politeness

Goffman, 1967; Lakoff, 1973; Brown & Levinson, 1987

Positive Politeness

Bolstering listener's self-image

+ Gratitude
+ Complements
+ Open-ended Questions
+ In-Group Identity
+ Formal Titles
- Informal Titles

Negative Politeness

Upholding listener's autonomy

+ Minimizing
+ Indirect requests
+ Apologies
+ Hedging
- Negations
- Bare Commands

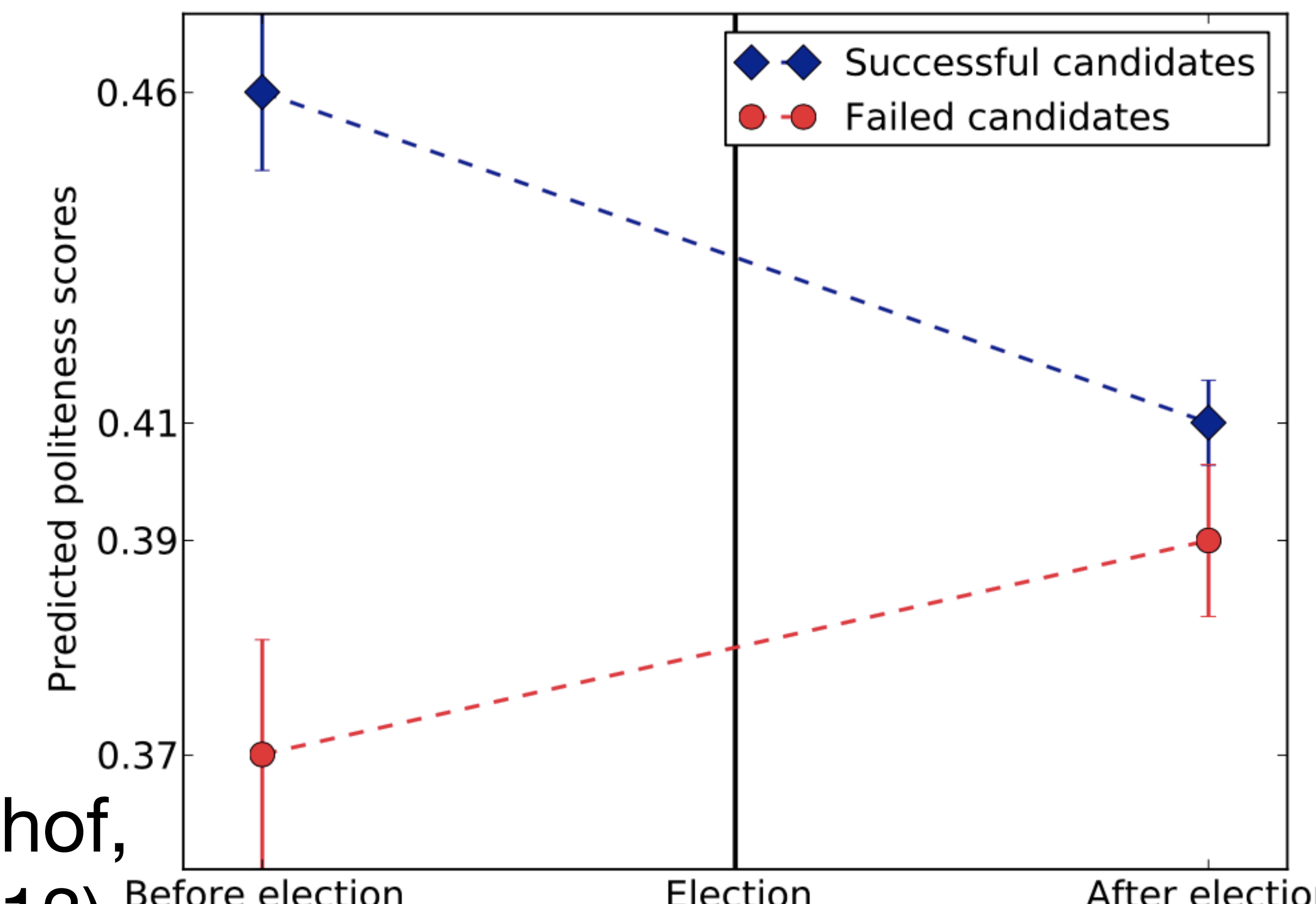
Markers of Politeness

A Computational Approach to Politeness with Applications to Social Factors

(Danescu-Niculescu-Mizil, Sudhof, Jurafsky, Leskovec & Potts, 2013)

Markers of Politeness

A Computational Approach to Politeness with Applications to Social Factors



(Danescu-Niculescu-Mizil, Sudhof, Jurafsky, Leskovec & Potts, 2013)

Markers of Politeness

Language from Police Body Camera Footage shows Racial Disparities in Officer Respect

(Voigt, Camp, Prabakharan, Hamilton, Hetey, Griffiths, Jurgens, Jurafsky & Eberhart, 2017)

Markers of Politeness

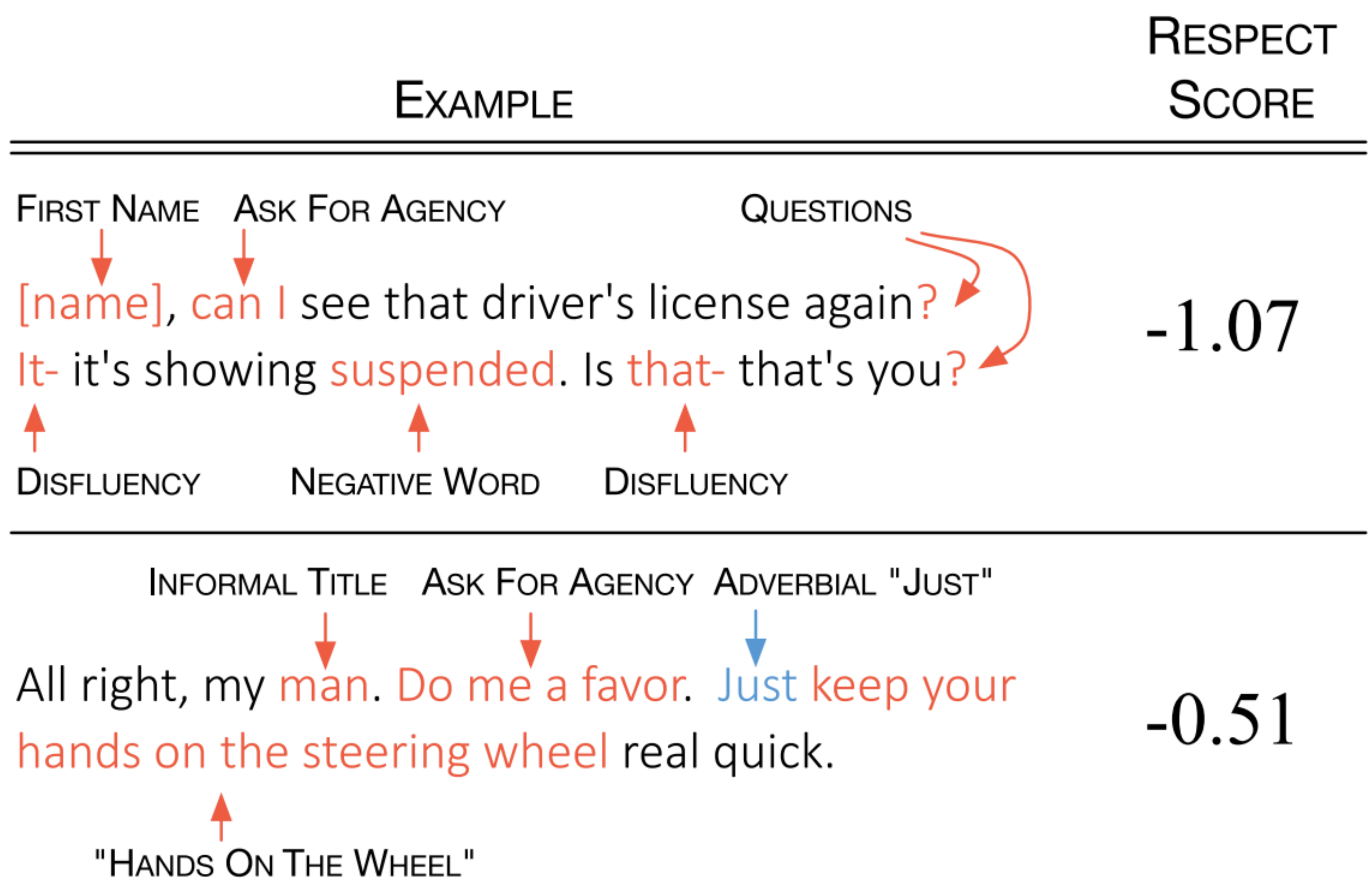
Language from Police Body Camera Footage shows Racial Disparities in Officer Respect

(Voigt, Camp, Prabakharan, Hamilton, Hetey, Griffiths, Jurgens, Jurafsky & Eberhart, 2017)

Markers of Politeness

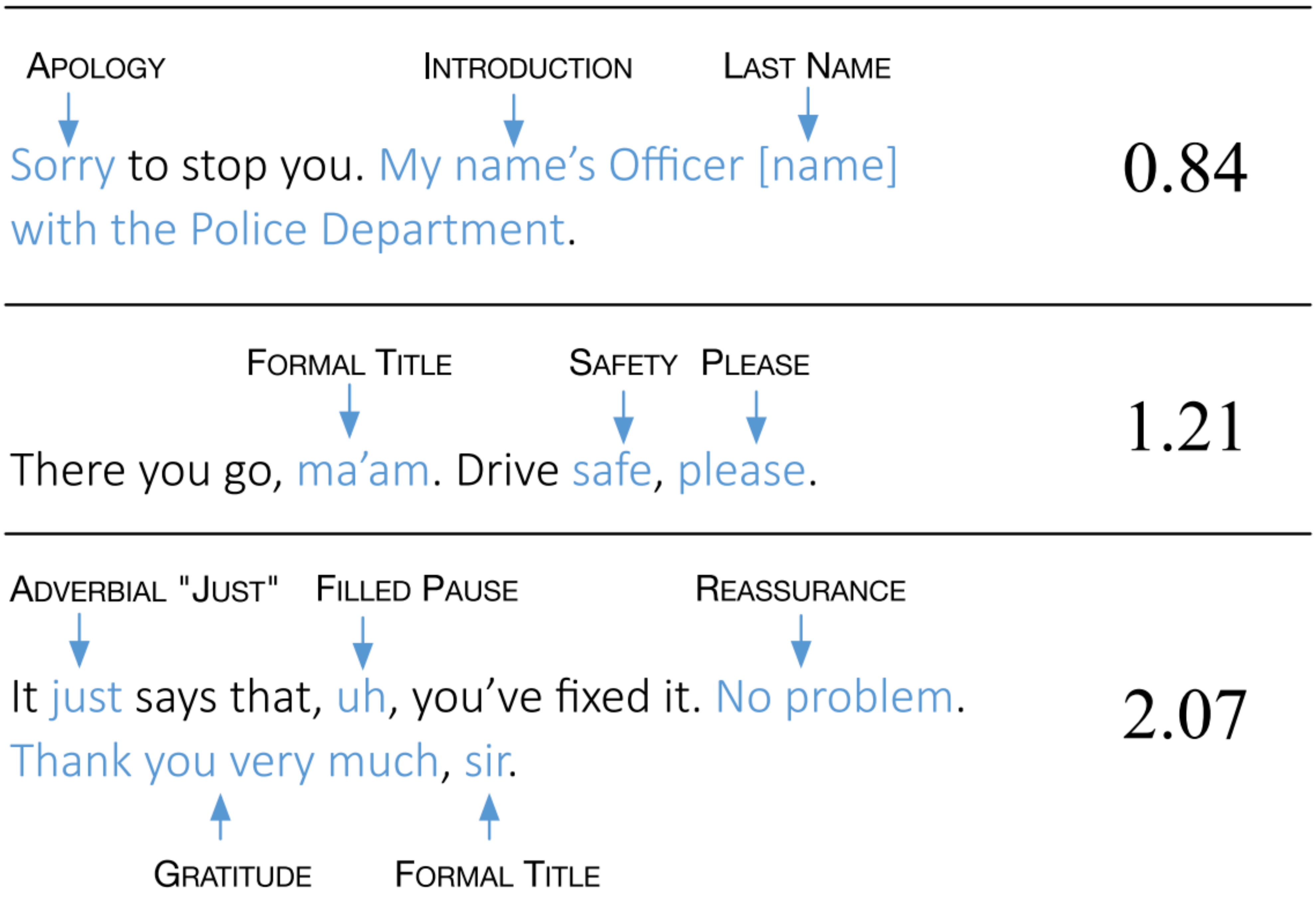
(Voigt et al., 2017)

Markers of Politeness



(Voigt et al., 2017)

Markers of Politeness



(Voigt et al., 2017)

Markers of Politeness

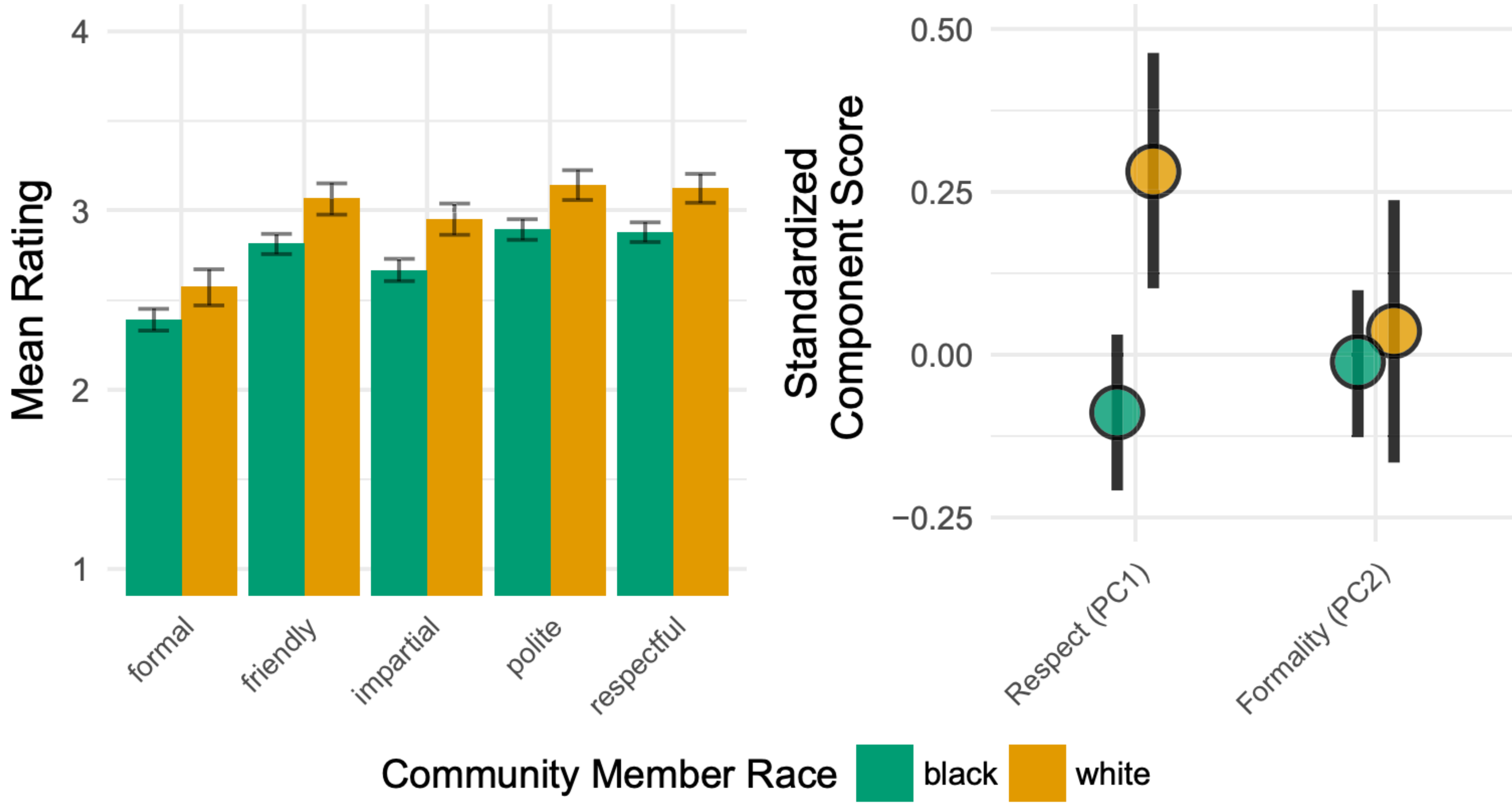
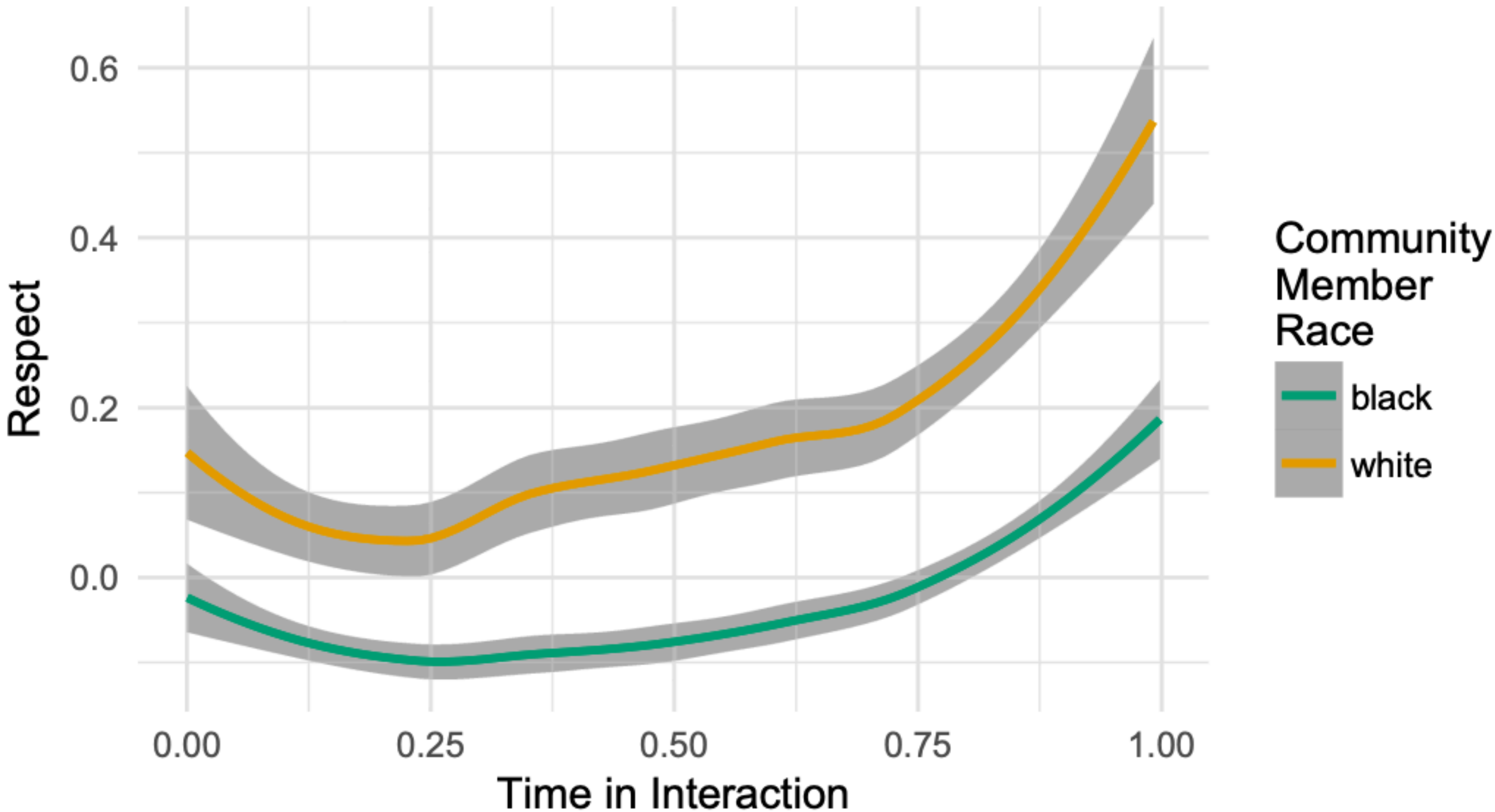


Fig. 1. (*Left*) Differences in raw participant ratings between interactions with black and white community members. (*Right*) When collapsed to two uncorrelated components, Respect and Formality, we find a significant difference for Respect but none for Formality. Error bars represent 95% confidence intervals. PC, principal component.

(Voigt et al., 2017)

Markers of Politeness



(Voigt et al., 2017)

Markers of Politeness

Apologies

Markers of Politeness

Apologies

"I apologize for my behavior"

Markers of Politeness

Apologies

"I apologize for my behavior"

"I apologise for my behaviour"

Markers of Politeness

Apologies

"I apologize for my behavior"

"I apologise for my behaviour"

"Not until you apologise for your behaviour"

Markers of Politeness

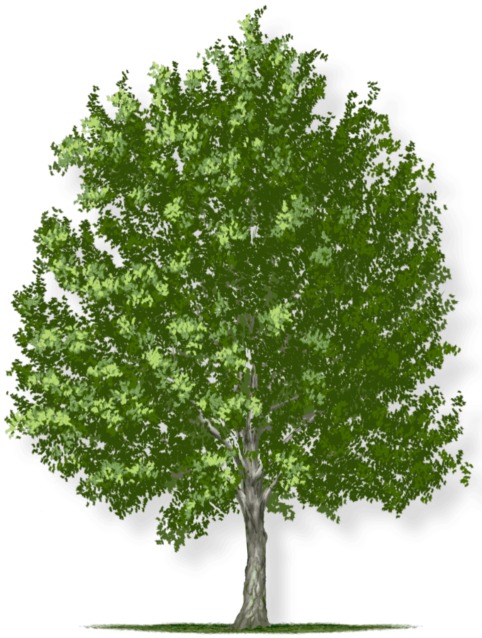
Apologies

"I apologize for my behavior"

"I apologise for my behaviour"

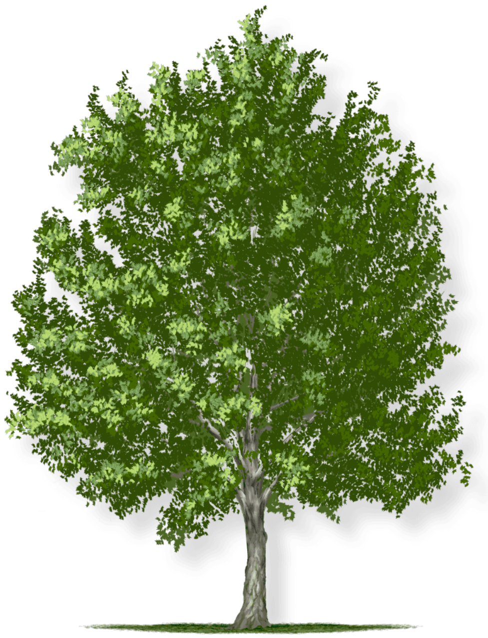
"Not until you apologise for your behaviour"

"I sincerely apologise for my behaviour"



Dependency Trees

I sincerely apologise for my behaviour



Dependency Trees

I sincerely apologise for my behaviour

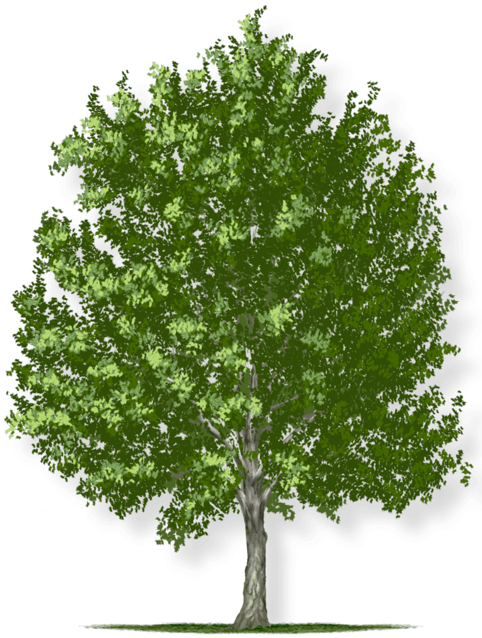
PRON ADV VERB ADP DET NOUN

Dependency Trees

The "root"

I sincerely apologise for my behaviour

PRON ADV VERB ADP DET NOUN

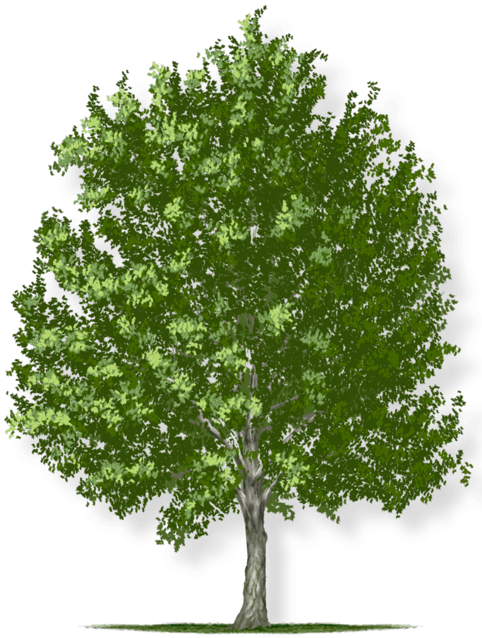


Dependency Trees

nsubj

I sincerely apologise for my behaviour

PRON ADV VERB ADP DET NOUN



Dependency Trees

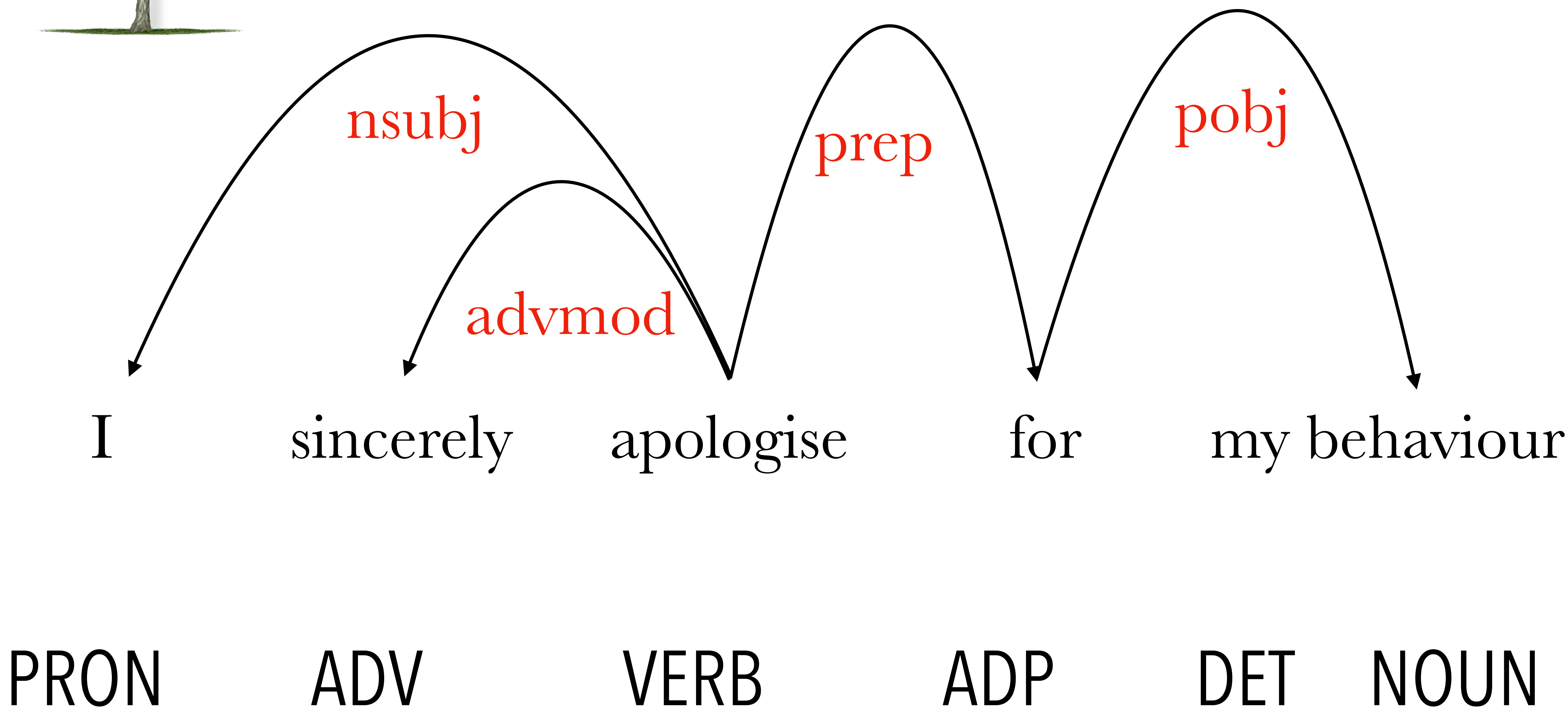
nsubj

advmod

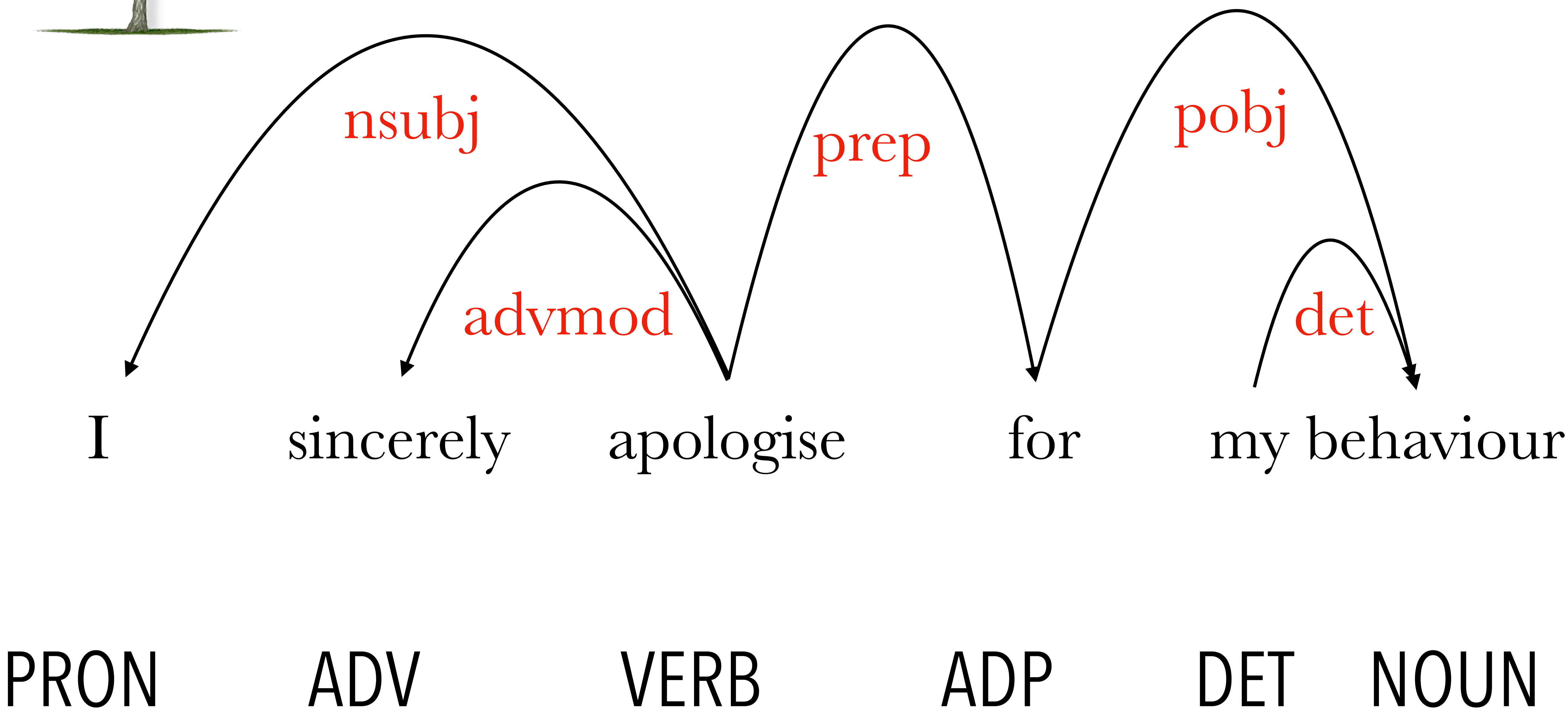
I sincerely apologise for my behaviour

PRON ADV VERB ADP DET NOUN

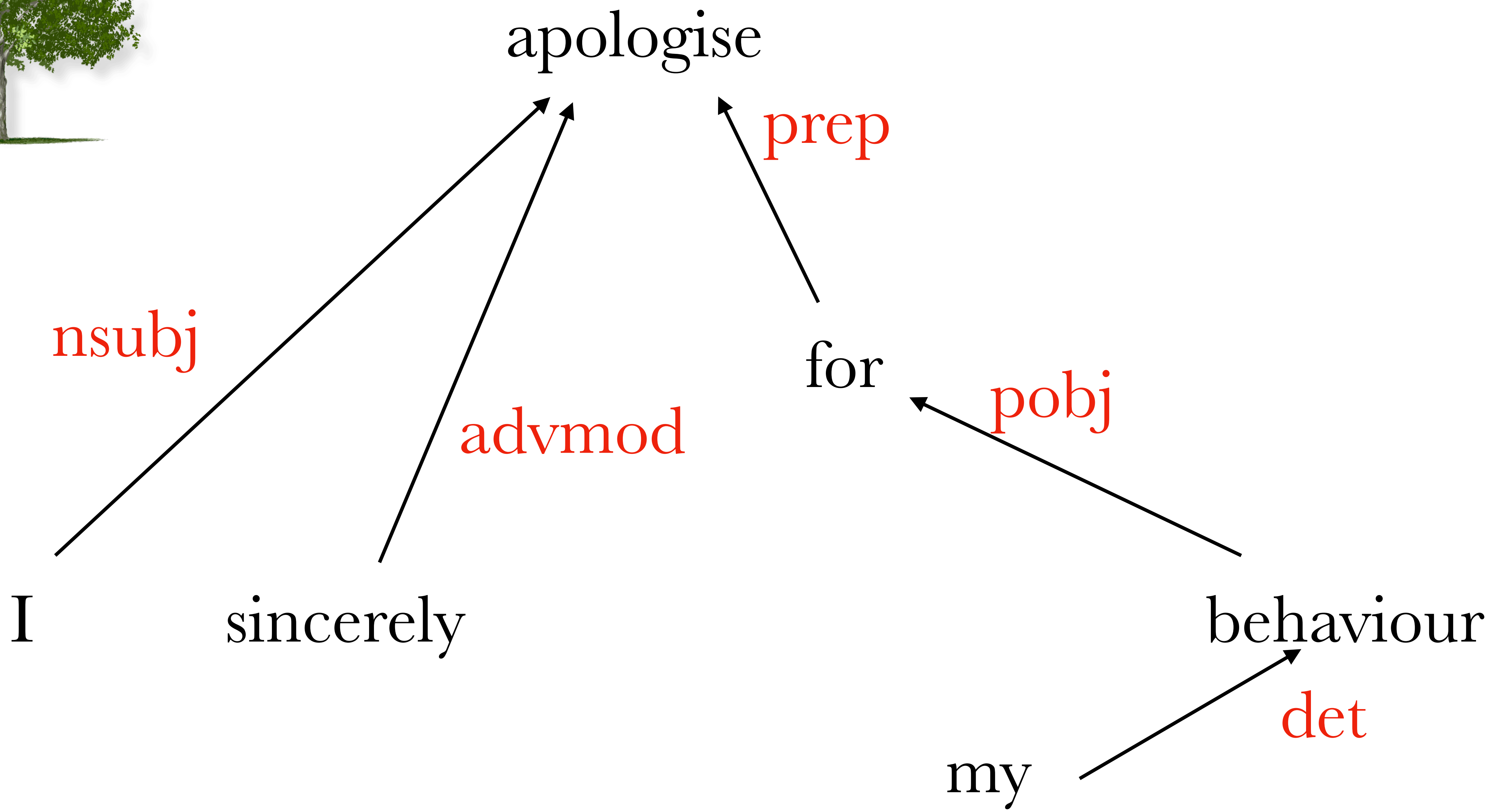
Dependency Trees

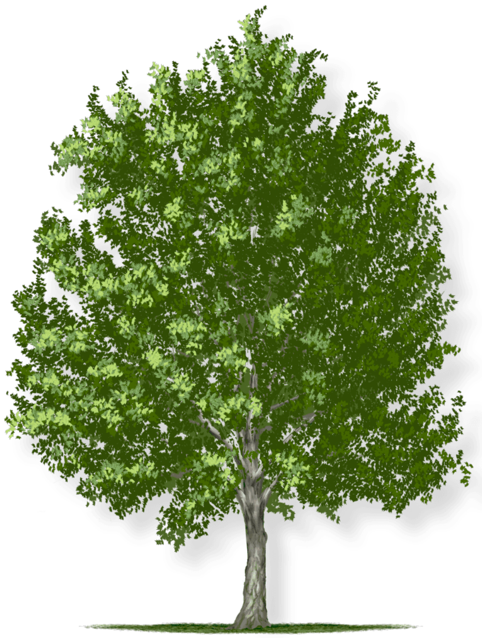


Dependency Trees



Dependency Trees





Dependency Trees

	Word
1	I
2	sincerely
3	apologise
4	for
5	my
6	behaviour

Dependency Trees

spaCy

	Word
1	I
2	sincerely
3	apologise
4	for
5	my
6	behaviour

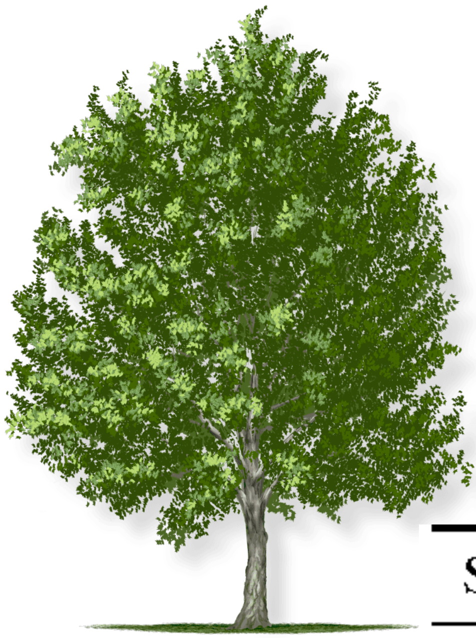
Dependency Trees

spaCy

	Word	Part of Speech	Head	Relation	Dependency Pair
1	I	PRON	3	nsubj	nsubj(apologize, I)
2	sincerely	ADV	3	advmod	advmod(apologize, sincerely)
3	apologise	VERB	3	ROOT	ROOT(--,apologise)
4	for	ADP	3	prep	prep(apologise,for)
5	my	DET	6	poss	det(behaviour, my)
6	behaviour	NOUN	4	pobj	pobj(for, behaviour)

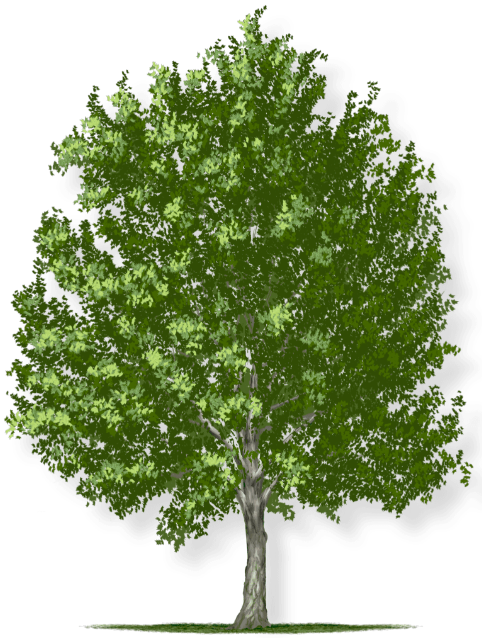
Dependency Trees

	Word	Part of Speech	Head	Relation	Dependency Pair
1	I	PRON	3	nsubj	**nsubj(apologize, I)**
2	sincerely	ADV	3	advmod	advmod(apologize, sincerely)
3	apologise	VERB	3	ROOT	ROOT(--,apologise)
4	for	ADP	3	prep	prep(apologise,for)
5	my	DET	6	poss	det(behaviour, my)
6	behaviour	NOUN	4	pobj	pobj(for, behaviour)



Dependency Trees

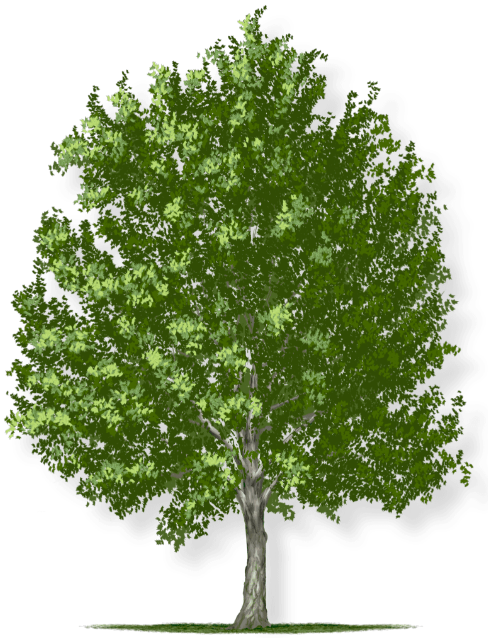
Syntactic tag	%	Gloss	Syntactic tag	%	Gloss
acl	1.89	adjectival clause	expl	0.00	expletive
advcl	0.70	adverbial clause modifier	foreign	0.01	foreign words
advmod	2.12	adverbial modifier	goeswith	0.08	goes with
amod	8.34	adjectival modifier	iobj	0.22	indirect object
appos	1.69	appositional modifier	list	0.00	list
aux	4.35	auxiliary	mark	3.59	marker
auxpass	0.71	passive auxiliary	mwe	0.32	multi-word expression
case	9.80	case marking	name	1.56	name
cc	3.09	coordinating conjunction	neg	0.30	negation modifier
ccomp	1.03	clausal complement	nmod	17.05	nominal modifier
compound	3.02	compound	nsubj	5.97	nominal subject
conj	3.80	conjunct	nsubjpass	0.65	passive nominal subject
cop	1.41	copula	nummod	2.05	numeric modifier
csubj	0.12	clausal subject	parataxis	1.47	parataxis
csubjpass	0.03	clausal passive subject	punct	12.86	punctuation
dep	0.01	unspecified dependency	remnant	0.14	remnant in ellipsis
det	0.98	determiner	root	4.51	root
discourse	0.71	discourse element	vocative	0.00	vocative
dislocated	0.01	dislocated elements	xcomp	1.50	open clausal complement
dobj	3.92	direct object			



Dependency Trees

Dependency Trees

But why though?



Dependency Trees

But why though?

The actual structure of communication

Dependency Trees

But why though?

The actual structure of communication

Word order matters!

"I don't understand you" vs. "I understand you don't"

Dependency Trees

But why though?

The actual structure of communication

Word order matters!
"I don't understand you" vs. "I understand you don't"

Distinguish homonyms (“I like you” vs. “It’s like trash”)

Dependency Trees

But why though?

The actual structure of communication

Word order matters!
"I don't understand you" vs. "I understand you don't"

Distinguish homonyms (“I like you” vs. “It’s like trash”)

Not in LIWC, dictionaries, topic models, word2vec...

Dependency Trees

But why though?

The actual structure of communication

Word order matters!
"I don't understand you" vs. "I understand you don't"

Distinguish homonyms (“I like you” vs. “It’s like trash”)

Not in LIWC, dictionaries, topic models, word2vec...

What about neural nets? (BERT, GPT-3...)

Emergent linguistic structure in artificial neural networks trained by self-supervision

Christopher D. Manning[a,1], Kevin Clark[a], John Hewitt[a], Urvashi Khandelwal[a], and Omer Levy[b]

[a]Computer Science Department, Stanford University, Stanford, CA 94305; and [b]Facebook Artificial Intelligence Research, Facebook Inc., Seattle, WA 98109

Edited by Matan Gavish, Hebrew University of Jerusalem, Jerusalem, Israel, and accepted by Editorial Board Member David L. Donoho April 13, 2020 (received for review June 3, 2019)

This paper explores the knowledge of linguistic structure learned by large artificial neural networks, trained via self-supervision, whereby the model simply tries to predict a masked word in a given context. Human language communication is via sequences of words, but language understanding requires constructing rich hierarchical structures that are never observed explicitly. The mechanisms for this have been a prime mystery of human language acquisition, while engineering work has mainly proceeded by supervised learning on treebanks of sentences hand labeled for this latent structure. However, we demonstrate that modern deep contextual language models learn major aspects of this structure, without any explicit supervision. We develop methods for identifying linguistic hierarchical structure emergent in artificial neural networks and demonstrate that components in these models focus on syntactic grammatical relationships and anaphoric coreference. Indeed, we show that a linear transformation of learned embeddings in these models captures parse tree distances to a surprising degree, allowing approximate reconstruction of the sentence tree structures normally assumed by linguists. These results help explain why these models have brought such large improvements across many language-understanding tasks.

own supervised learning problems by choosing to interpret some of the data as a "label" to be predicted.† The canonical case for human language is the language-modeling task of trying to predict the next word in an utterance based on the temporally preceding words (Fig. 2). Variant tasks include the masked language-modeling task of predicting a masked word in a text [a.k.a. the cloze task (11)] and predicting the words likely to occur around a given word (12, 13). Autoencoders (14) can also be thought of as self-supervised learning systems. Since no explicit labeling of the data is required, self-supervised learning is a type of unsupervised learning, but the approach of self-generating supervised learning objectives differentiates it from other unsupervised learning techniques such as clustering.

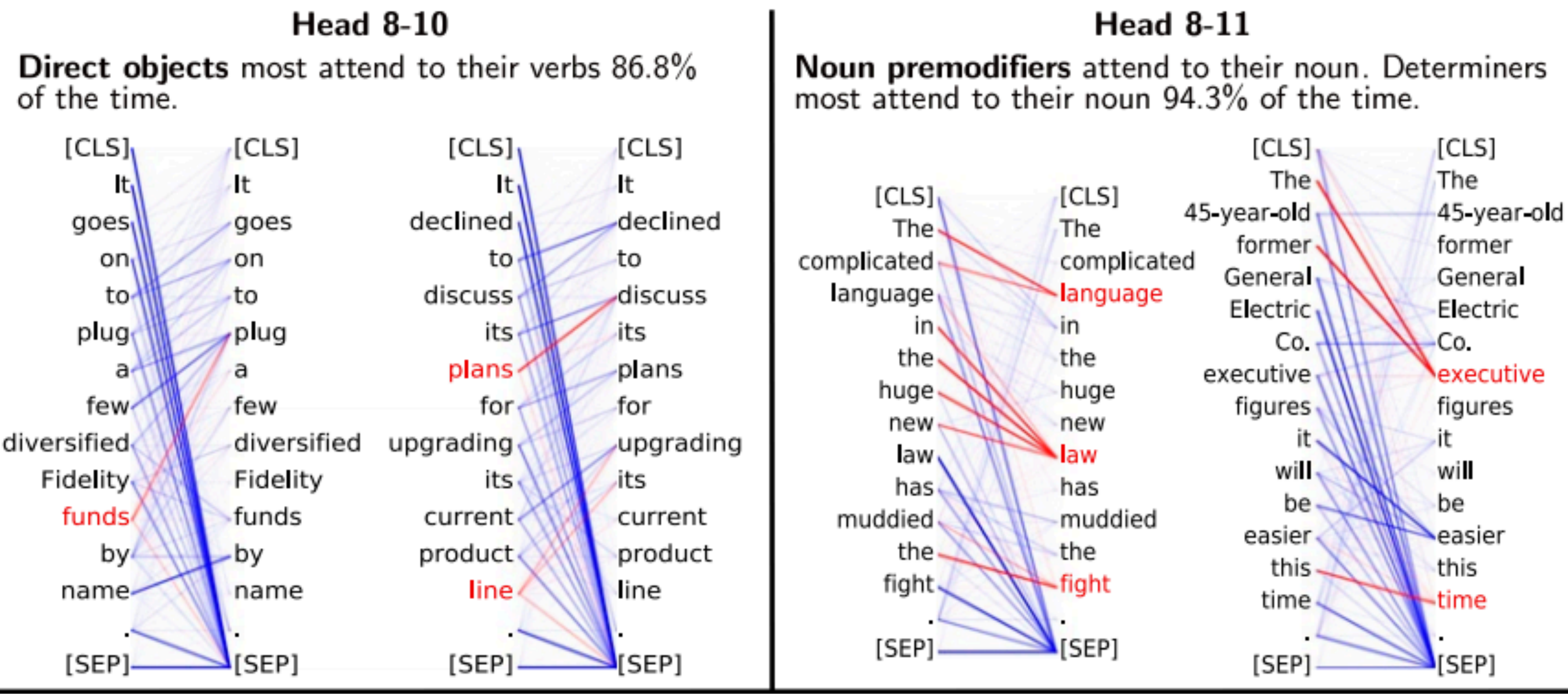
One might expect that a machine-learning model trained to predict the next word in a text will just be a giant associational learning machine, with lots of statistics on how often the word restaurant is followed by kitchen and perhaps some basic abstracted sequence knowledge such as knowing that adjectives are commonly followed by nouns in English. It is not at all clear that such a system can develop interesting knowledge of the linguistic structure of whatever human language the system is trained on. Indeed, this has been the dominant perspective in linguistics, where language models have long been seen as inadequate

Emergent linguistic structure in artificial neural networks trained by self-supervision

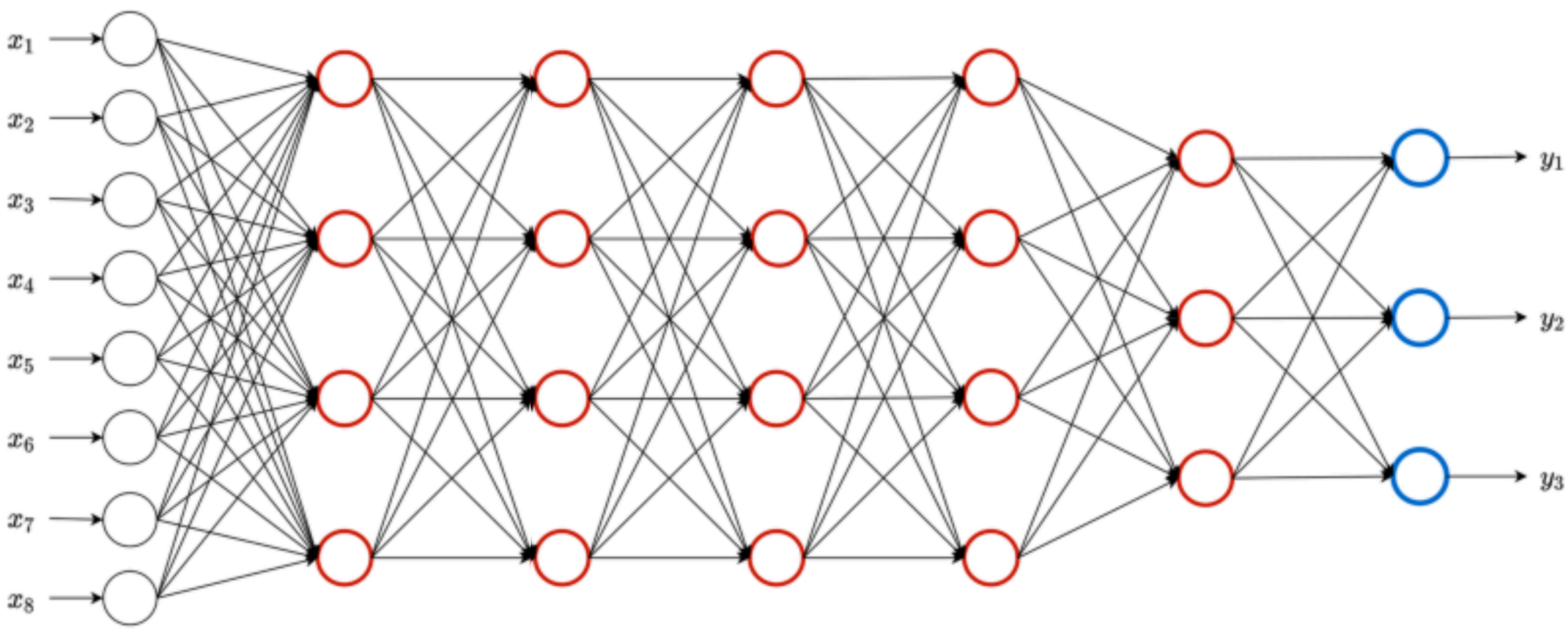
Christopher D. Manning[a,1], Kevin Clark[a], John Hewitt[a], Urvashi Khandelwal[a], and Omer Levy[b]

[a]Computer Science Department, Stanford University, Stanford, CA 94305; and [b]Facebook Artificial Intelligence Research, Facebook Inc., Seattle, WA 98109

Edited by Matan Gavish, Hebrew University of Jerusalem, Jerusalem, Israel, and accepted by Editorial Board Member David L. Donoho April 13, 2020

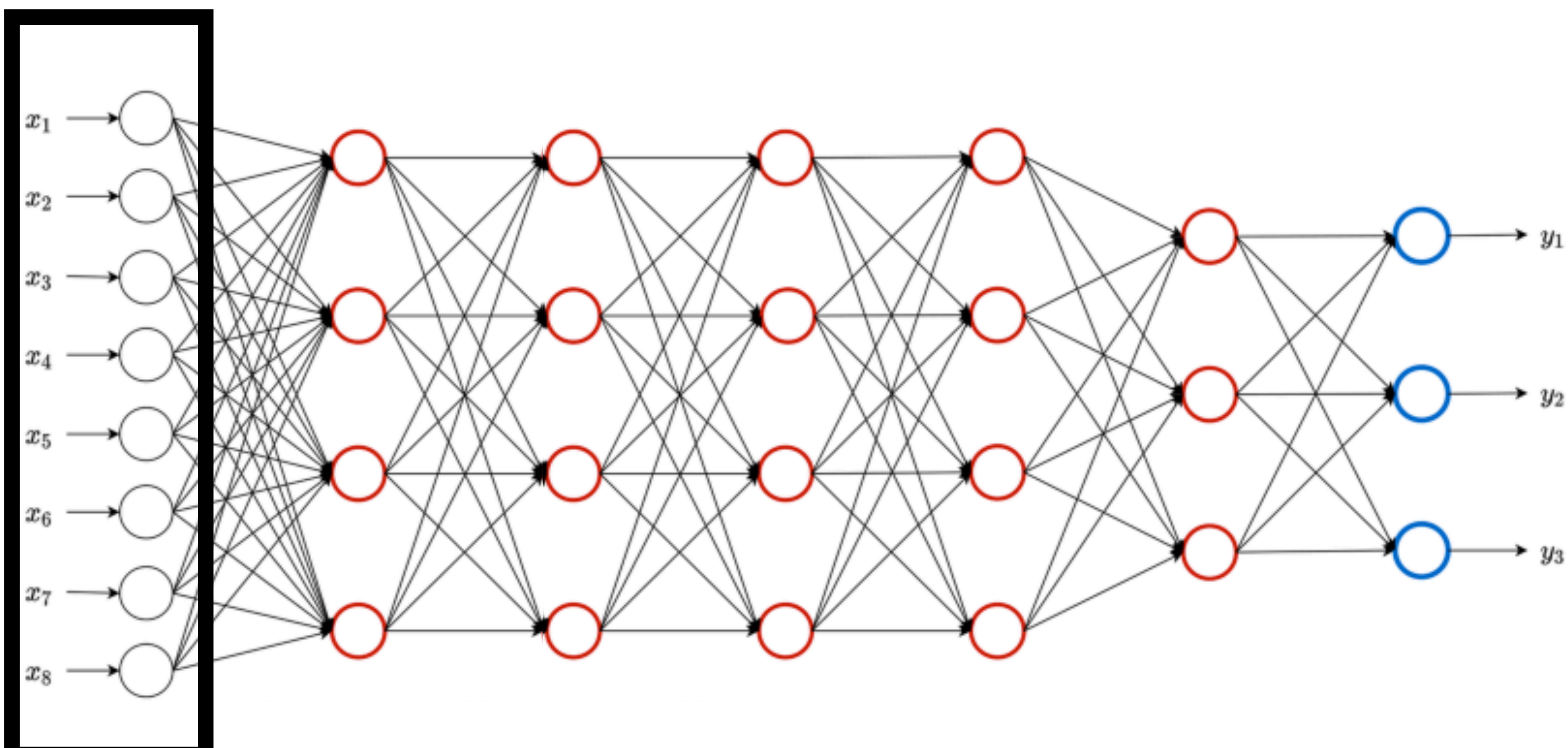


Pre-Trained Models



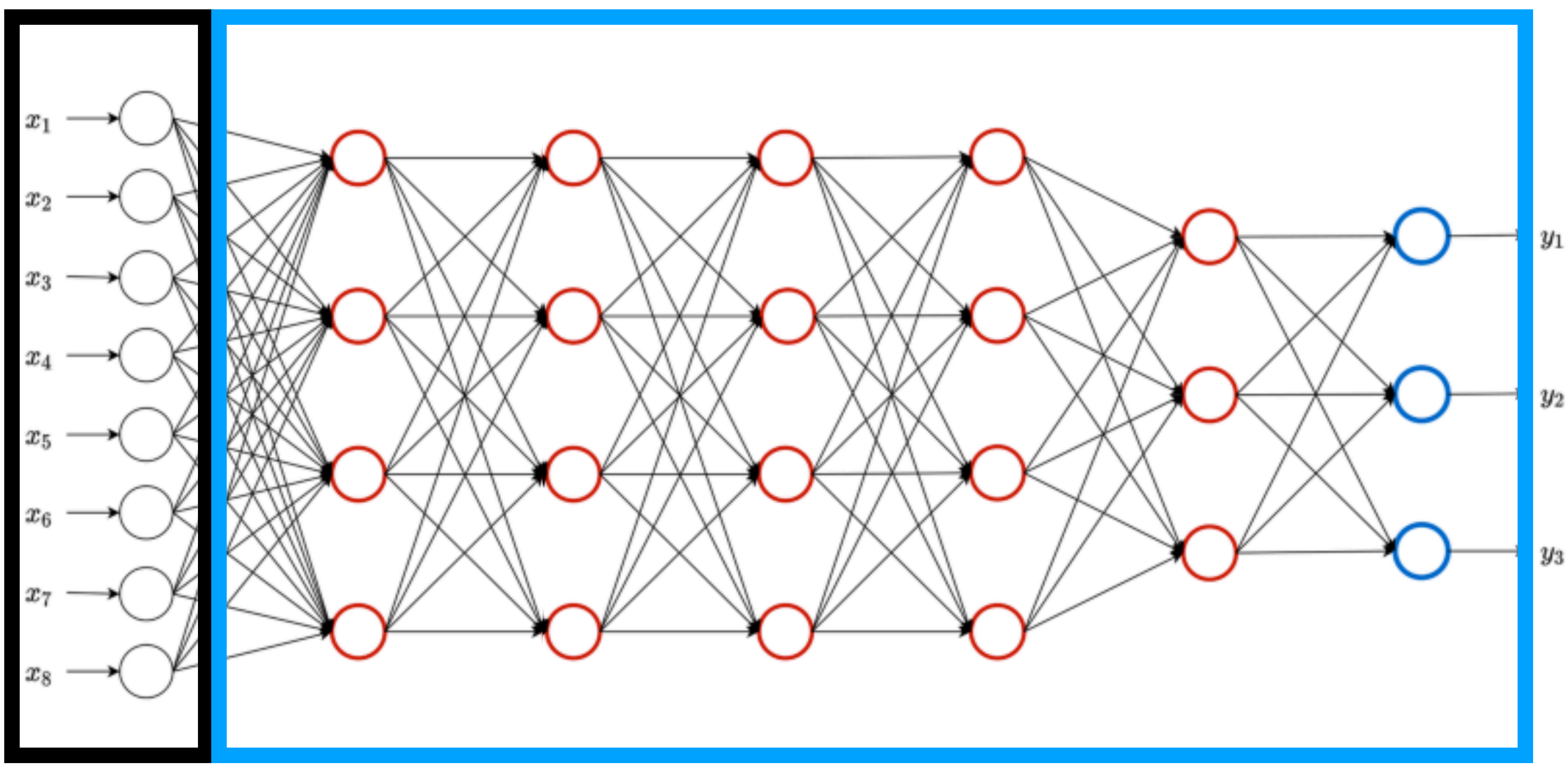
Pre-Trained Models

Words in a Document



Pre-Trained Models

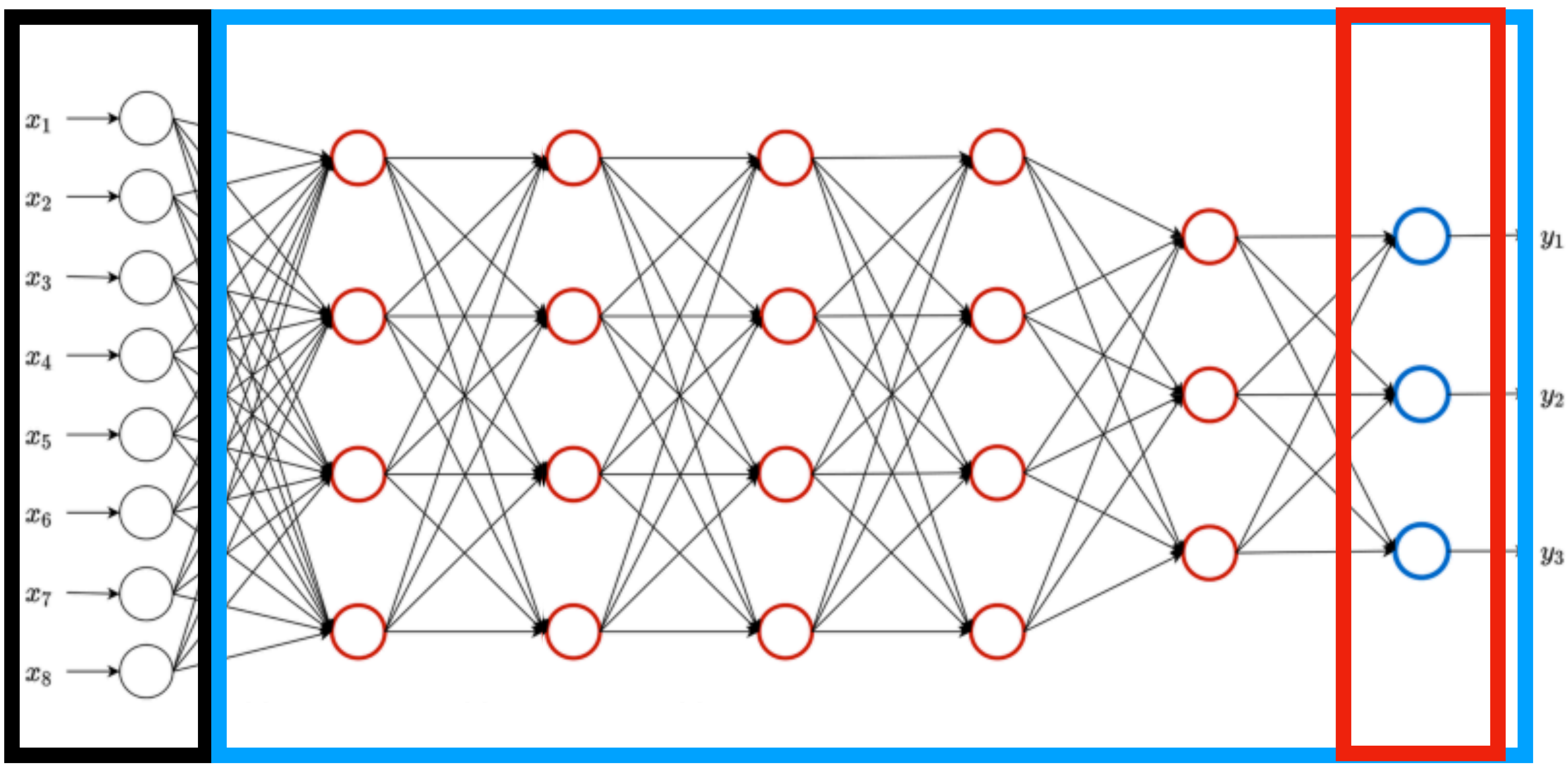
spaCy



Pre-Trained Models

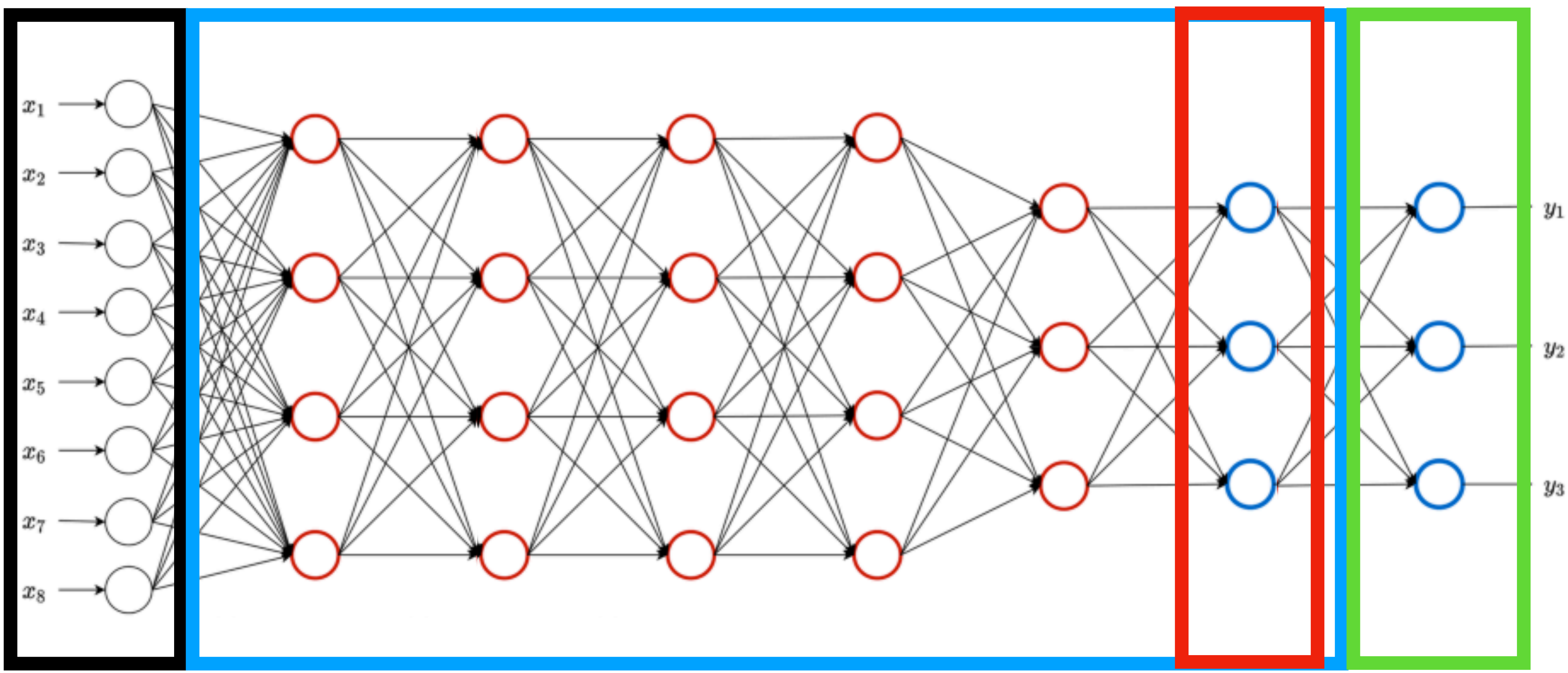
spaCy

Grammar Labels



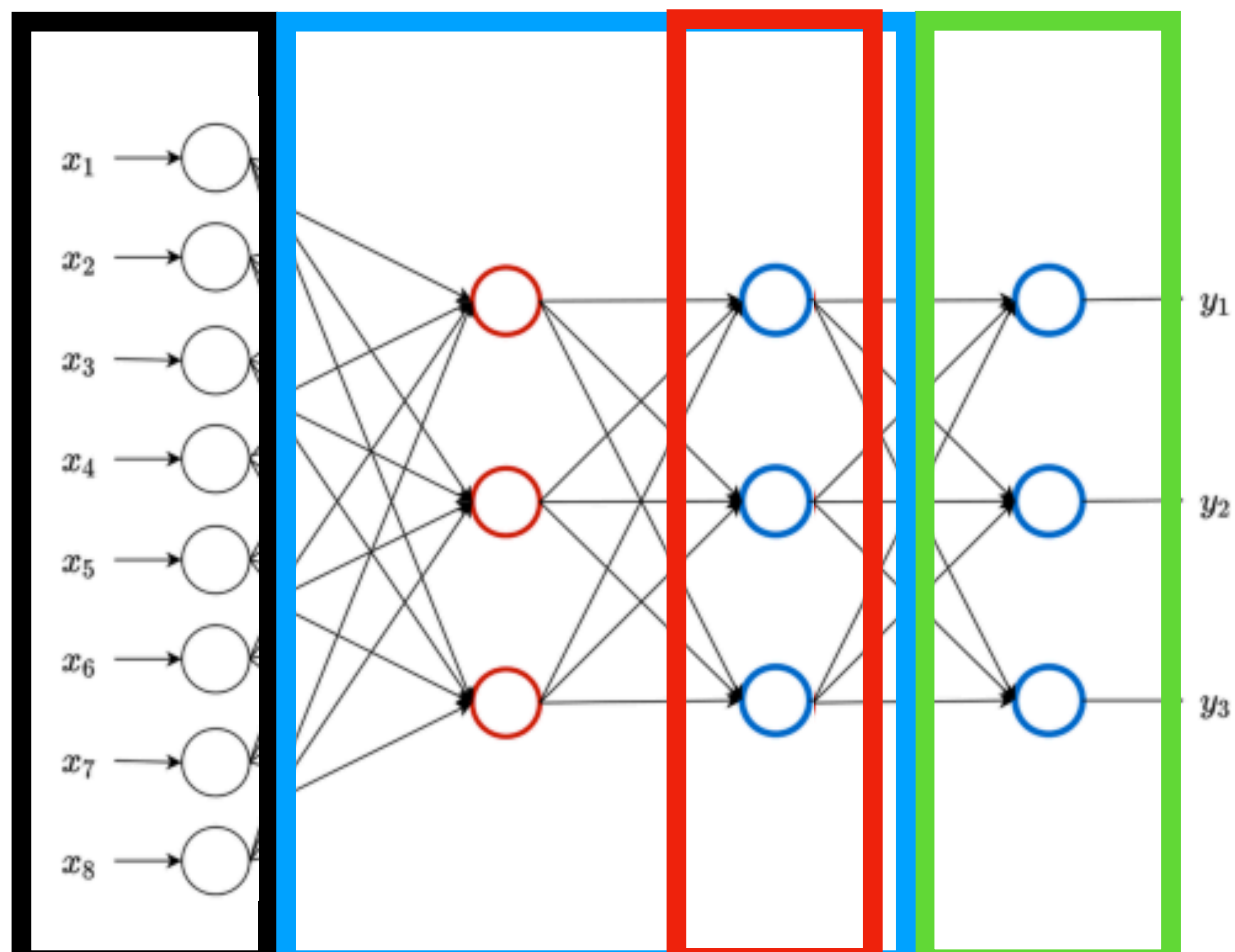
Pre-Trained Models

New quantity of interest



Pre-Trained Models

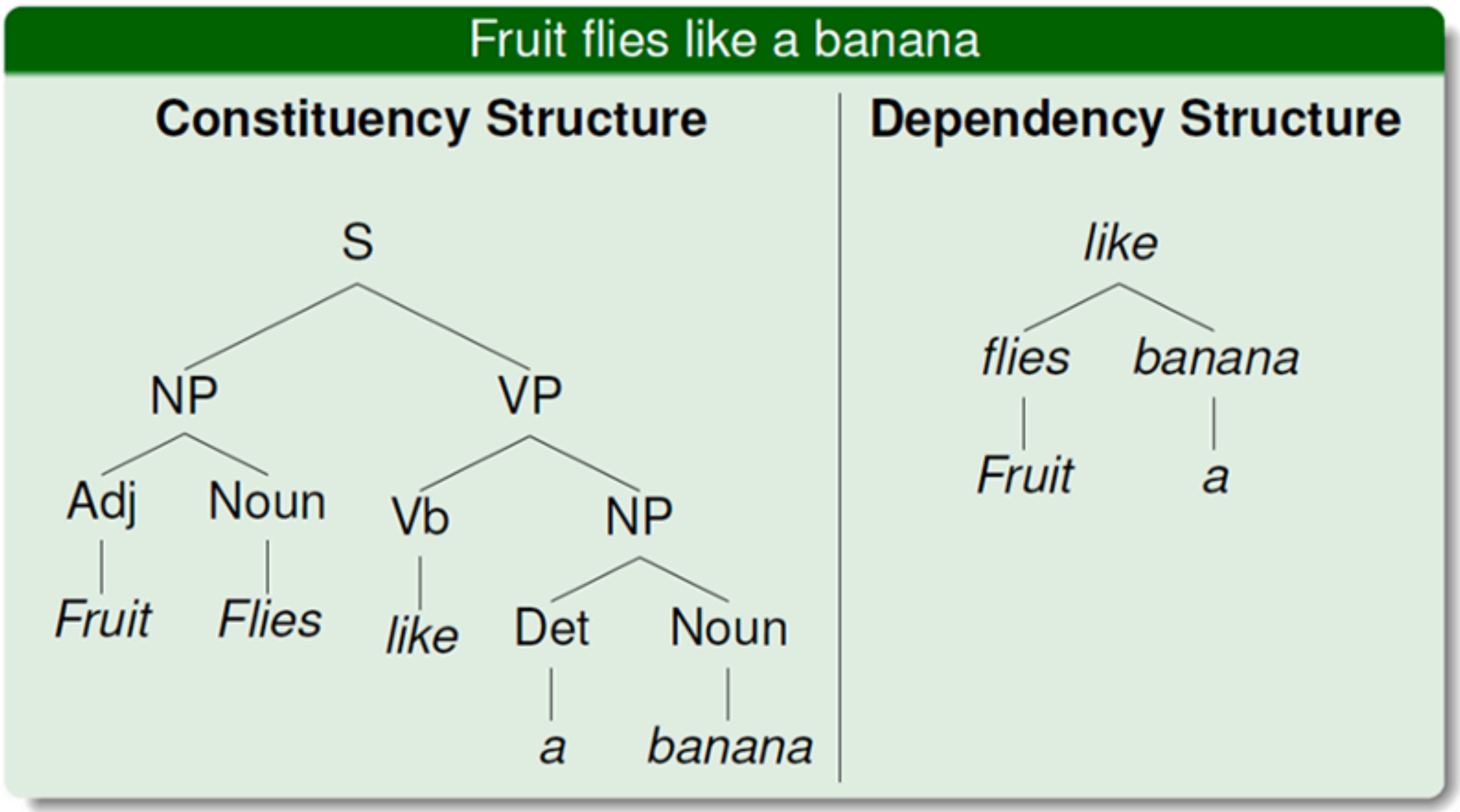
New quantity of interest



Pre-Trained Models

spaCy

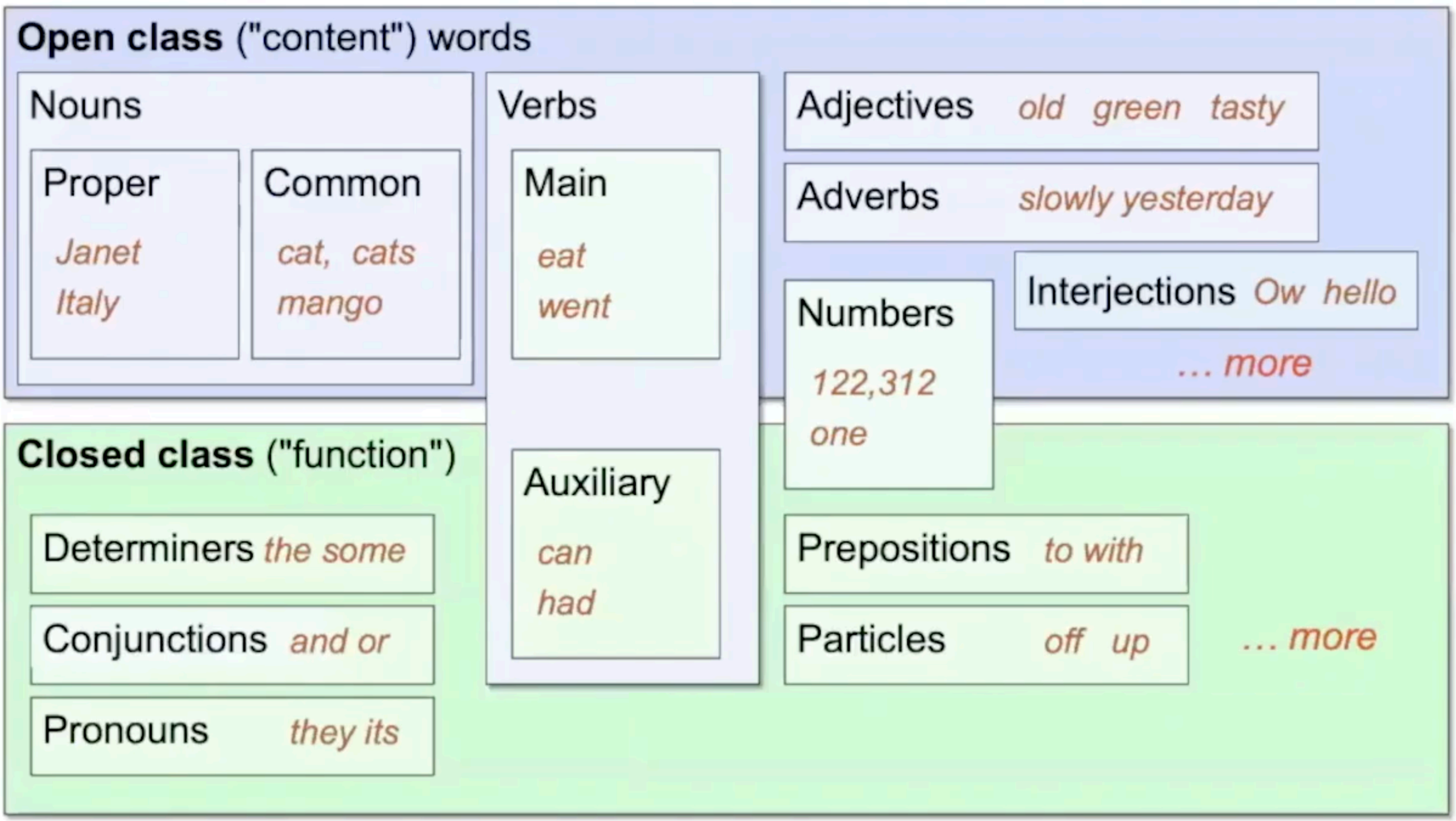
Two Types of Structure



Pre-Trained Models

Part of Speech Tagging

spaCy



Pre-Trained Models

Part of Speech Tagging

spaCy

Roughly 15% of word types are ambiguous

- Hence 85% of word types are unambiguous
- *Janet* is always PROPN, *hesitantly* is always ADV

Pre-Trained Models

Part of Speech Tagging

spaCy

Roughly 15% of word types are ambiguous

- Hence 85% of word types are unambiguous
- *Janet* is always PROPN, *hesitantly* is always ADV

But those 15% tend to be very common.

So ~60% of word tokens are ambiguous

Pre-Trained Models

Part of Speech Tagging

spaCy

Roughly 15% of word types are ambiguous

- Hence 85% of word types are unambiguous
- *Janet* is always PROPN, *hesitantly* is always ADV

But those 15% tend to be very common.

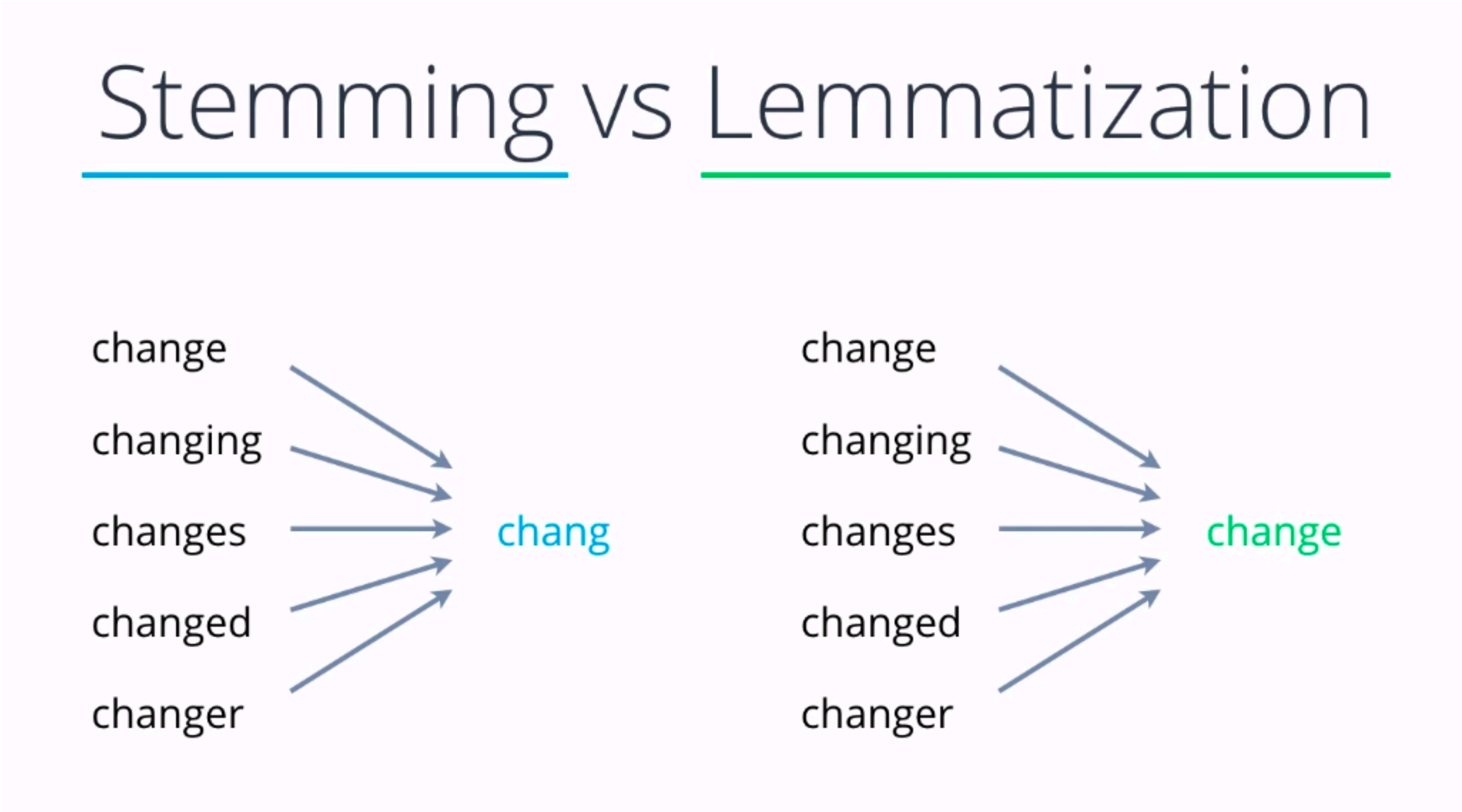
So ~60% of word tokens are ambiguous

E.g., *back*

earnings growth took a back/ADJ seat
a small building in the back/NOUN
a clear majority of senators back/VERB the bill
enable the country to buy back/PART debt
I was twenty-one back/ADV then

Pre-Trained Models

Lemmatization



Pre-Trained Models

spaCy

Lemmatization

Stemming

adjustable → adjust
formality → formaliti
formaliti → formal
airliner → airlin

Lemmatization

was → (to) be
better → good
meeting → meeting

Pre-Trained Models

Named Entity Recognition

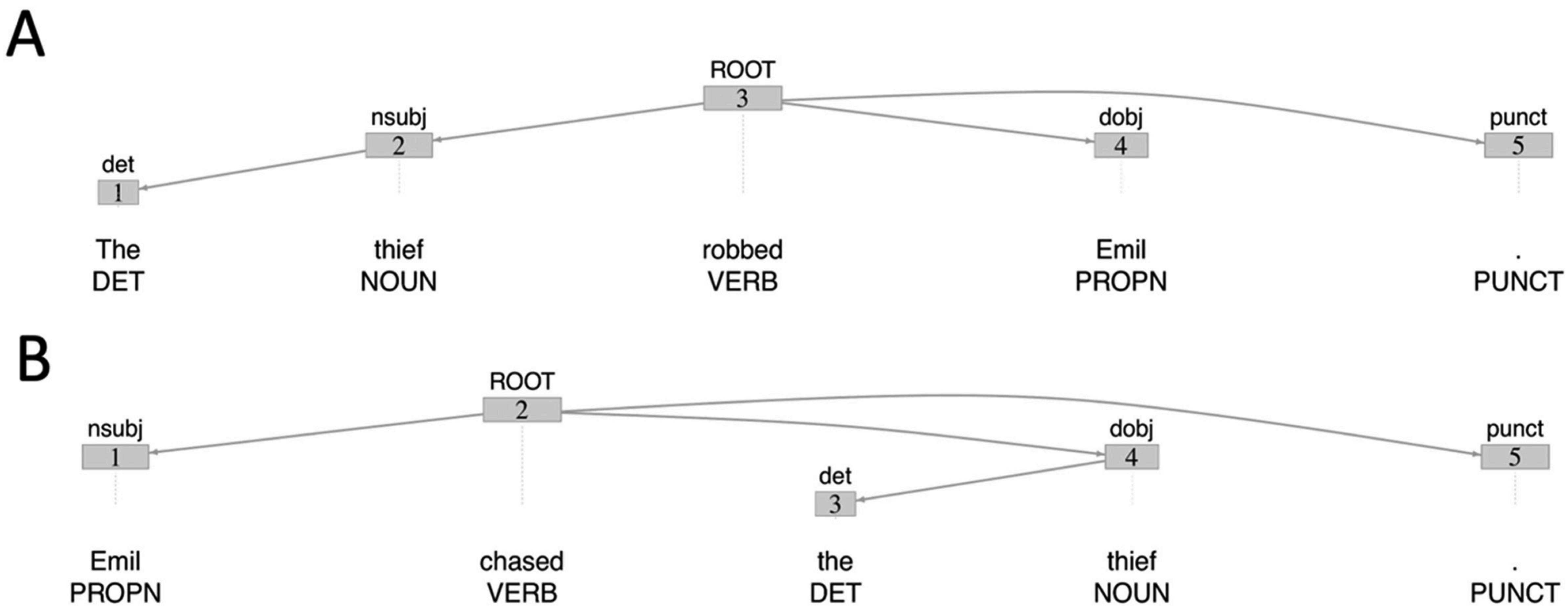
spaCy

Barack Obama Person the 44th President of the United States Title , was born in Honolulu, Hawaii Location . He graduated from Columbia University Org and Harvard Law School Org . In 2009 Date , Obama was elected as the first African American Ethnicity President of the United States Location . During his presidency, Obama implemented the Affordable Care Act Law and strengthened diplomatic relations with Cuba Location . He served two terms in office before being succeeded by President Donald Trump Title in 2017 Date .

Pre-Trained Models

Knowledge Extraction

spaCy



Stuhler, 2022

Pre-Trained Models

spaCy

Knowledge Extraction

actions of an entity

treatments of an entity

agents acting upon an entity

patients acted upon by an entity

characterizations of an entity

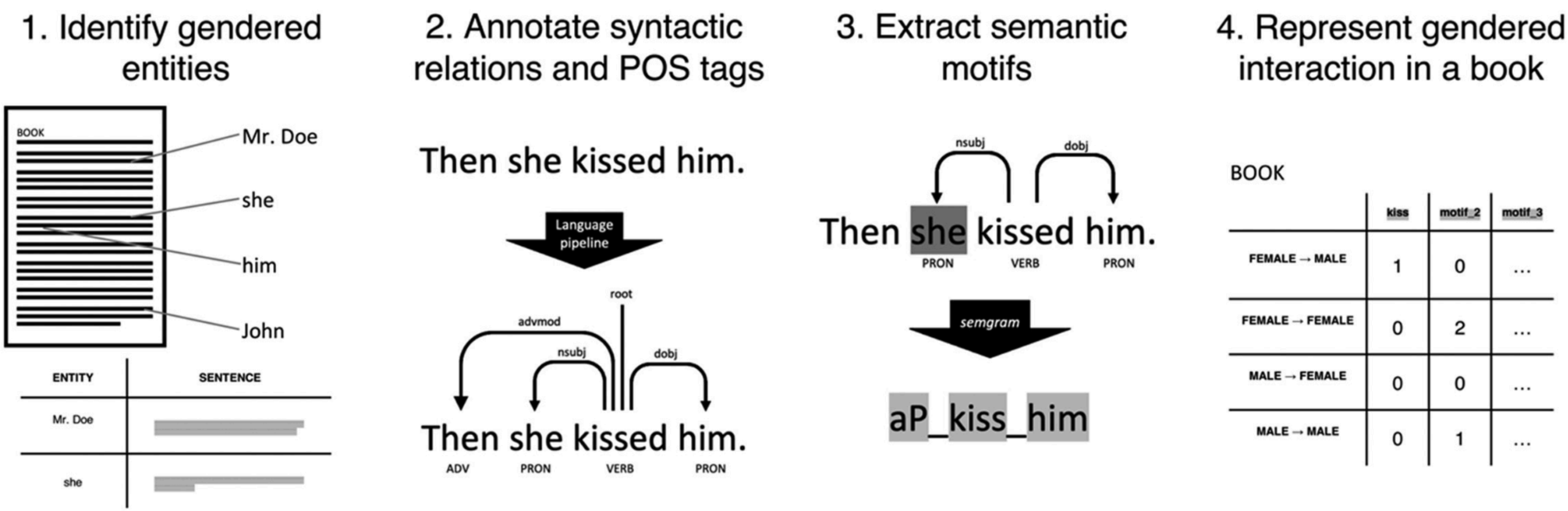
possessions of an entity

Stuhler, 2022

Pre-Trained Models

Knowledge Extraction

spaCy

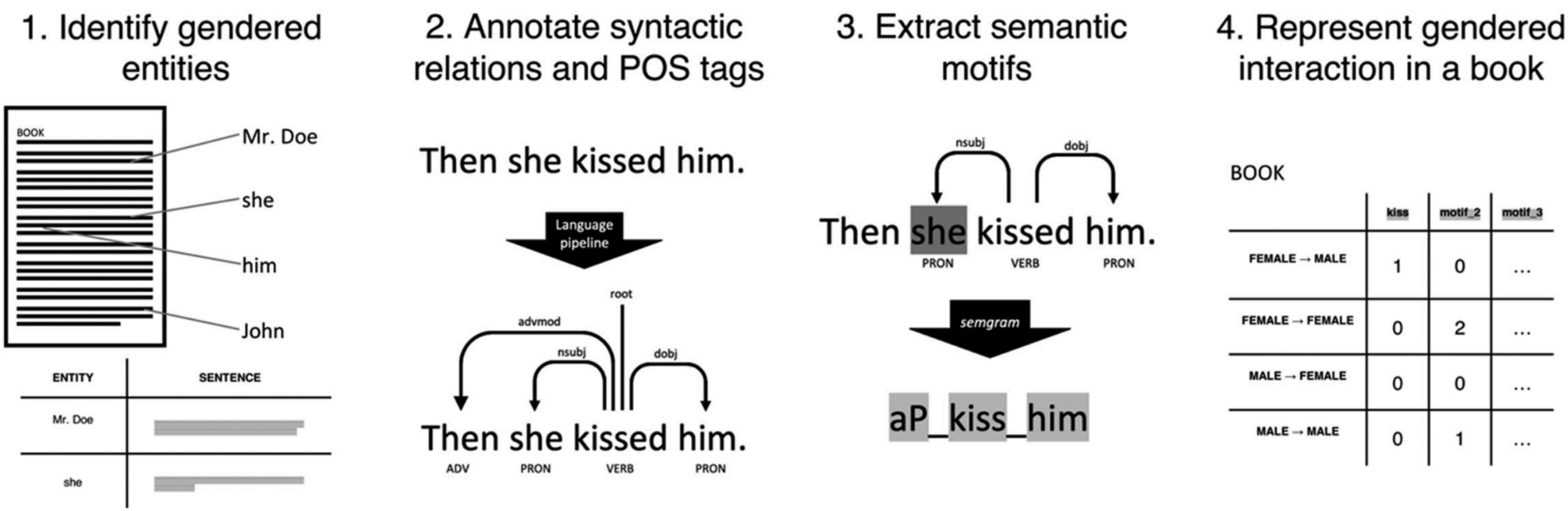


Stuhler, 2022

Pre-Trained Models

Knowledge Extraction

spaCy



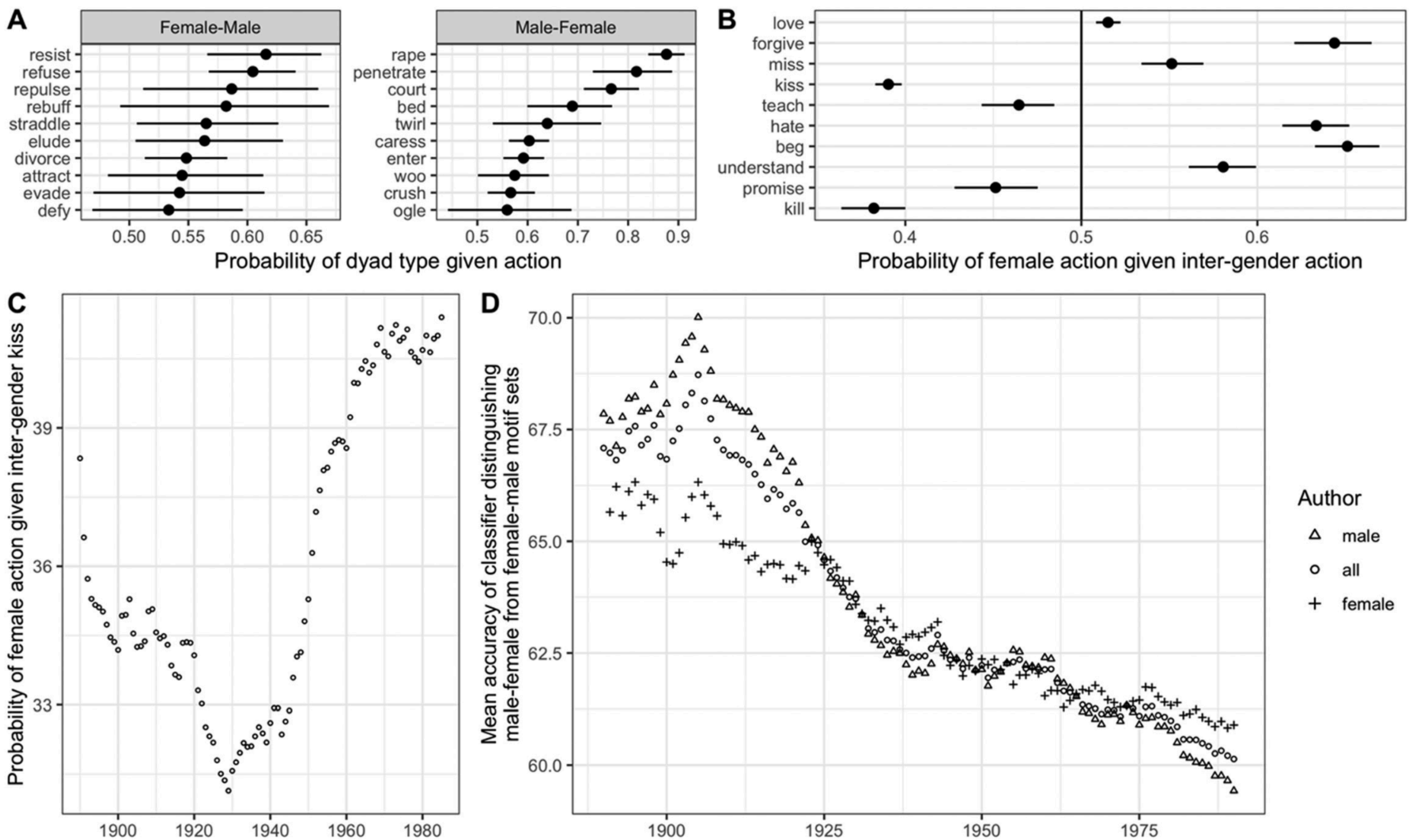
	Male patient	Female patient	Total
Male actions	17.3%	39.1%	56.4%
Female actions	32.2%	11.5%	43.7%
Total	49.5%	50.6%	

Stuhler, 2022

Pre-Trained Models

Knowledge Extraction

spaCy



Stuhler, 2022

Pre-Trained Models

Knowledge Extraction

spaCy

Most Frequent Subject-Modal-Verb Tuples

Subject - Modal - Verb	Subject - Modal - Verb	Subject - Modal - Verb
agreement_shall_be	employee_shall_be	employee_shall_receive
arbitrator_shall_have	employee_shall_be_allow	employee_shall_retain
board_shall_have	employee_shall_be_consider	employee_will_be
case_may_be	employee_shall_be_entitle	employee_will_be_allow
committee_shall_meet	employee_shall_be_give	employee_will_be_entitle
company_shall_pay	employee_shall_be_grant	employee_will_be_give
company_shall_provide	employee_shall_be_lay_off	employee_will_be_grant
company_will_pay	employee_shall_be_pay	employee_will_be_pay
company_will_provide	employee_shall_be_require	employee_will_be_require
decision_shall_be	employee_shall_continue	employee_will_have
employee_may_request	employee_shall_lose	employer_shall_grant

Ash et al., 2019

Pre-Trained Models

Knowledge Extraction

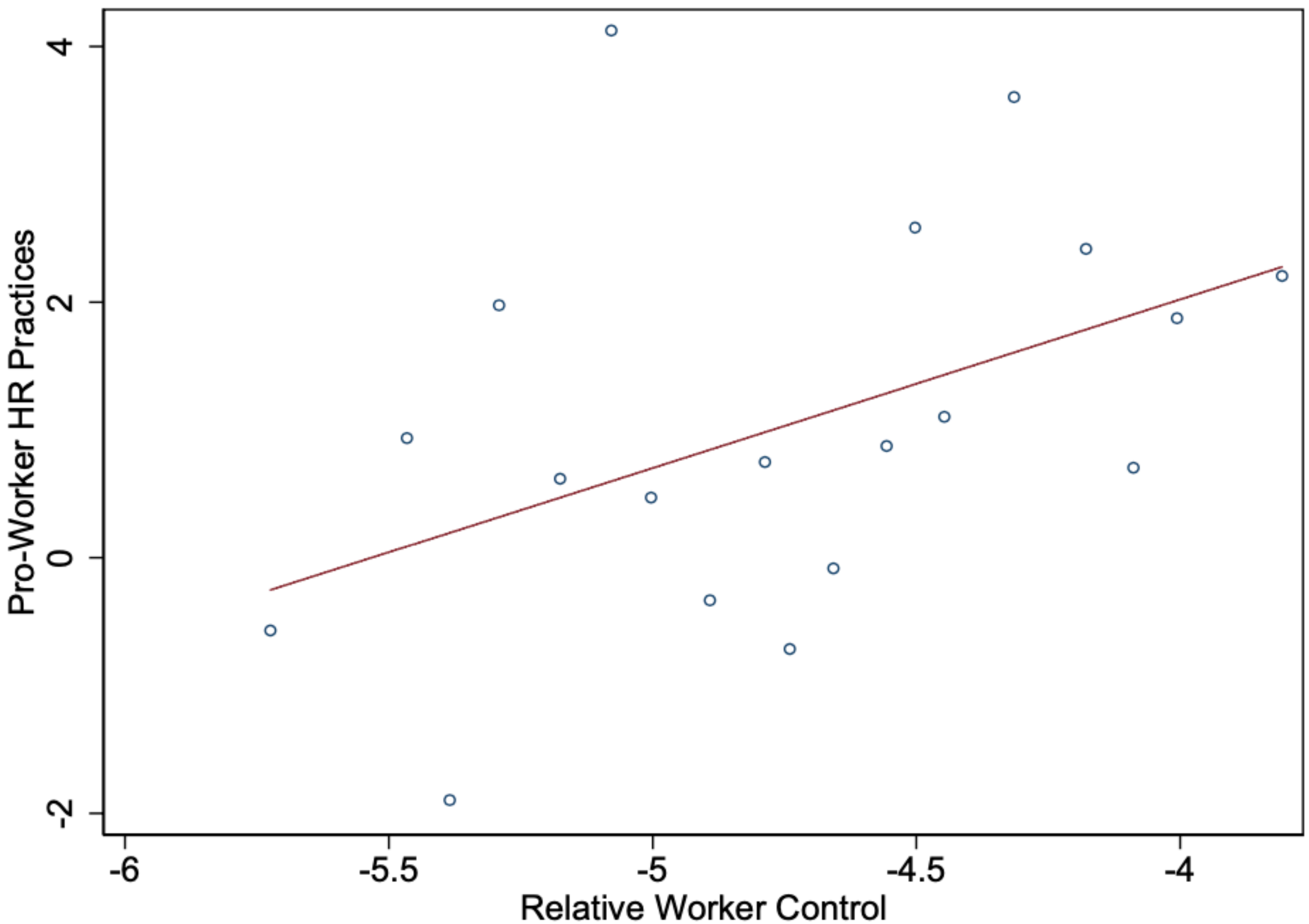
spaCy

Categorization Logic	Examples
Obligations	
Positive, Strict Modal, Active Verb	shall be, shall provide, shall include, shall notify, shall continue
Positive, Strict Modal, Obligation Verb	shall be required, shall be expected, shall be obliged
Positive, Non-Modal, Obligation Verb	is required, is expected
Prohibitions	
Negative, Any Modal, Active Verb	shall not exceed, shall not use, shall not apply, shall not discriminate
Negative, Permission Verb	shall not be allowed, is not permitted
Positive, Strict Modal, Constraint Verb	shall be prohibited, shal be restricted
Permissions	
Positive, Non-Modal, Permission Verb	is allowed, is permitted, is authorized
Positive, Strict Modal, Permission Verb	shall be allowed, shall be permitted
Positive, Permissive Modal, Active Verb	may be, may request, may use, may require, may apply
Negative, Any Modal, Constraint Verb	shall not be restricted, shall not be prohibited
Entitlements	
Strict Modal, Passive Verb	shall be paid, shall be given, shall not be discharged
Positive, Strict Modal, Entitlement Verb	shall have, shall receive, shall retain
Negative, Any Modal, Obligation Verb	may not be required

Ash et al., 2019

Pre-Trained Models

Knowledge Extraction



Ash et al., 2019

Single-Voice Documents

Product Reviews	News Articles	Books
Speeches	Legal Documents	Scientific Articles
Writing Samples	Survey Responses	Memory Recollections
Press Releases	Song Lyrics	Social Media Posts

Reviews of SVD Analysis

Psychology

Pennebaker, Mehl & Niederhoffer (2003)
Tasuszcik & Pennebaker (2010)
Jackson et al. (2021)
Boyd & Schwartz (2021)
Boyd & Dehghani (2021)

Economics

Gentzkow, Kelly & Taddy (2018)
Hansen & Ash (2022)

Marketing

Berger et al. (2020)

Political Sciences

Grimmer & Stewart (2013)
Benoit (2020)
Grimmer, Stewart & Roberts (2022)

Computer Sciences

Hirschberg & Manning (2015)

The SVD Model of Conversations

Dump all text into a single document

Sometimes separating each speaker, sometimes not

"Everything, everywhere, all at once"

The SVD Model of Conversations

Dump all text into a single document

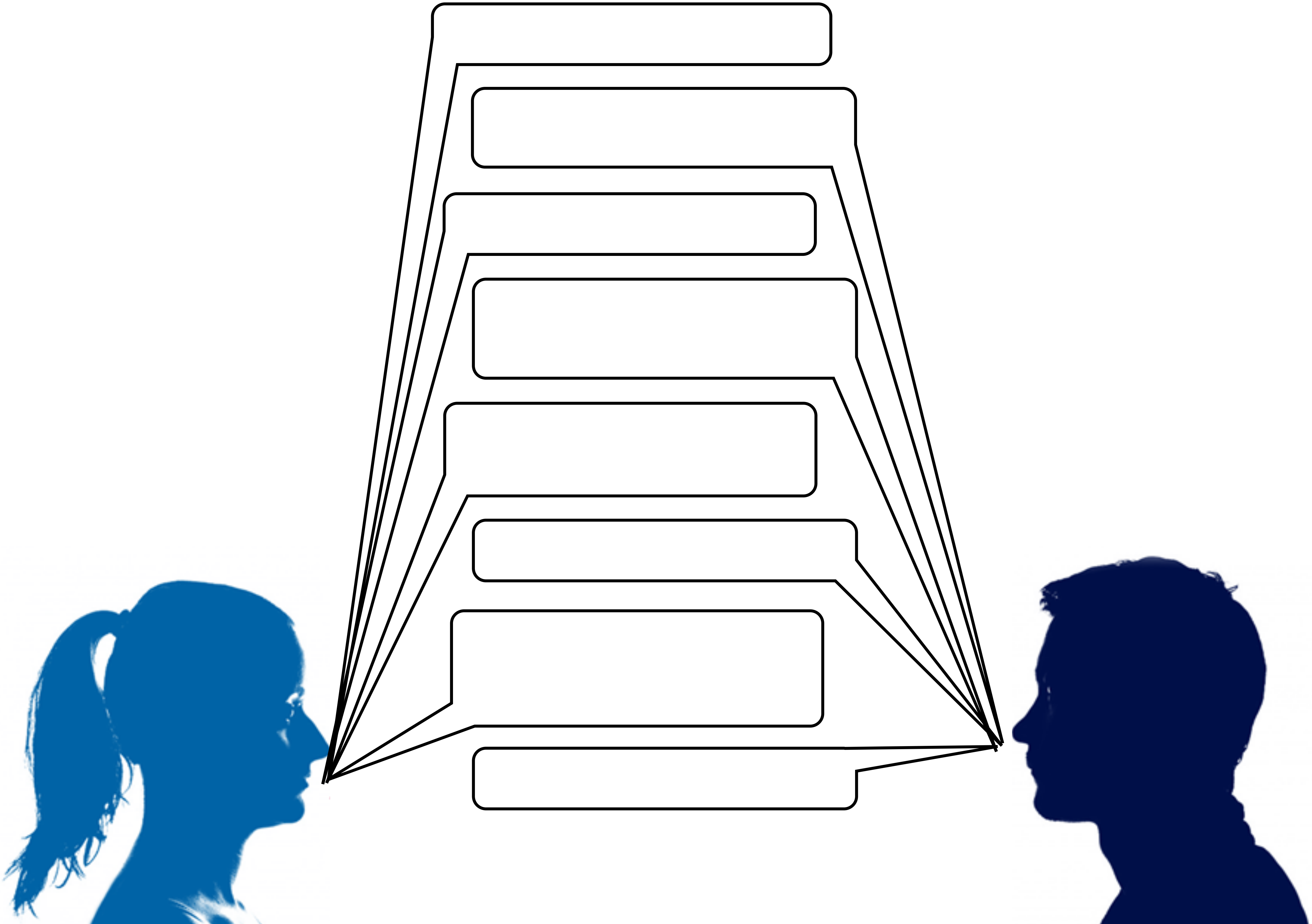
Sometimes separating each speaker, sometimes not

"Everything, everywhere, all at once"

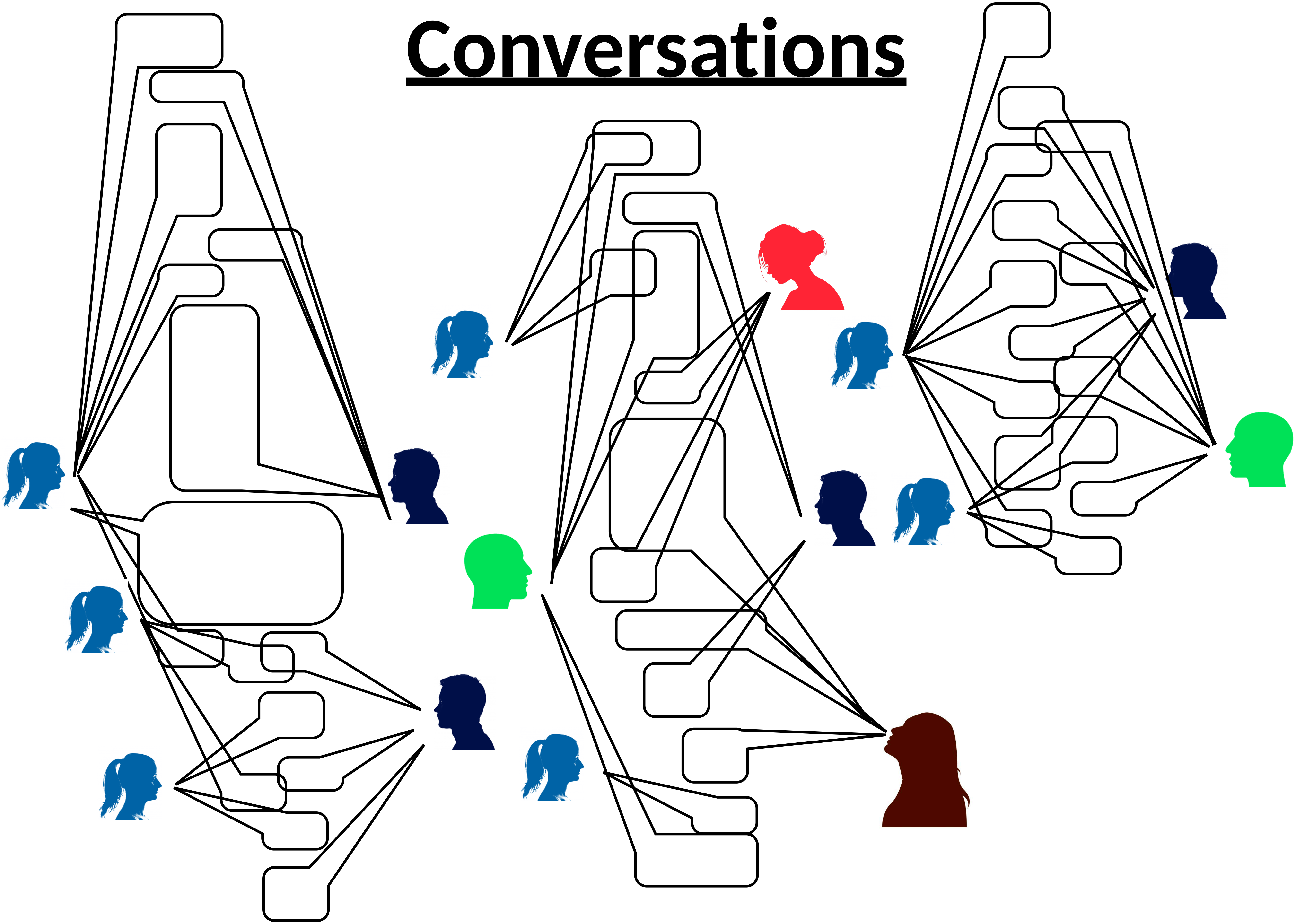
Analyse it using tools built for SVDs

Dictionaries, topic models, word counts, etc.

Conversations



Conversations



Single-Voice Documents

Product Reviews	News Articles	Books
Speeches	Legal Documents	Scientific Articles
Writing Samples	Survey Responses	Memory Recollections
Press Releases	Song Lyrics	Social Media Posts

Reviews of SVD Analysis

Psychology

Pennebaker, Mehl & Niederhoffer (2003)
Tasuszcik & Pennebaker (2010)
Jackson et al. (2021)
Boyd & Schwartz (2021)
Boyd & Dehghani (2021)

Economics

Gentzkow, Kelly & Taddy (2018)
Hansen & Ash (2022)

Marketing

Berger et al. (2020)

Political Sciences

Grimmer & Stewart (2013)
Benoit (2020)
Grimmer, Stewart & Roberts (2022)

Computer Sciences

Hirschberg & Manning (2015)

The SVD Model of Conversations

Dump all text into a single document

Sometimes separating each speaker, sometimes not

"Everything, everywhere, all at once"

The SVD Model of Conversations

Dump all text into a single document

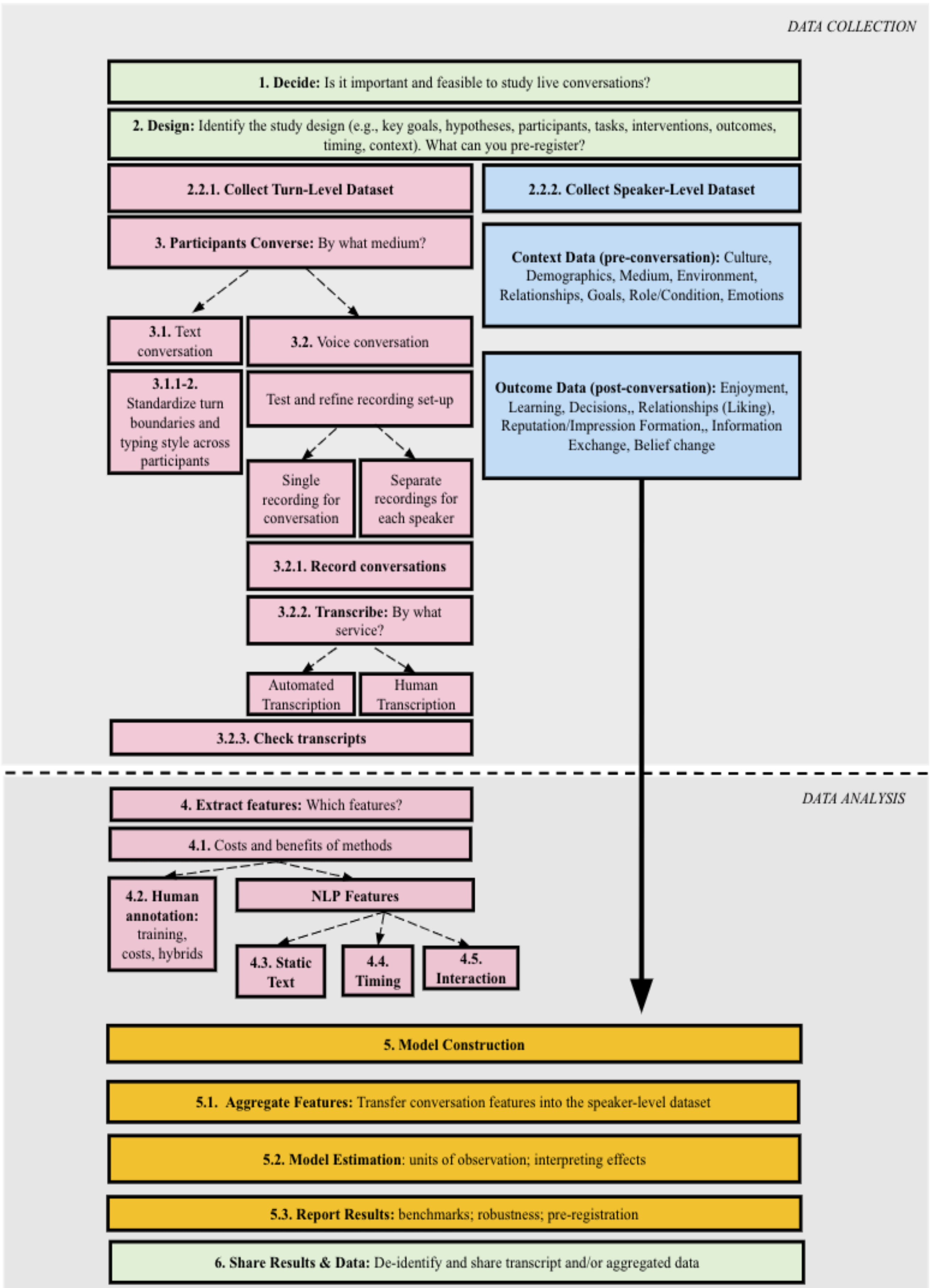
Sometimes separating each speaker, sometimes not

"Everything, everywhere, all at once"

Analyse it using tools built for SVDs

Dictionaries, topic models, word counts, etc.

The Whole Paper in one Figure



A Practical Guide To Conversation Research

(Yeomans, Boland, Collins, Abi-Esber & Brooks, 2023)

Conversation Data - Two tables

Conversation Data - Two tables

Turn-Level dataset - *words*, *speaker/convo ID*, *timestamps*, *features*

Conversation Data - Two tables

Turn-Level dataset - *words*, *speaker/convo ID*, *timestamps*, *features*

One row for every turn in conversation

Conversation Data - Two tables

Turn-Level dataset - *words*, *speaker/convo ID*, *timestamps*, *features*

One row for every turn in conversation

	A	B	C	D	E	F	G	H	I
1	conversation_id	turn	start_time	end_time	participant_id	text	question	laughter	wordcount
2	1	1	0:00:01	0:00:03	A1	Hey, how are you? My name is Bruce, but my friends call me Barry.	1	0	14
3	1	2	0:00:04	0:00:06	B1	Nice to meet you Bruce, I'm Sarah. Where are you from?	1	0	11
4	1	3	0:00:06	0:00:12	A1	Thanks for asking! I'm from a small town outside of Chicago actually, you probably haven't heard of it. What about you?	1	0	21
5	1	4	0:00:13	0:00:19	B1	Probably not [laughter]. I've never been to Chicago. I'm from upstate Portland, Oregon. Have you ever been to Portland?	1	1	19
6	1	5	0:00:20	0:00:22	A1	No, I haven't! I've been to Seattle, but that's all.	0	0	10
7	1	6	0:00:22	0:00:27	B1	Seattle is ok. In Portland, we actually call it Vancouver's shoe [laughter].	0	1	12
8	1	7	0:00:27	0:00:28	A1	[laughter] That's funny.	0	1	3
9	1	8	0:00:29	0:00:36	B1	Umm. What's your favourite food?	1	0	5
10	1	9	0:00:37	0:00:45	A1	Hmm. That's a hard question, I really like all different foods. I made this really good stew the other day that I think might be the best thing I've eaten lately. But I'm always partial to a good hamburger.	0	0	39
11	1	10	0:00:46	0:00:55	B1	Cool. What was in your stew?	1	0	6
12	2	1	0:00:01	0:00:05	A2	Hi, nice to meet you. My name is Darla and I'm from Boston. Though I was born in Texas.	0	0	19
13	2	2	0:00:06	0:00:10	B2	Hey, I'm Tony. I'm also from Boston, but I've never been to Texas. When did you leave Texas?	1		18
14	2	3	0:00:06	0:00:12	A2	When I was really young, I think maybe 3 or 4. So I don't really remember it. But I've been back to visit once or twice.	0	0	26
15	2	4	0:00:13	0:00:19	B2	Cool. I'd love to visit someday	0	0	6

Conversation Data - Two tables

Person-Level dataset - *speaker/convo ID,*
feature summaries from self and partner,
metadata on goals/demographics/outcomes/condition etc

Conversation Data - Two tables

Person-Level dataset - *speaker/convo ID,*
feature summaries from self and partner,
metadata on goals/demographics/outcomes/condition etc

One row for each speaker in each conversation
(multiple rows per conversation, and perhaps speaker as well)

Conversation Data - Two tables

Person-Level dataset - *speaker/convo ID,*
feature summaries from self and partner,
metadata on goals/demographics/outcomes/condition etc

One row for each speaker in each conversation
(multiple rows per conversation, and perhaps speaker as well)

person_id	convo_id	partner_id	turns	condition	laughter	questions	liking	partner_likin	gender	partner_gender
100001	100001	100002	16	control	2	4	5	1	male	female
100002	100001	100001	17	control	0	1	1	5	female	male
100001	100002	100003	43	treated	1	7	6	3	male	male
100003	100002	100001	42	treated	4	2	3	6	male	male
100004	100003	100005	35	control	3	3	6	5	female	female
100005	100003	100004	35	control	3	0	5	6	female	female
100006	100004	100002	22	treated	6	1	5	6	male	female
100002	100004	100006	22	treated	0	5	6	5	female	male

Conversation Data - Two tables

Person-Level dataset - *speaker/convo ID,*
feature summaries from self and partner,
metadata on goals/demographics/outcomes/condition etc

One row for each speaker in each conversation
(multiple rows per conversation, and perhaps speaker as well)

person_id	convo_id	turn	laughter
100001	100001	1	0
100002	100001	2	0
100001	100001	3	1
100002	100001	4	0
100001	100001	5	1
100002	100001	6	0
100001	100001	7	0
100002	100001	8	0

Conversation Data - Two tables

Person-Level dataset - *speaker/convo ID,*
feature summaries from self and partner,
metadata on goals/demographics/outcomes/condition etc

One row for each speaker in each conversation
(multiple rows per conversation, and perhaps speaker as well)

person_id	convo_id	turn	laughter
100001	100001	1	0
100002	100001	2	0
100001	100001	3	1
100002	100001	4	0
100001	100001	5	1
100002	100001	6	0
100001	100001	7	0
100002	100001	8	0

```
group_by(person_id) %>%
summarize(laughter=sum(laughter))
```

Conversation Data - Two tables

Person-Level dataset - *speaker/convo ID,*
feature summaries from self and partner,
metadata on goals/demographics/outcomes/condition etc

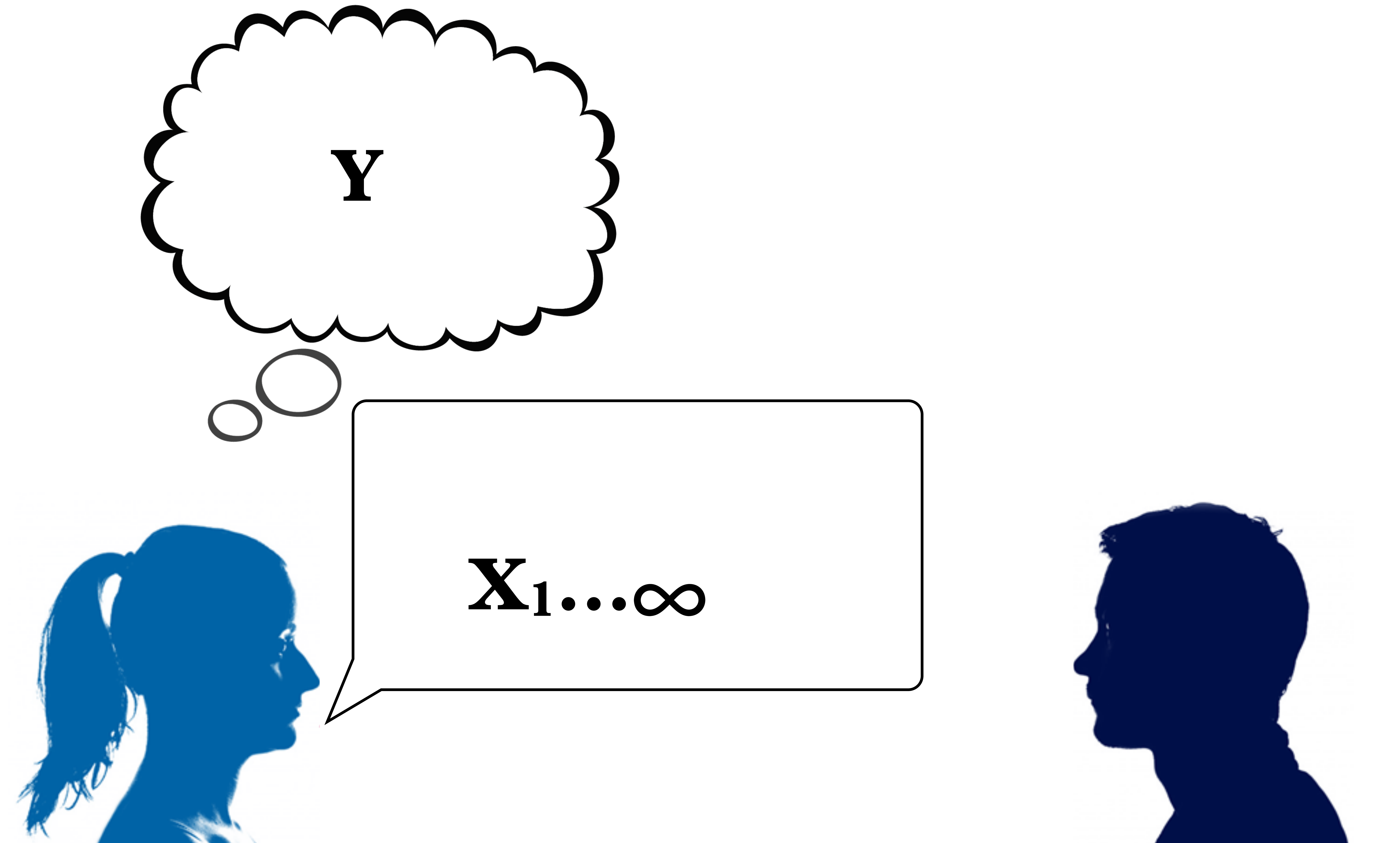
One row for each speaker in each conversation
(multiple rows per conversation, and perhaps speaker as well)

person_id	convo_id	turn	laughter
100001	100001	1	0
100002	100001	2	0
100001	100001	3	1
100002	100001	4	0
100001	100001	5	1
100002	100001	6	0
100001	100001	7	0
100002	100001	8	0

group_by(person_id) %>%
summarize(laughter=sum(laughter))

person_id	partner_id	convo_id	laughter
100001	100001	100001	2
100002	100002	100001	0
100001	100001	100002	1
100003	100003	100002	1
100004	100004	100003	3
100005	100005	100003	1
100006	100006	100004	0
100002	100002	100004	4

Effect of Own Features



Conversation Data - Two tables

Person-Level dataset - *speaker/convo ID,*
feature summaries from self and partner,
metadata on goals/demographics/outcomes/condition etc

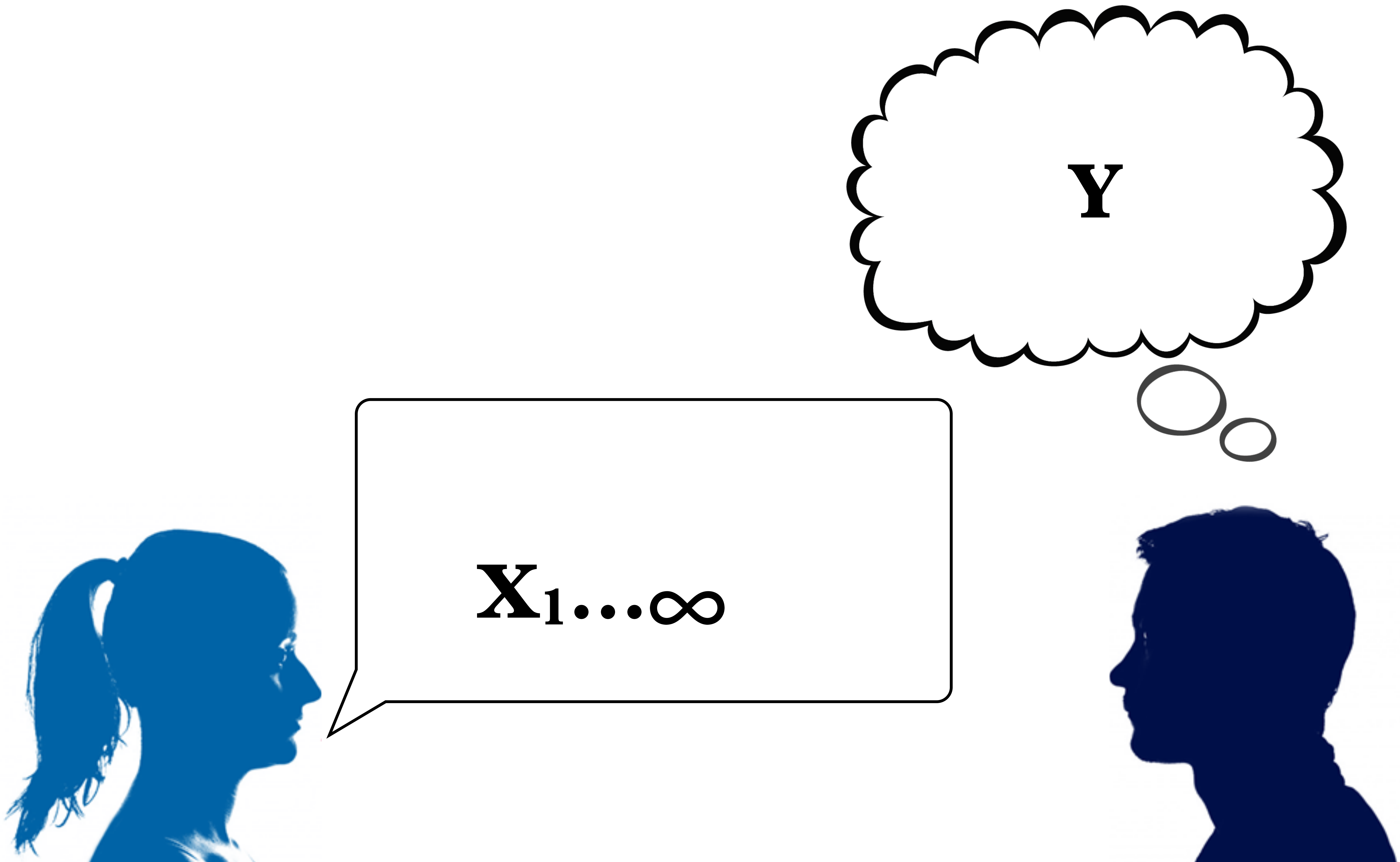
One row for each speaker in each conversation
(multiple rows per conversation, and perhaps speaker as well)

person_id	convo_id	turn	laughter
100001	100001	1	0
100002	100001	2	0
100001	100001	3	1
100002	100001	4	0
100001	100001	5	1
100002	100001	6	0
100001	100001	7	0
100002	100001	8	0

rename(partner_id="person_id") %>%
group_by(partner_id) %>%
summarize(laughter=sum(laughter))

person_id	partner_id	convo_id	laughter	partner_laughter
100001	100001	100001	2	0
100002	100002	100001	0	2
100001	100001	100002	1	1
100003	100003	100002	1	1
100004	100004	100003	3	1
100005	100005	100003	1	3
100006	100006	100004	0	4
100002	100002	100004	4	0

Effect of Partner Features



Conversations versus SVDs

1. Multiple speakers: each with different goals, personalities, styles, etc

Why do we converse?

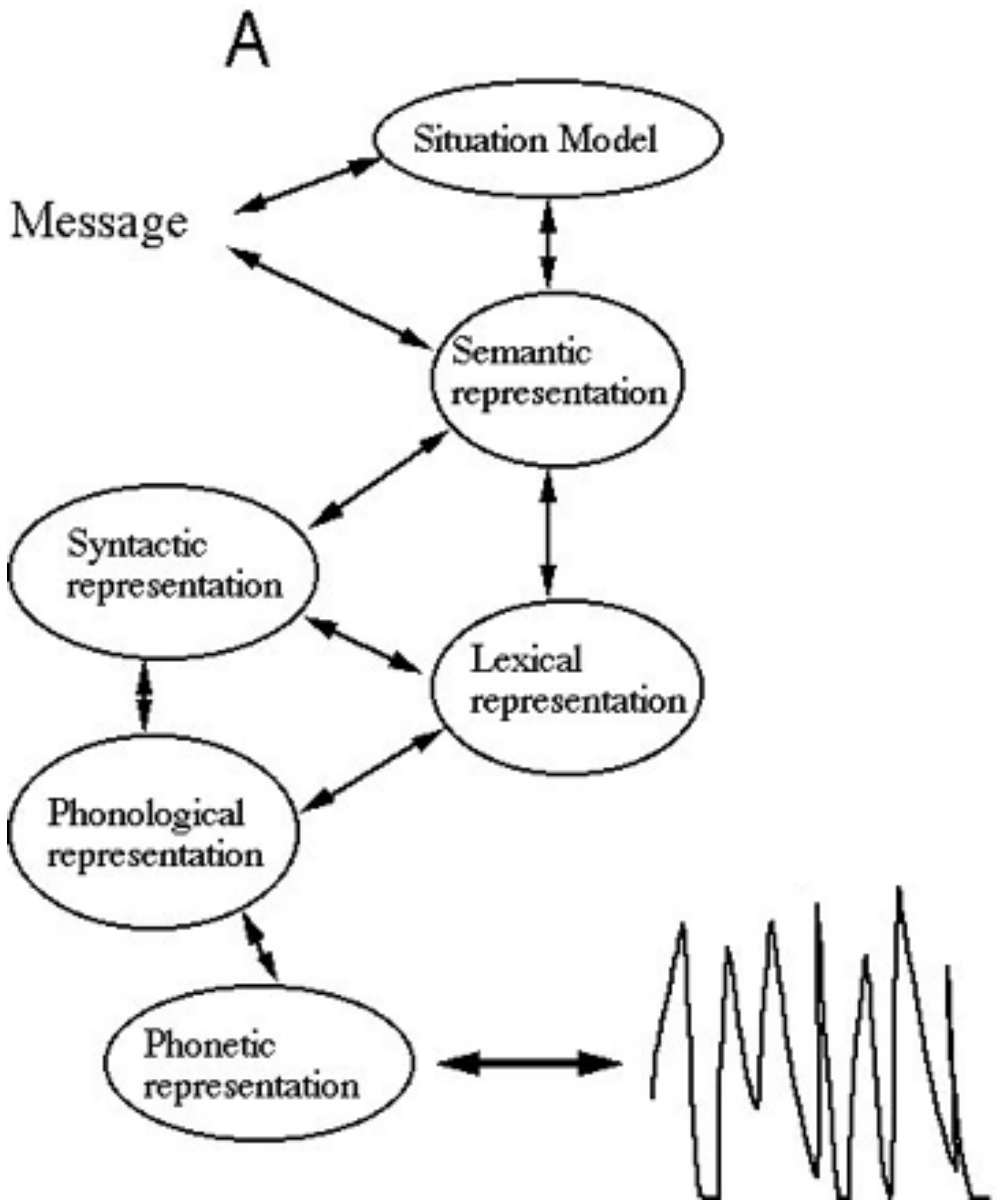
Why do we converse?

Traditionally: generate and maintain common ground

Toward A Mechanistic Psychology of Dialogue
(Pickering & Garrod, 2004)

Why do we converse?

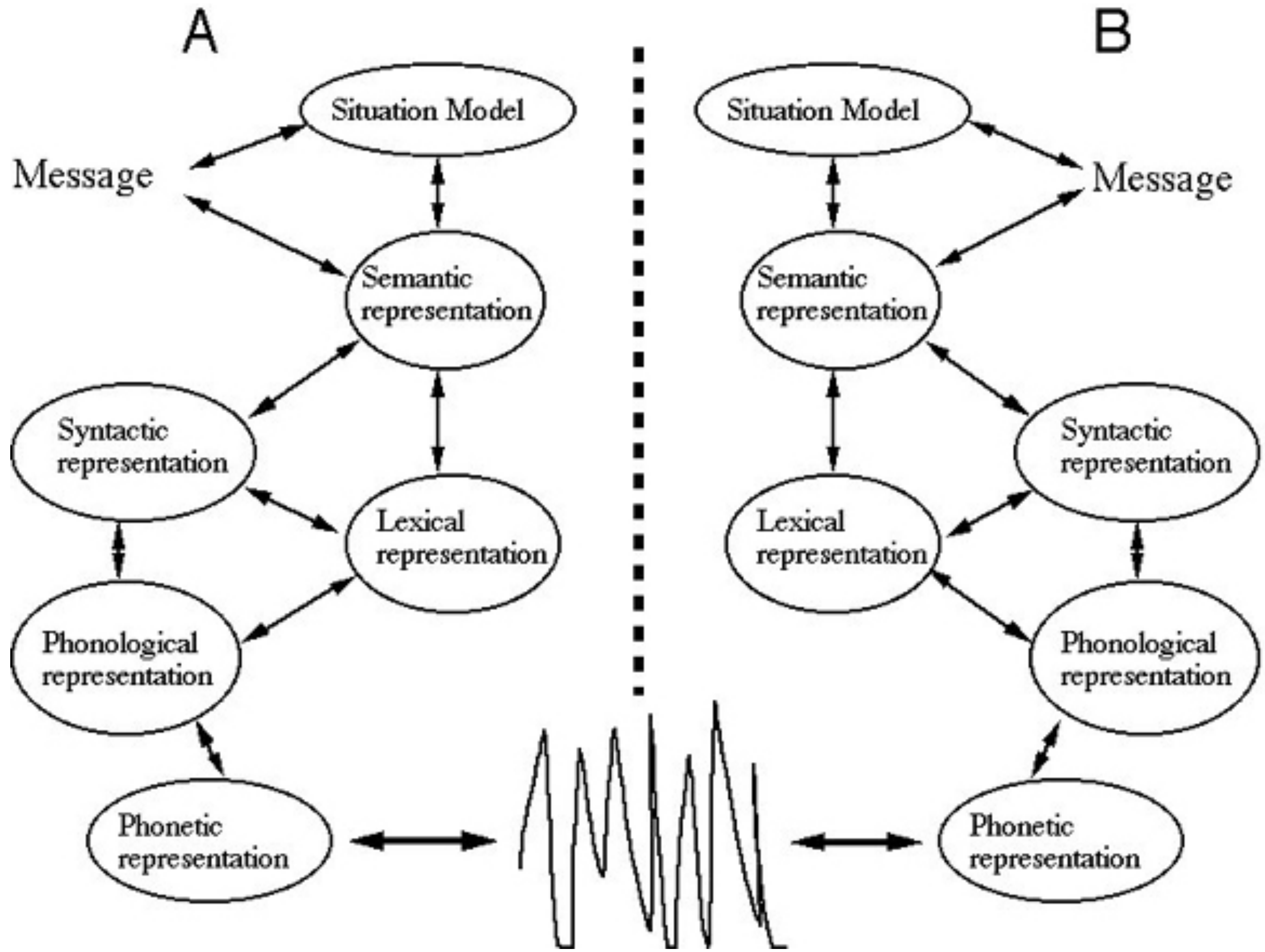
Traditionally: generate and maintain common ground



Toward A Mechanistic Psychology of Dialogue
(Pickering & Garrod, 2004)

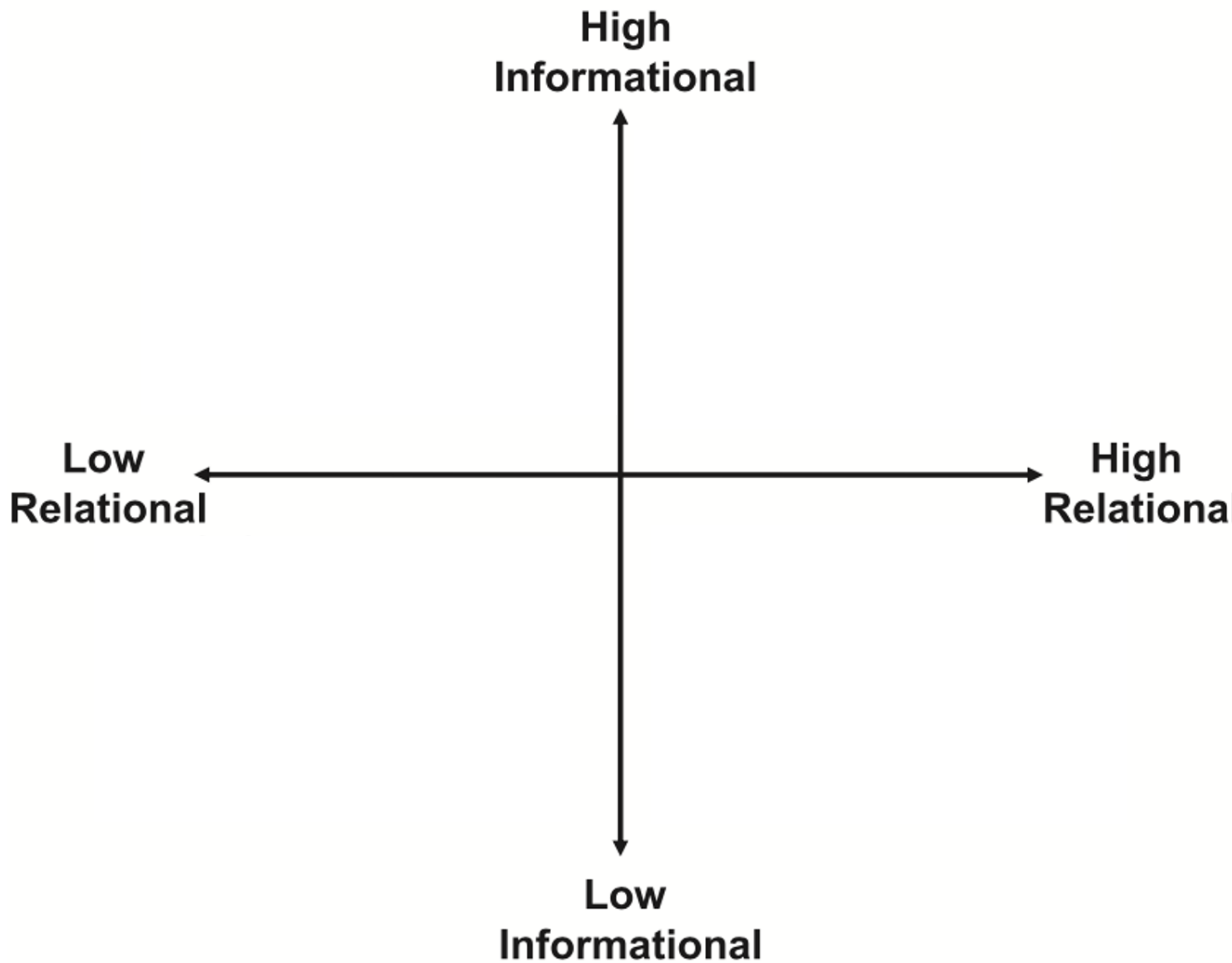
Why do we converse?

Traditionally: generate and maintain common ground



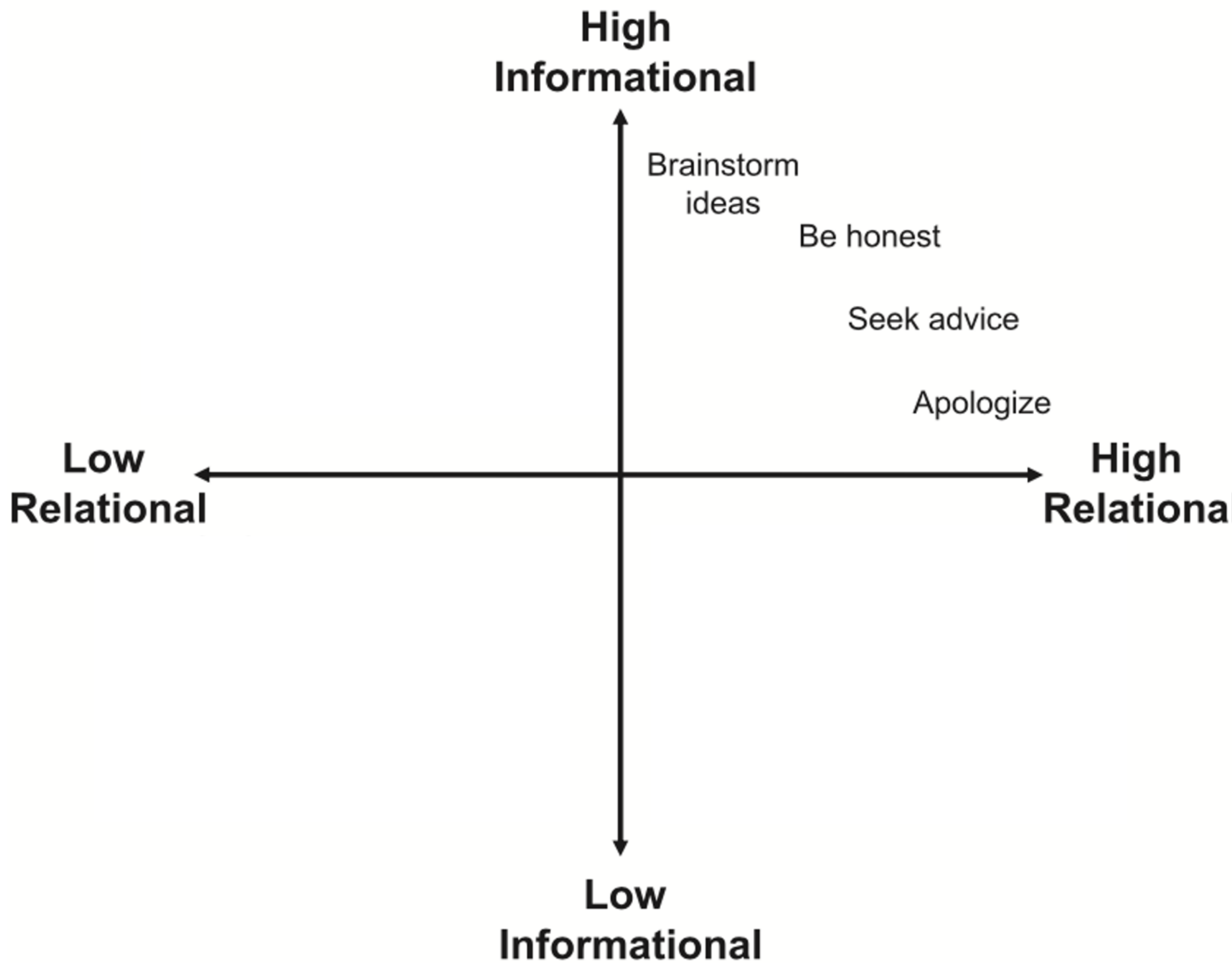
Toward A Mechanistic Psychology of Dialogue
(Pickering & Garrod, 2004)

Goals of Conversation



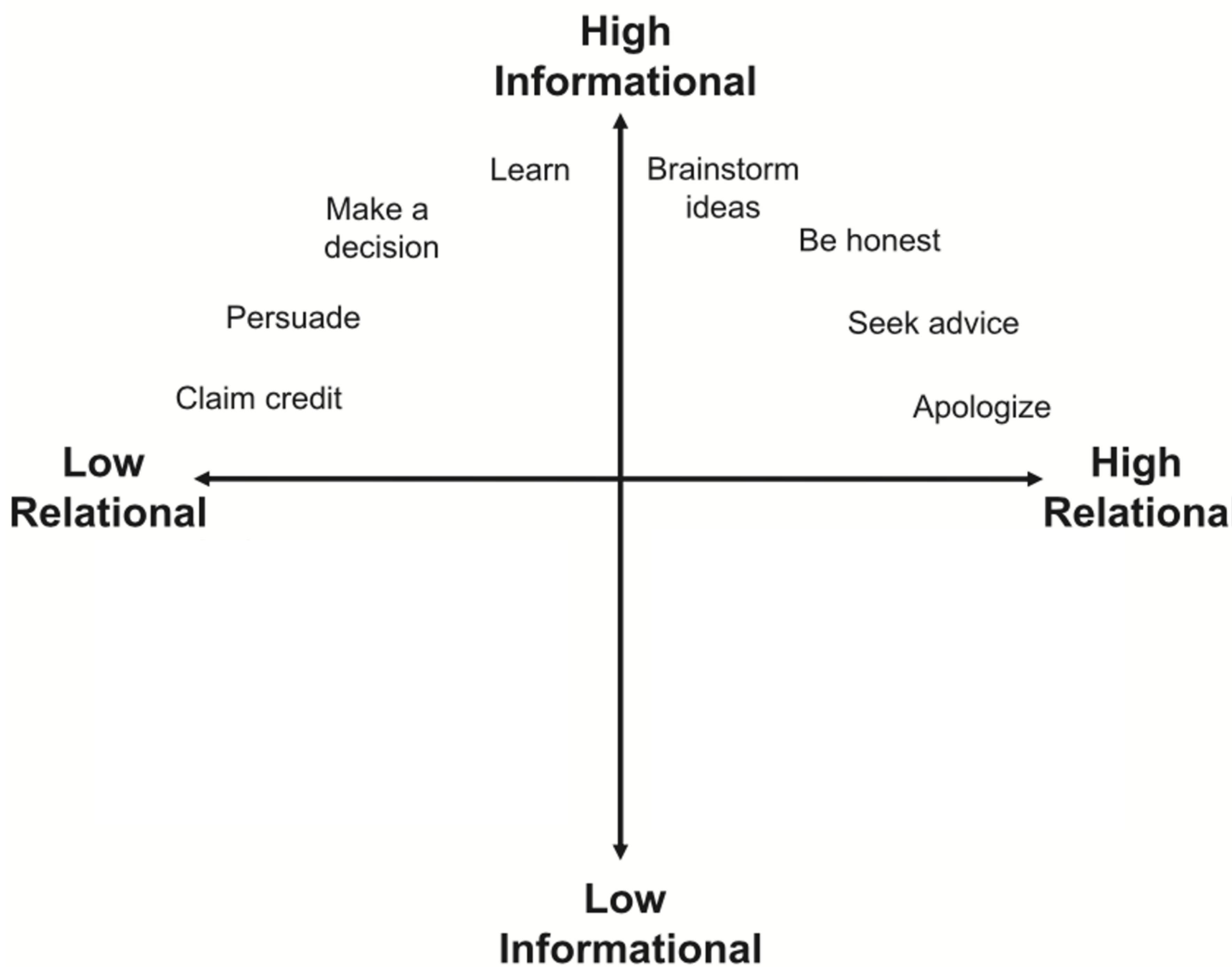
The Conversational Circumplex
(Yeomans, Schweitzer & Brooks, 2022)

Goals of Conversation



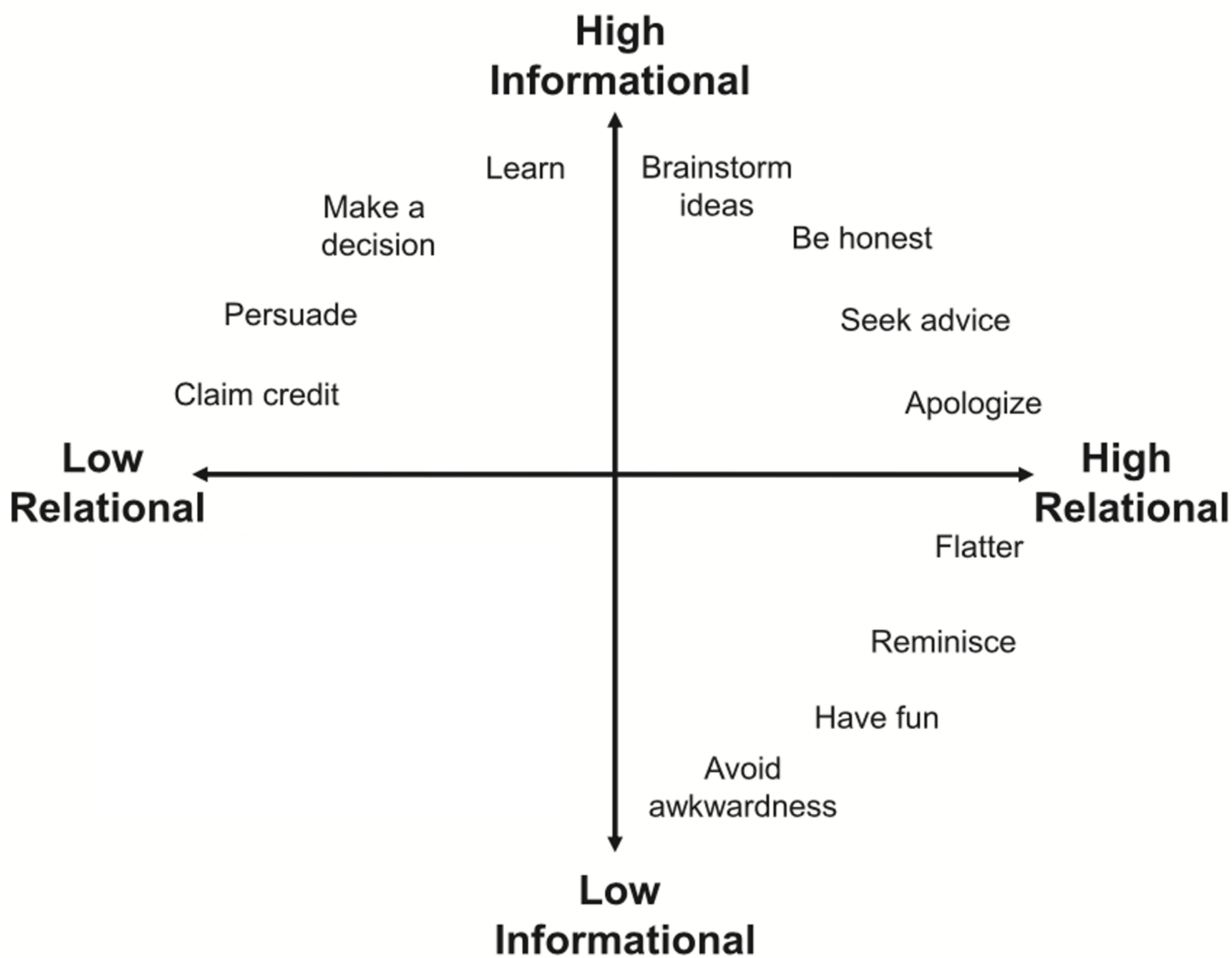
The Conversational Circumplex
(Yeomans, Schweitzer & Brooks, 2022)

Goals of Conversation



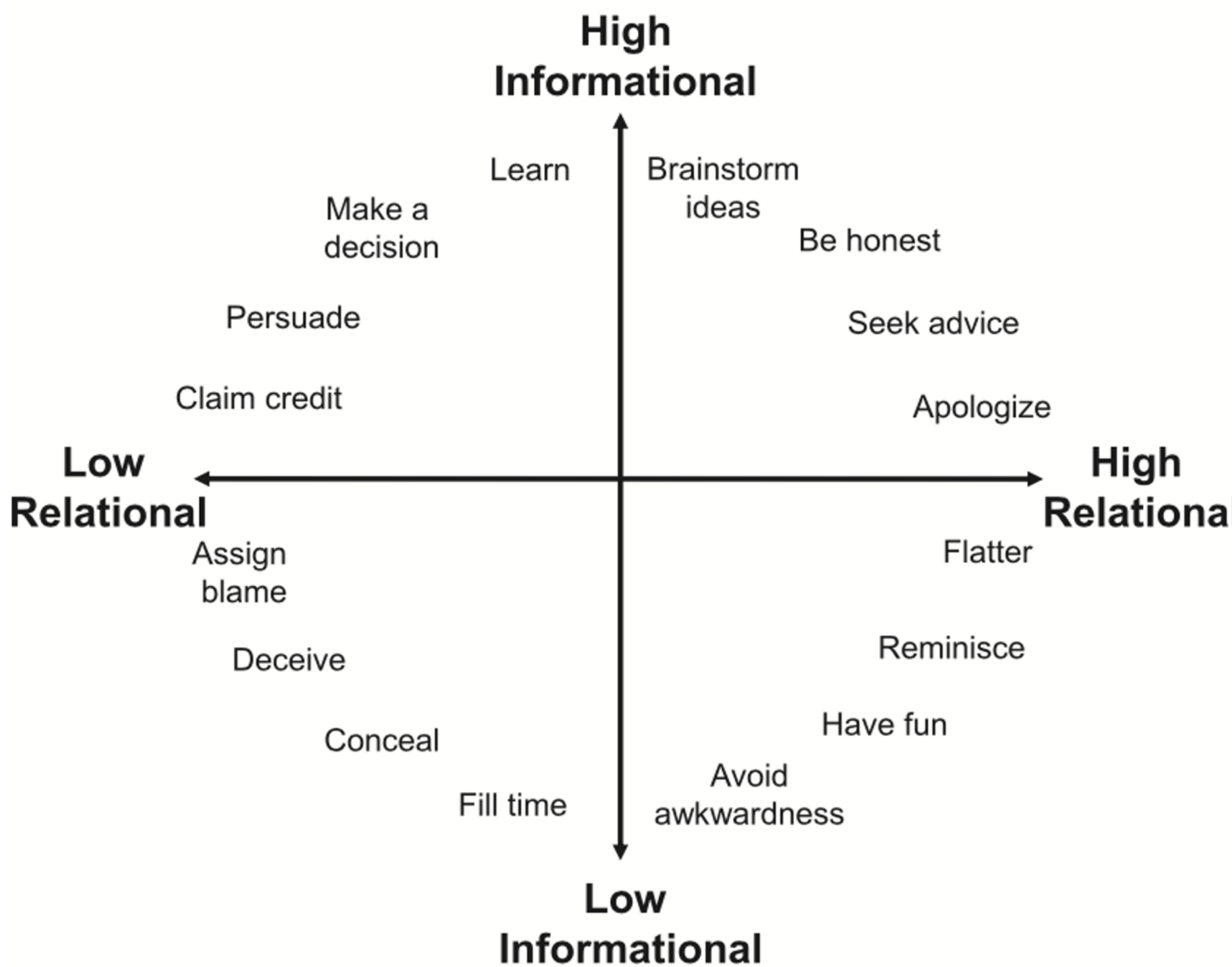
The Conversational Circumplex
(Yeomans, Schweitzer & Brooks, 2022)

Goals of Conversation



The Conversational Circumplex
(Yeomans, Schweitzer & Brooks, 2022)

Goals of Conversation



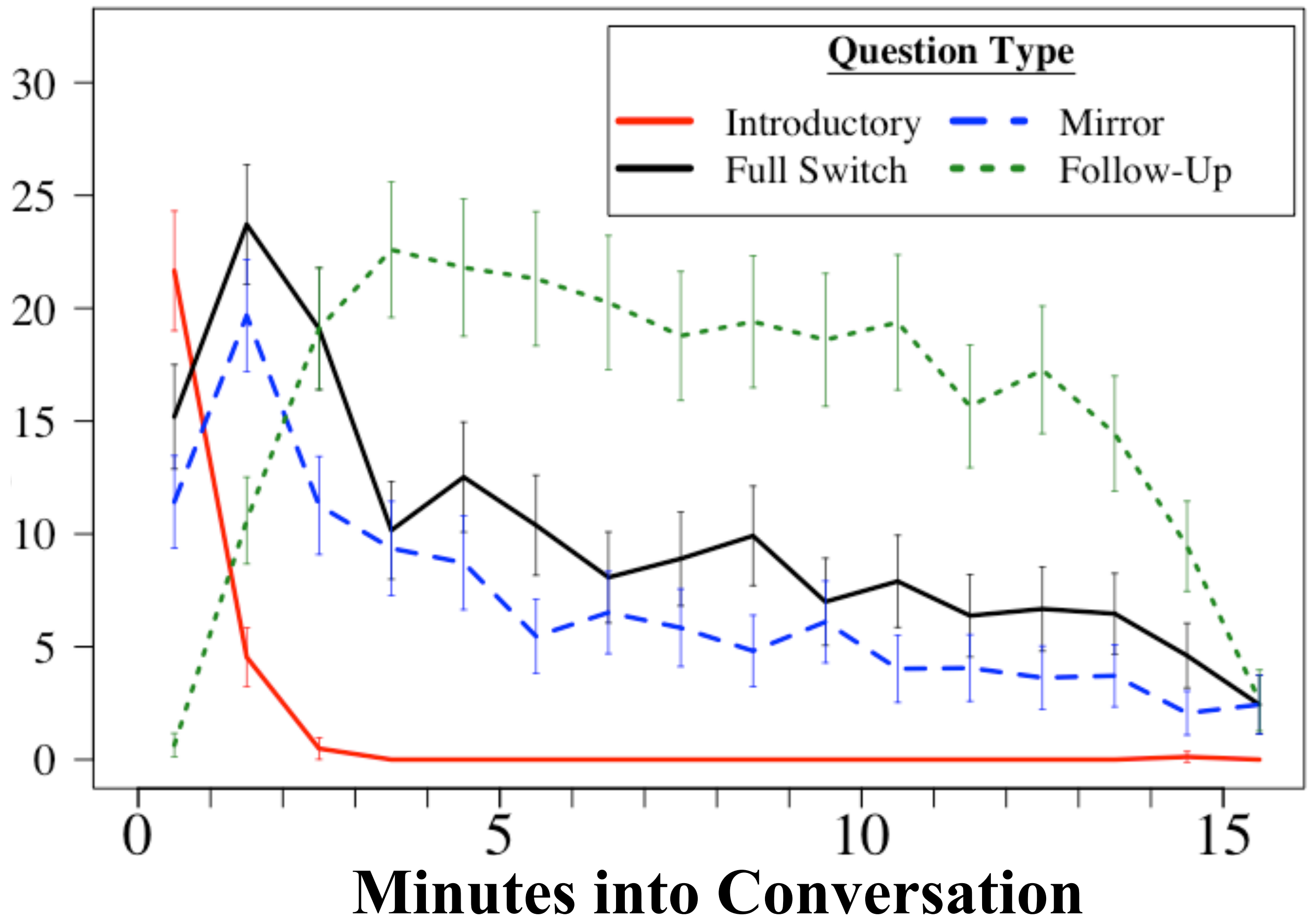
The Conversational Circumplex
(Yeomans, Schweitzer & Brooks, 2022)

Conversations versus SVDs

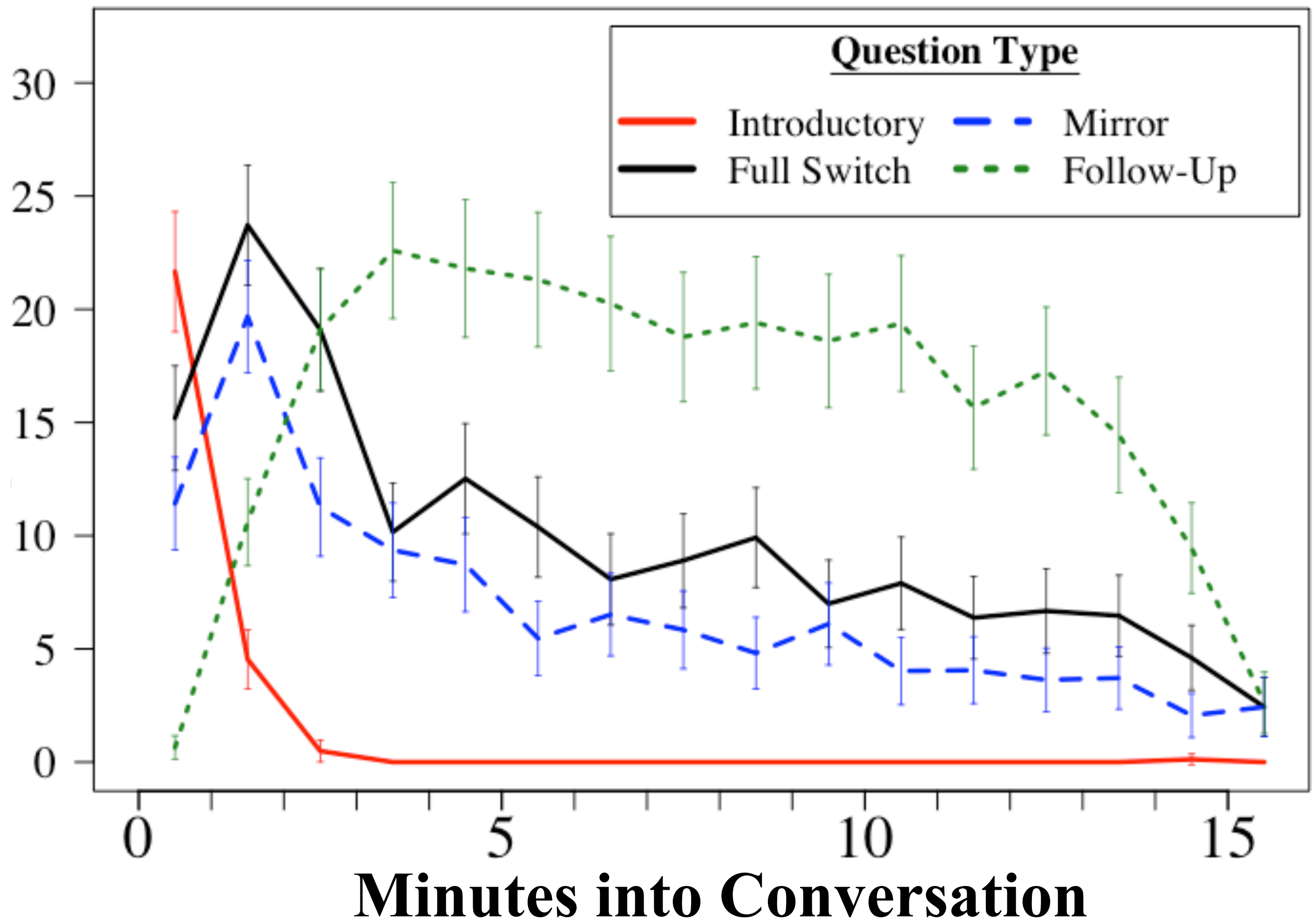
1. Multiple speakers: each with different goals, personalities, styles, etc

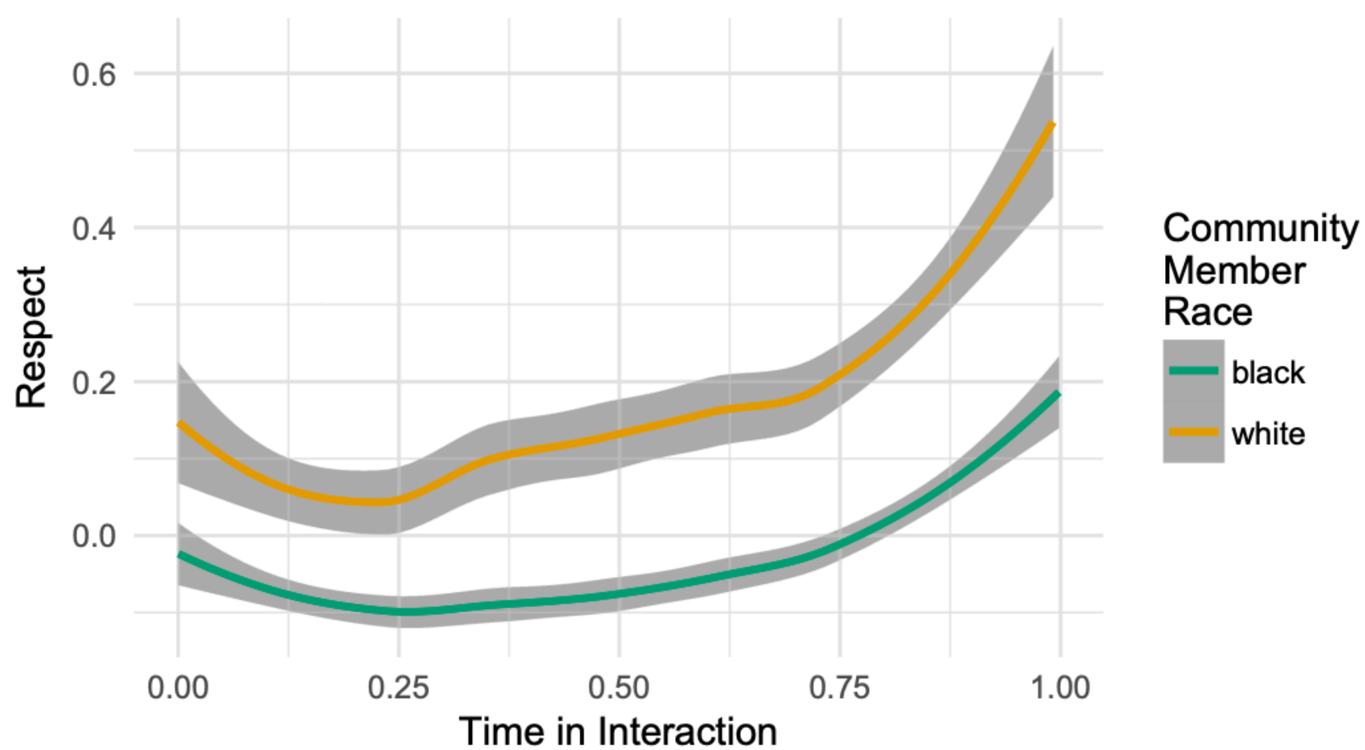
2. Generated responsively over time: transcripts can contain causes, correlates, and/or outcomes

The Time Course of Conversation

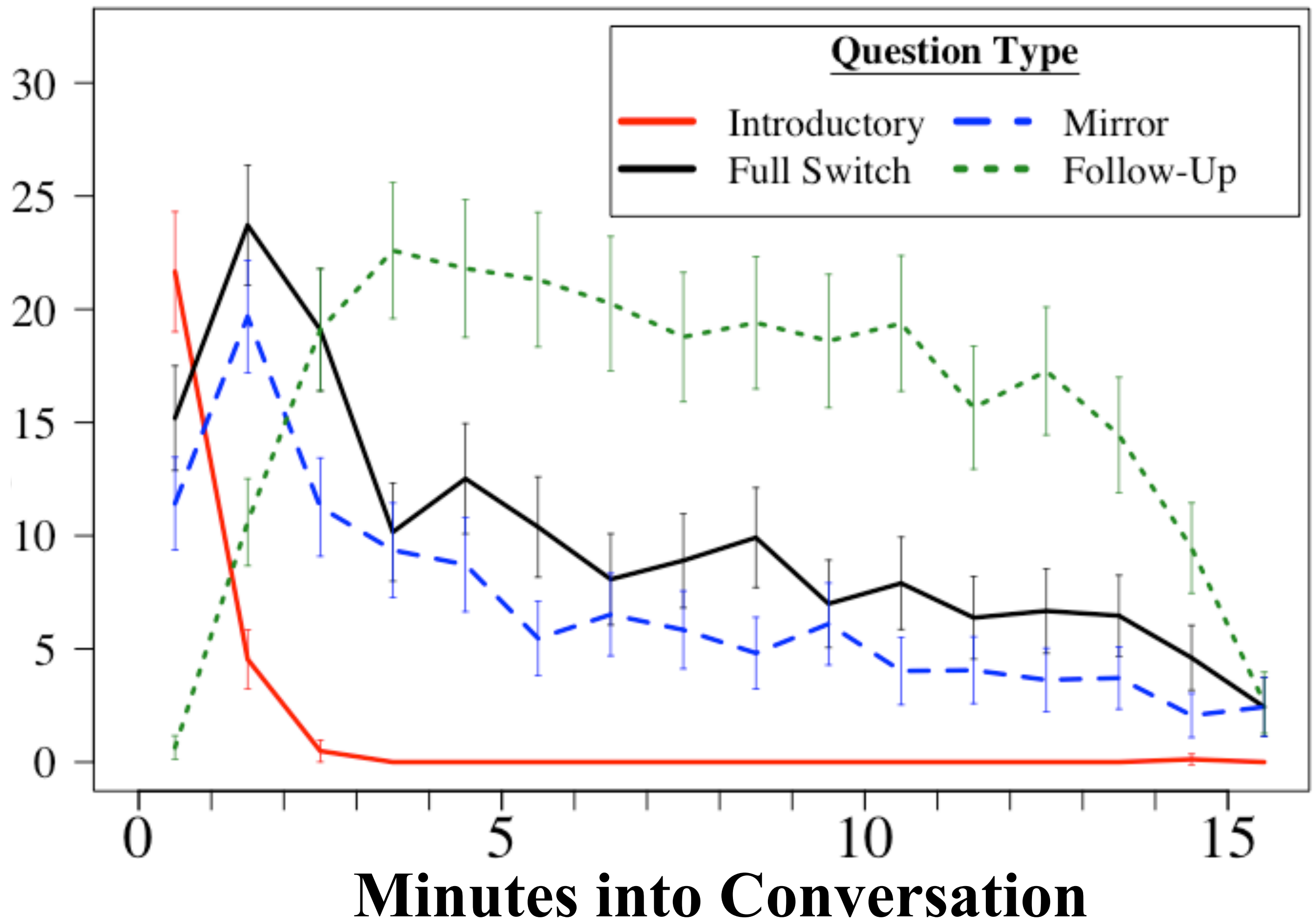


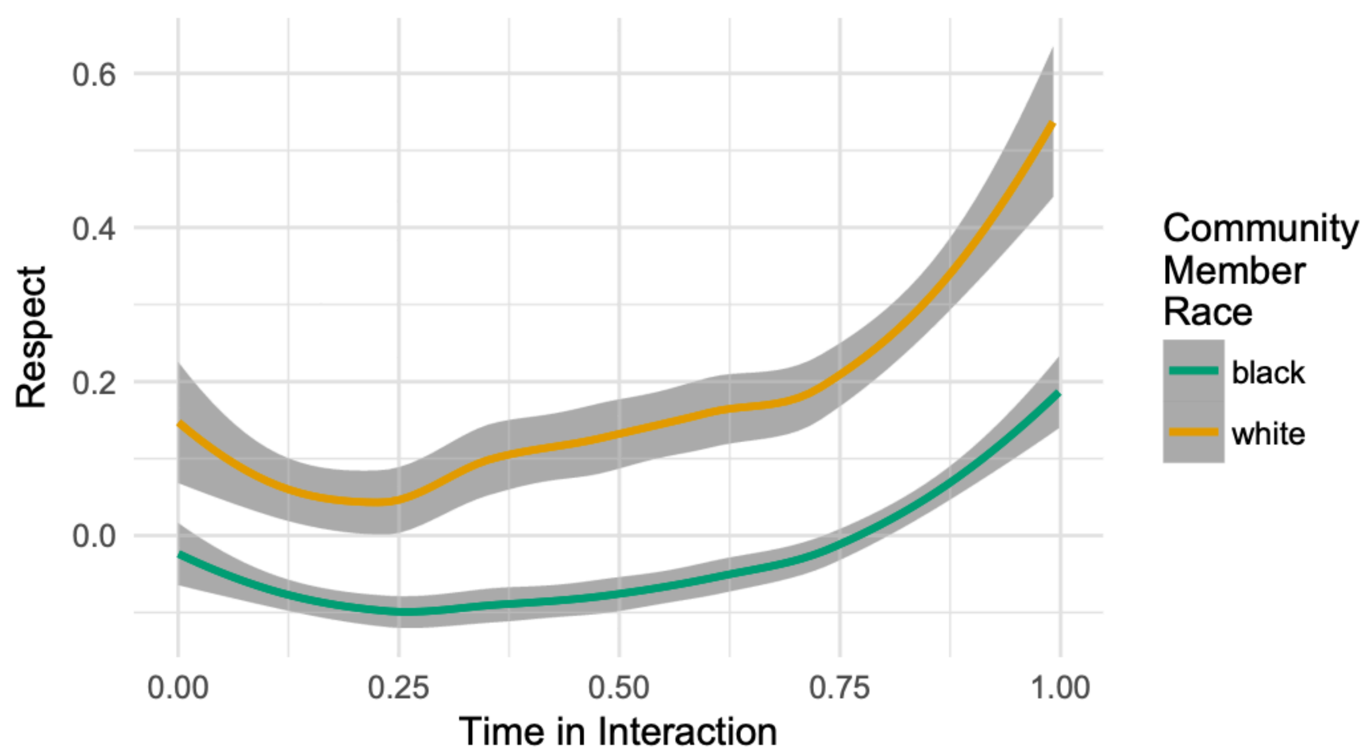
The Time Course of Conversation

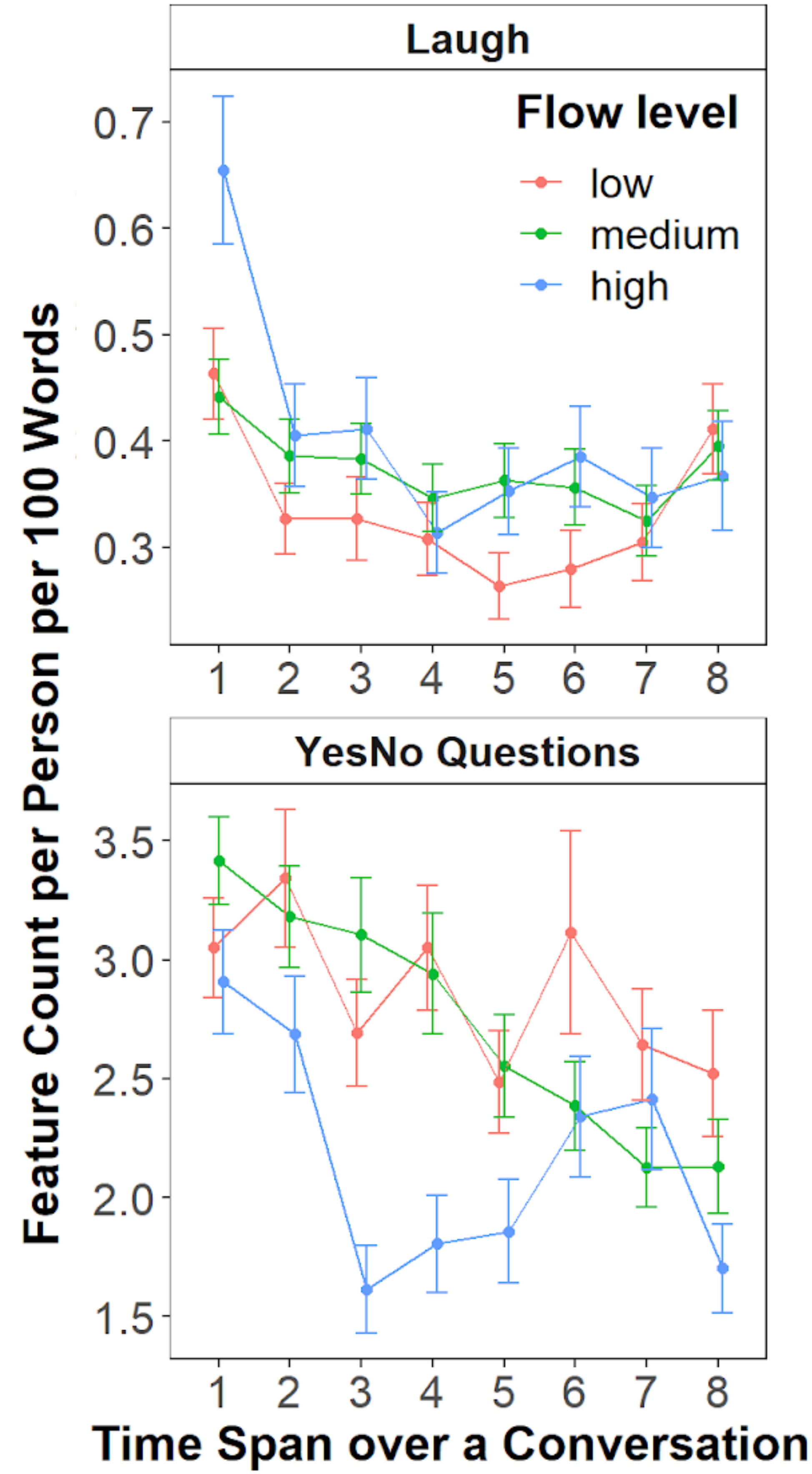




The Time Course of Conversation







Conversations versus SVDs

1. Multiple speakers: each with different goals, personalities, styles, etc

2. Generated responsively over time: transcripts can contain causes, correlates, and/or outcomes

3. Less formal: more pronouns, disfluencies, interruptions, backchannels, repairs, etc.

Conversations versus SVDs

1. Multiple speakers: each with different goals, personalities, styles, etc

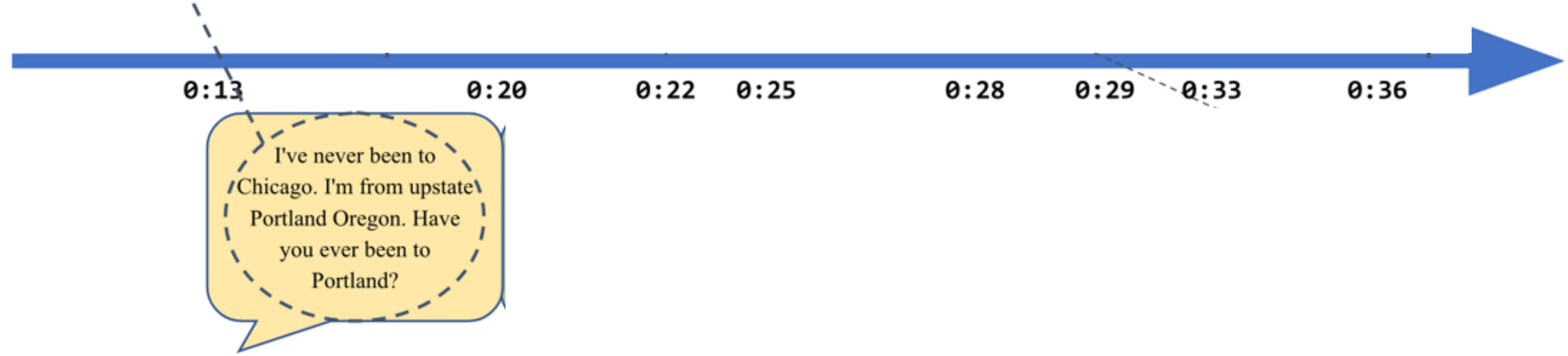
2. Generated responsively over time: transcripts can contain causes, correlates, and/or outcomes

3. Less formal: more pronouns, disfluencies, interruptions, backchannels, repairs, etc.

4. Dialogue acts: questions & answers, topic switching, agreement, affirmation, orders, acknowledgment, etc

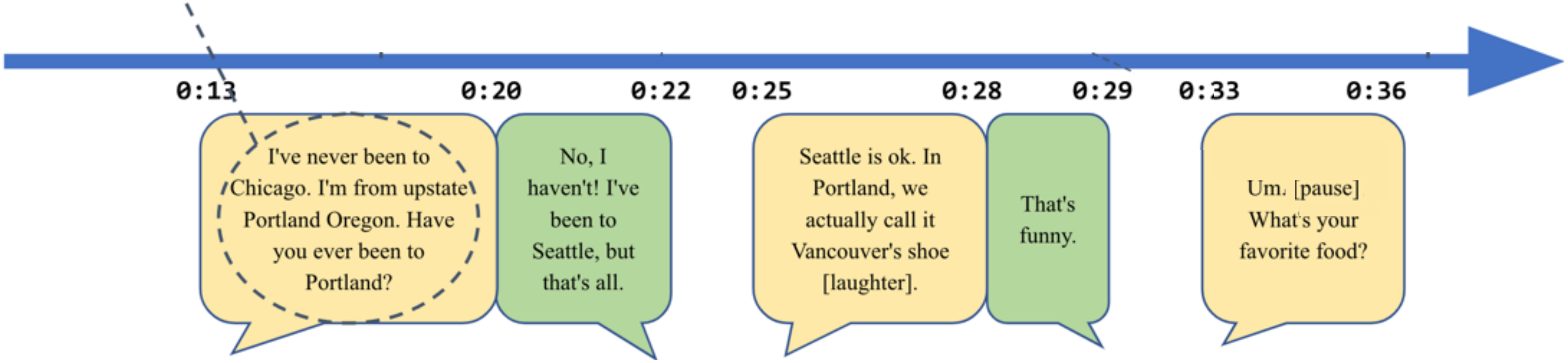
Conversation Features

4.3. Static Text Features
Word Counts
Dictionaries
Sentence Structure
Embeddings



Conversation Features

4.3. Static Text Features
Word Counts
Dictionaries
Sentence Structure
Embeddings



Conversation Features

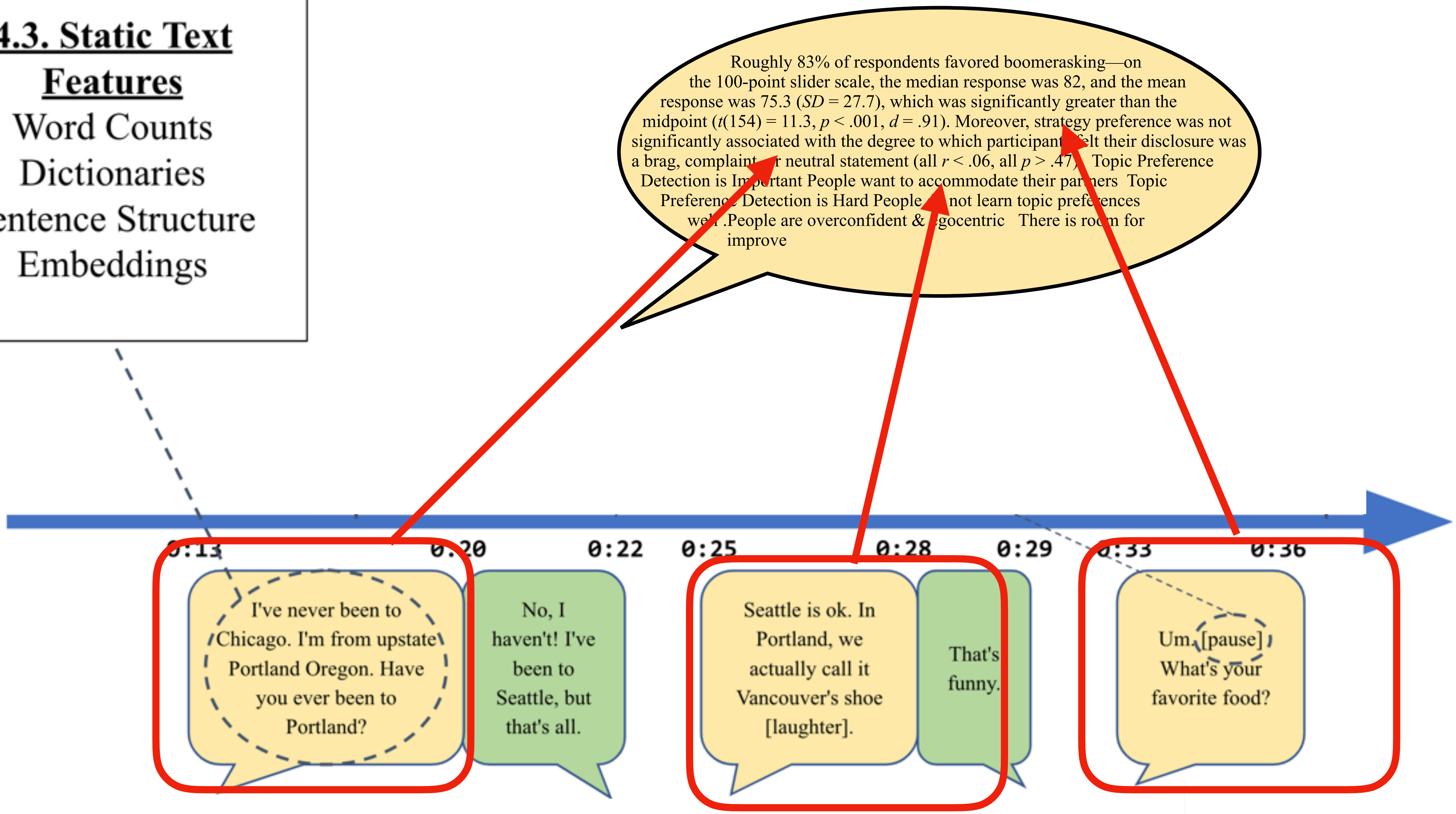
4.3. Static Text Features

Word Counts

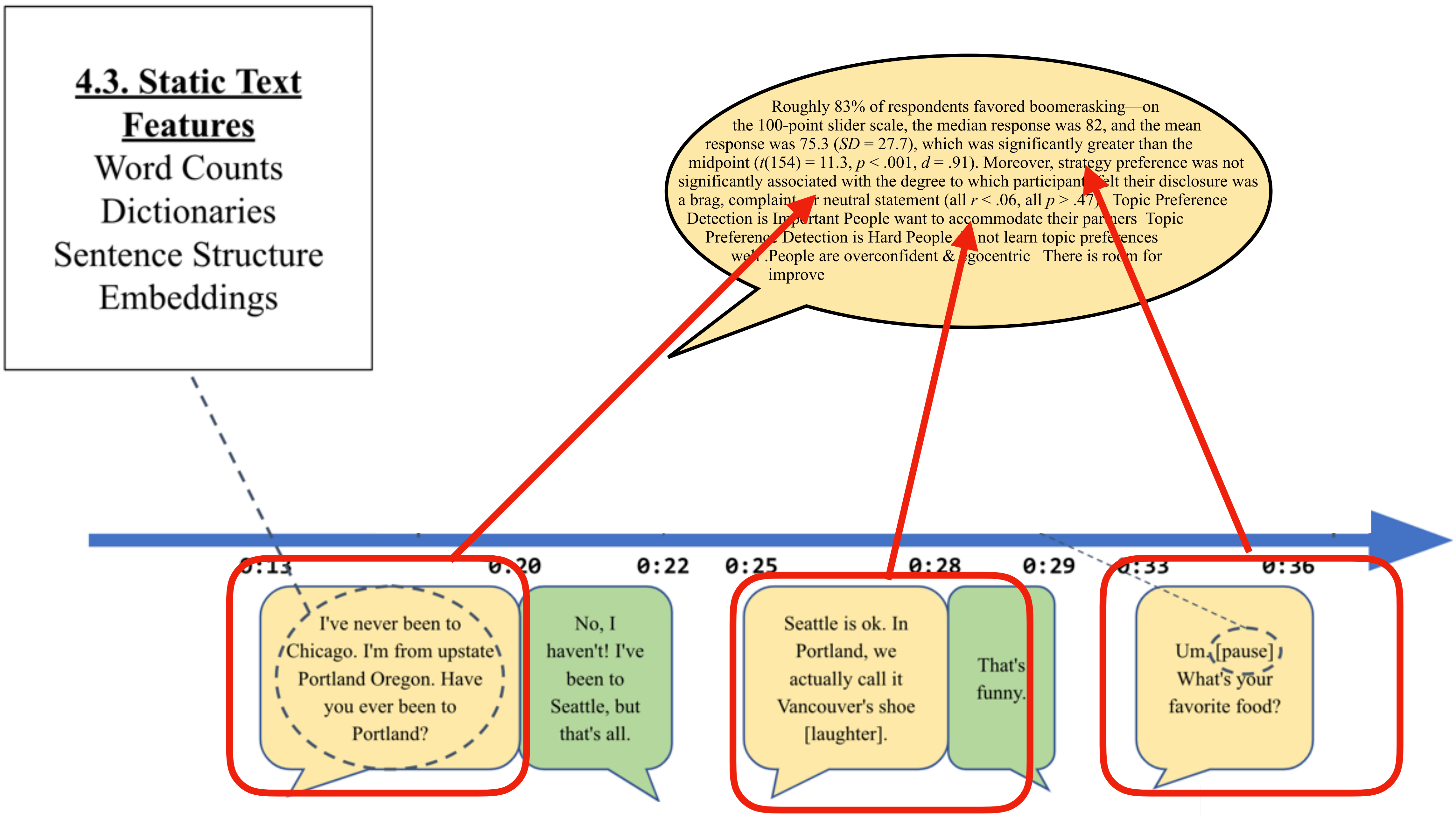
Dictionaries

Sentence Structure

Embeddings

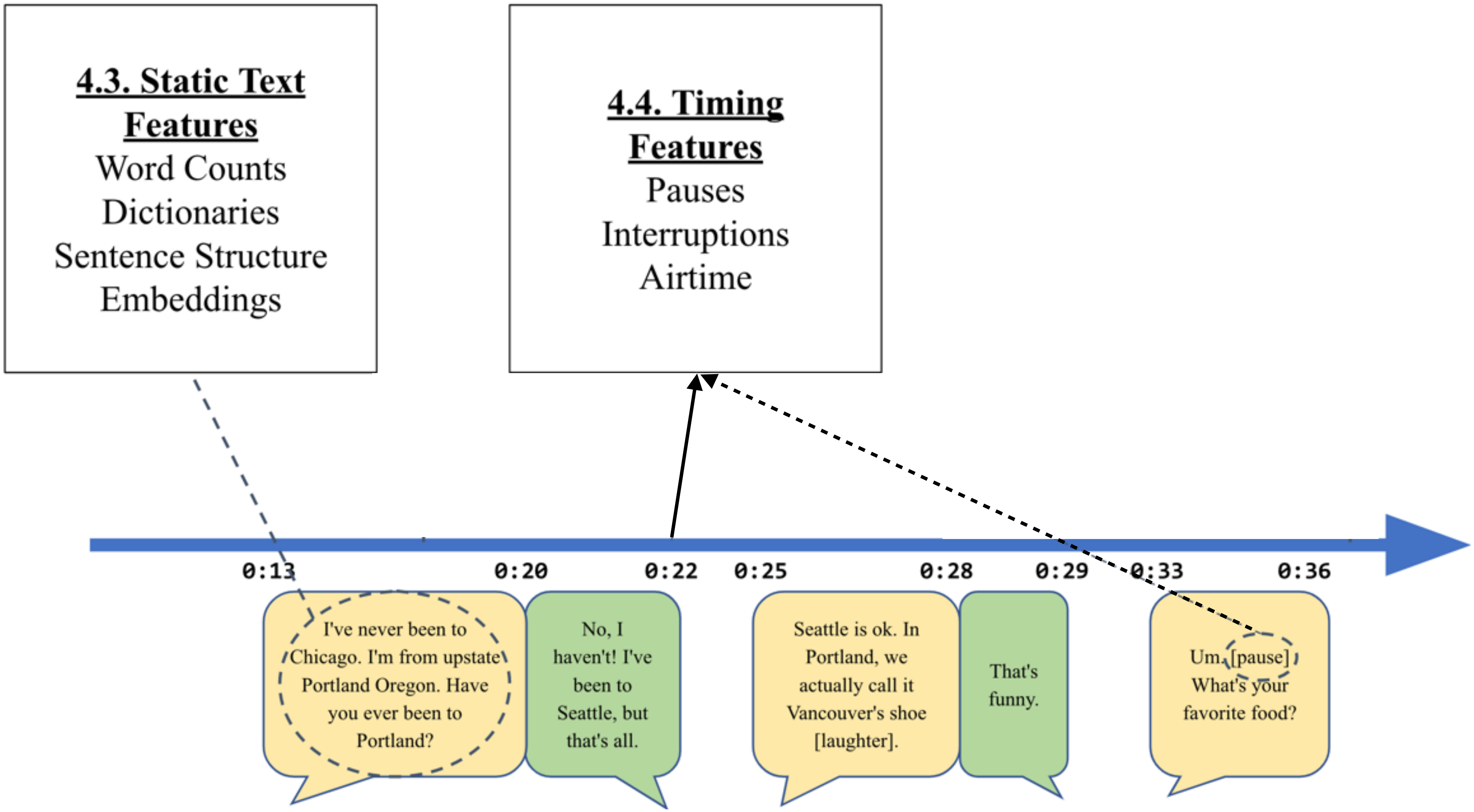


Conversation Features

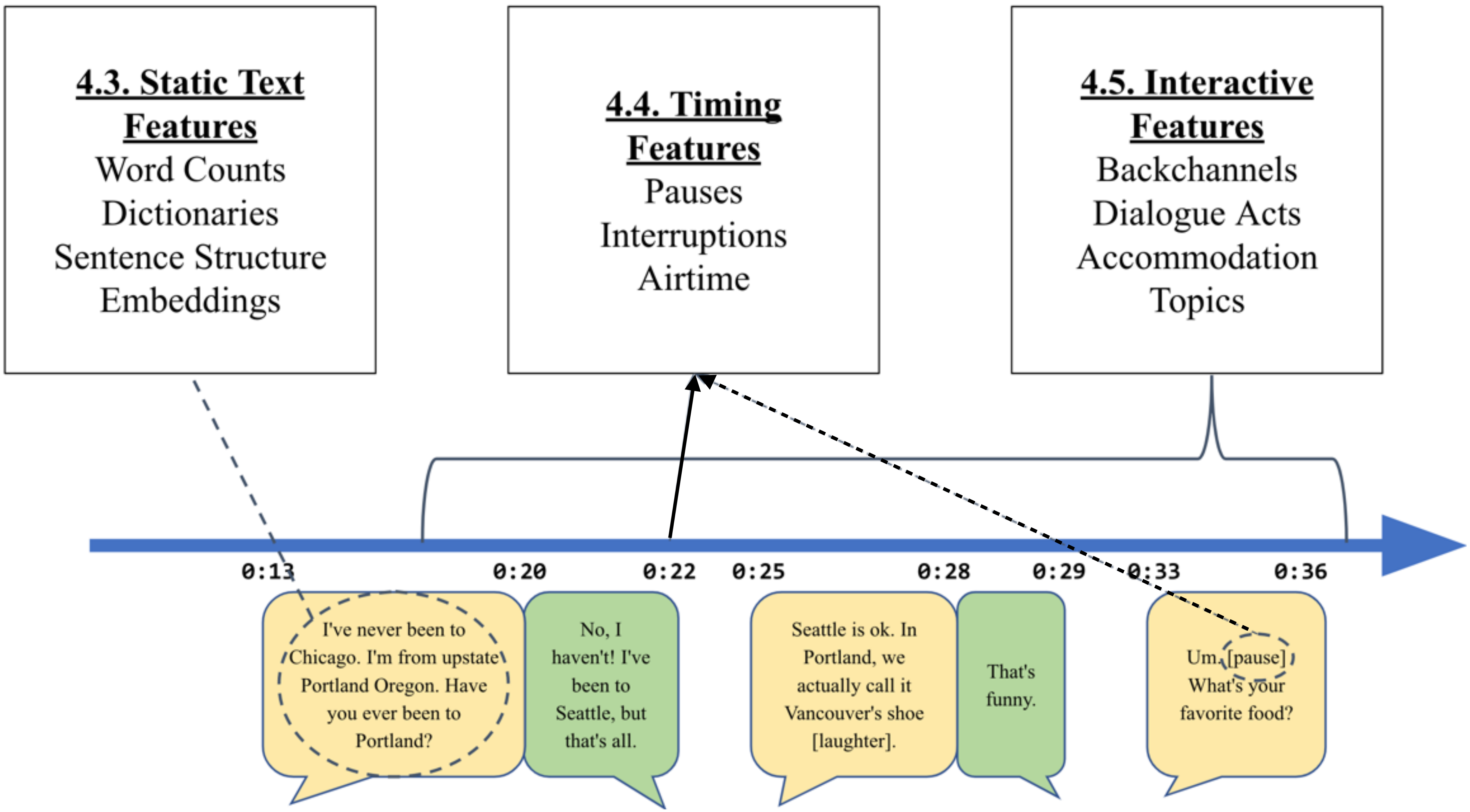


“Everything, Everywhere, All At Once”

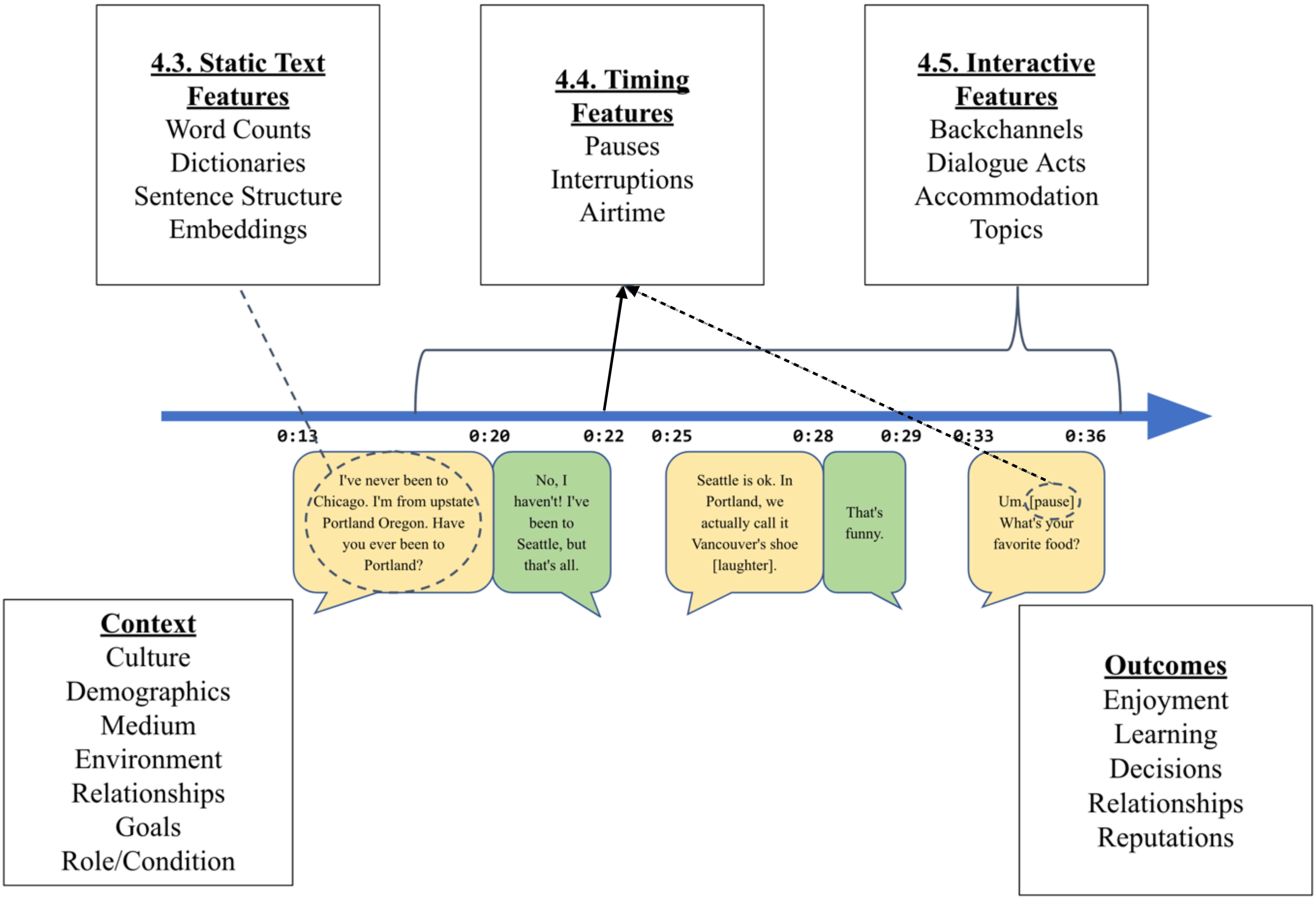
Conversation Features



Conversation Features



Conversation Features



Dialogue Acts

Tag	Example	%
Statement	*Me, I'm in the legal department.*	36%
Backchannel/Acknowledge	*Uh-huh.*	19%
Opinion	*I think it's great*	13%
Abandoned/Uninterpretable	*So, -/*	6%
Agreement/Accept	*That's exactly it.*	5%
Appreciation	*I can imagine.*	2%
Yes-No-Question	*Do you have to have any special training?*	2%
Non-verbal	*<Laughter>,<Throat_clearing>*	2%
Yes answers	*Yes.*	1%
Conventional-closing	*Well, it's been nice talking to you.*	1%
Wh-Question	*What did you wear to work today?*	1%
No answers	*No.*	1%
Response Acknowledgment	*Oh, okay.*	1%
Hedge	*I don't know if I'm making any sense or not.*	1%
Declarative Yes-No-Question	*So you can afford to get a house?*	1%
Other	*Well give me a break, you know.*	1%
Backchannel-Question	*Is that right?*	1%
Quotation	*You can't be pregnant and have cats*	.5%
Summarize/reformulate	*Oh, you mean you switched schools for the kids.*	.5%
Affirmative non-yes answers	*It is.*	.4%
Action-directive	*Why don't you go first*	.4%
Collaborative Completion	*Who aren't contributing.*	.4%
Repeat-phrase	*Oh, fajitas*	.3%
Open-Question	*How about you?*	.3%
Rhetorical-Questions	*Who would steal a newspaper?*	.2%
Hold before answer/agreement	*I'm drawing a blank.*	.3%
Reject	*Well, no*	.2%
Negative non-no answers	*Uh, not a whole lot.*	.1%
Signal-non-understanding	*Excuse me?*	.1%
Other answers	*I don't know*	.1%
Conventional-opening	*How are you?*	.1%
Or-Clause	*or is it more of a company?*	.1%
Dispreferred answers	*Well, not so much that.*	.1%
3rd-party-talk	*My goodness, Diane, get down from there.*	.1%
Offers, Options & Commits	*I'll have to check that out*	.1%
Self-talk	*What's the word I'm looking for*	.1%
Downplayer	*That's all right.*	.1%
Maybe/Accept-part	*Something like that*	<.1%
Tag-Question	*Right?*	<.1%
Declarative Wh-Question	*You are what kind of buff?*	<.1%
Apology	*I'm sorry.*	<.1%
Thanking	*Hey thanks a lot*	<.1%

Stolcke et al., 2000

Distinctive Question Features

hello, what things do you enjoy doing?
i love travelling
me too
are you a boston native ?
no, I am from Colombia. and you?

Distinctive Question Features

Contextual Feature	Follow-Up	Switch	Intro	Mirror
Word Count of turn	.20	—	—	—
Time into Conversation	.25	—	-1.34	—
Question in askee's last turn	—	—	-.43	.61
Pre-question statement in turn	-.10	—	—	.37
Question in asker's last turn	.30	.12	-.33	-.12
Multiple questions in the turn	.08	-.04	—	—

Distinctive Question Features

Do you play an instrument or enjoy listening to music?
I played the piano pretty seriously for about 13 years
whoa that's awesome. Did you like that?

Distinctive Question Features

Do you play an instrument or enjoy listening to music?
I played the piano pretty seriously for about 13 years
whoa that's awesome. Did you like that?
I don't play that much anymore, just occasionally when I go home. Yeah I loved it. Only classical music but I still enjoyed it so much. And now obviously I still really like listening to music, and producing it.
Classical? or other genres
you should try out Snarky Puppy. it's an instrumental based music. my favorite band
Cool! I'll definitely check it out. What kind of music do you mostly produce? And how'd you get started doing it?

The Detection Model

Q1
Q2
Q3
Q4
Q5

The Detection Model

	who	what	when	why	how	hobbies
Q1	0	1	0	0	0	0
Q2	0	0	1	0	0	0
Q3	0	1	0	1	1	1
Q4	1	0	0	1	0	1
Q5	0	0	0	0	0	0

Ngram Features

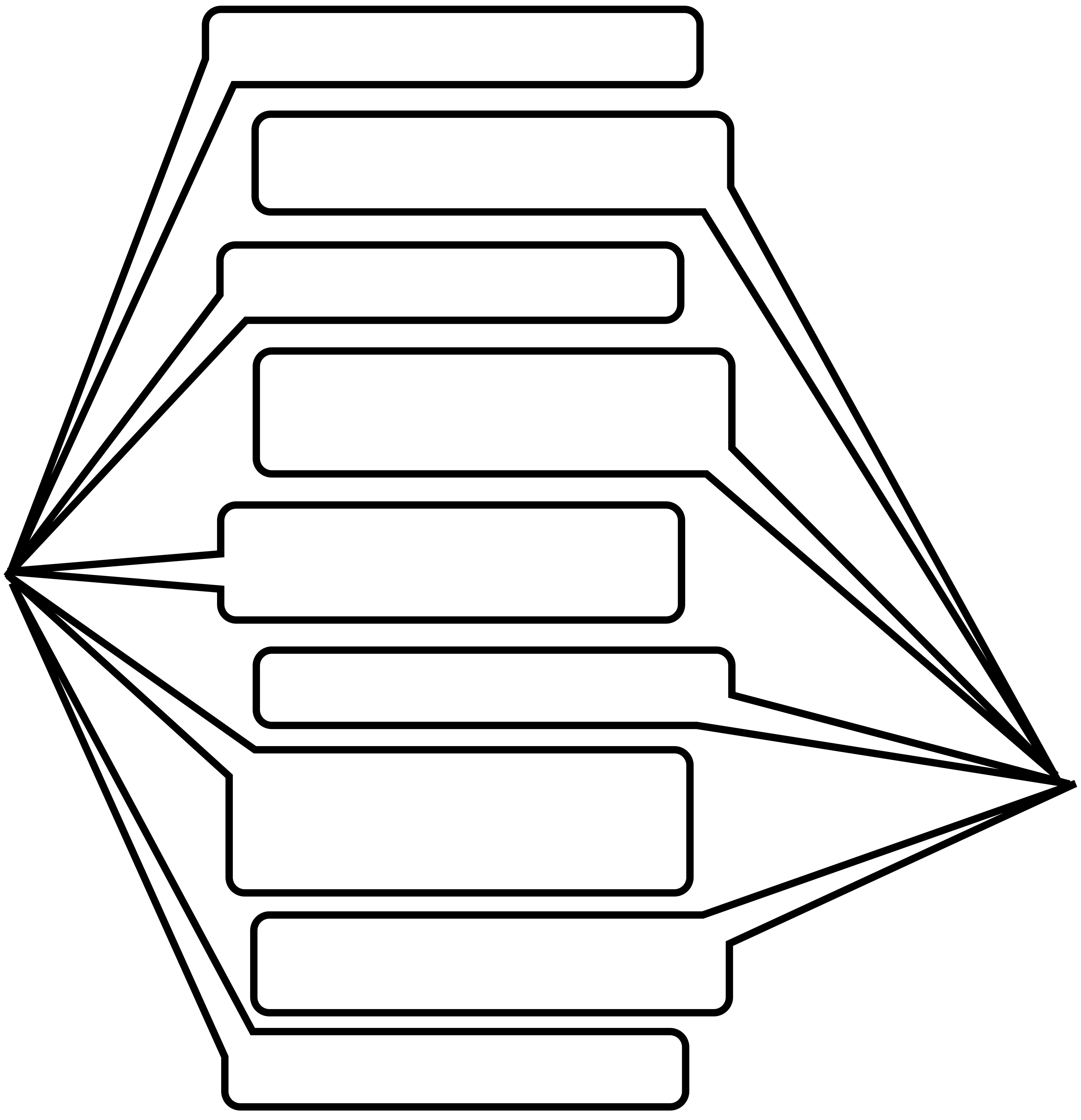
The Detection Model

	who	what	when	why	how	hobbies	Word count	prevQ asker	prevQ askee	turn Q count
Q1	0	1	0	0	0	0	11	0	0	1
Q2	0	0	1	0	0	0	26	0	1	2
Q3	0	1	0	1	1	1	8	1	0	1
Q4	1	0	0	1	0	1	15	1	0	1
Q5	0	0	0	0	0	0	19	0	0	2

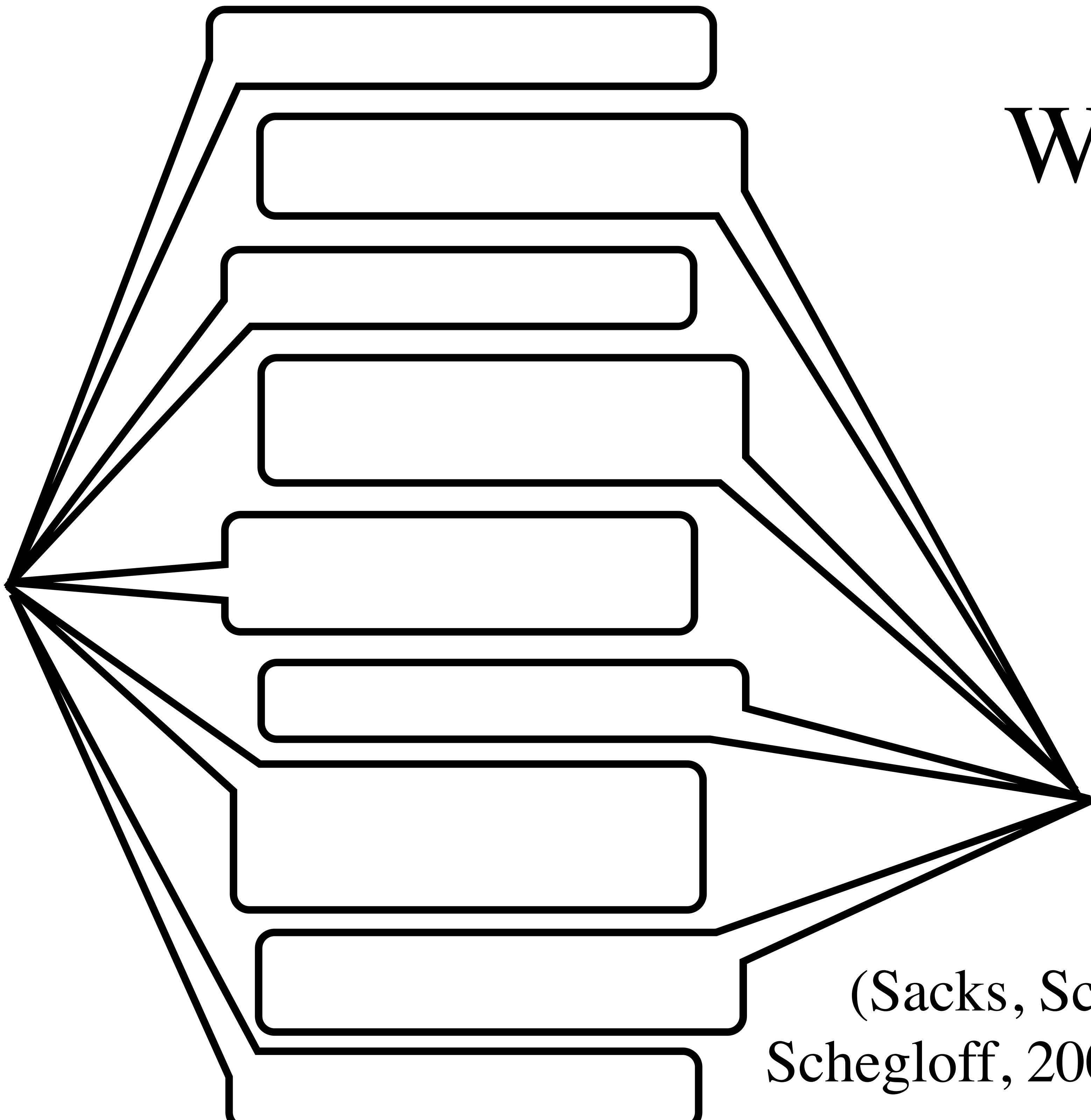
Ngram Features

Context Features

The Structure of Conversation



The Structure of Conversation

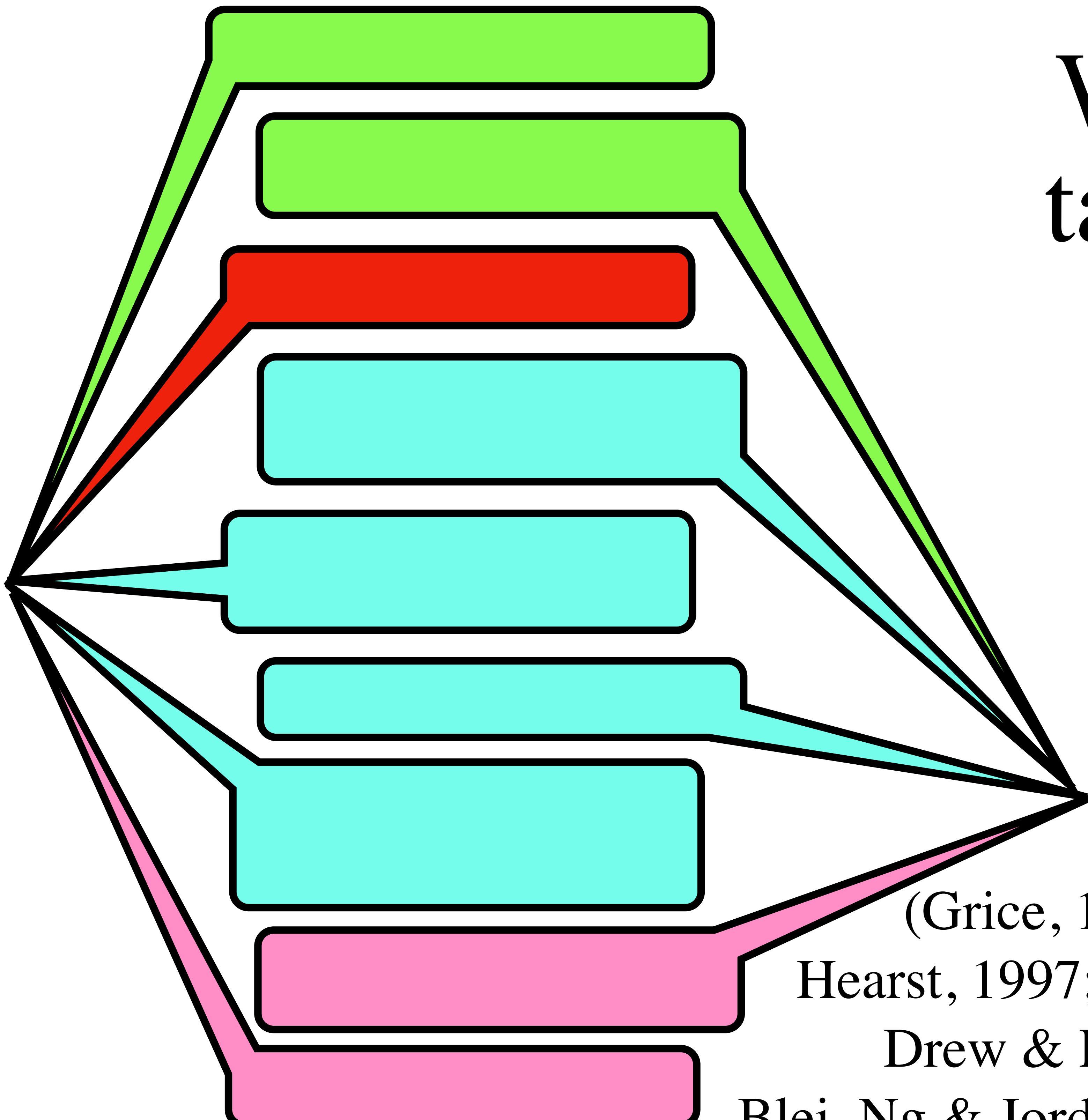


Who is talking

Turn-Taking

(Sacks, Schegloff & Jackson, 1979; Schegloff, 2007; Jurafsky & Martin, 2019)

The Structure of Conversation

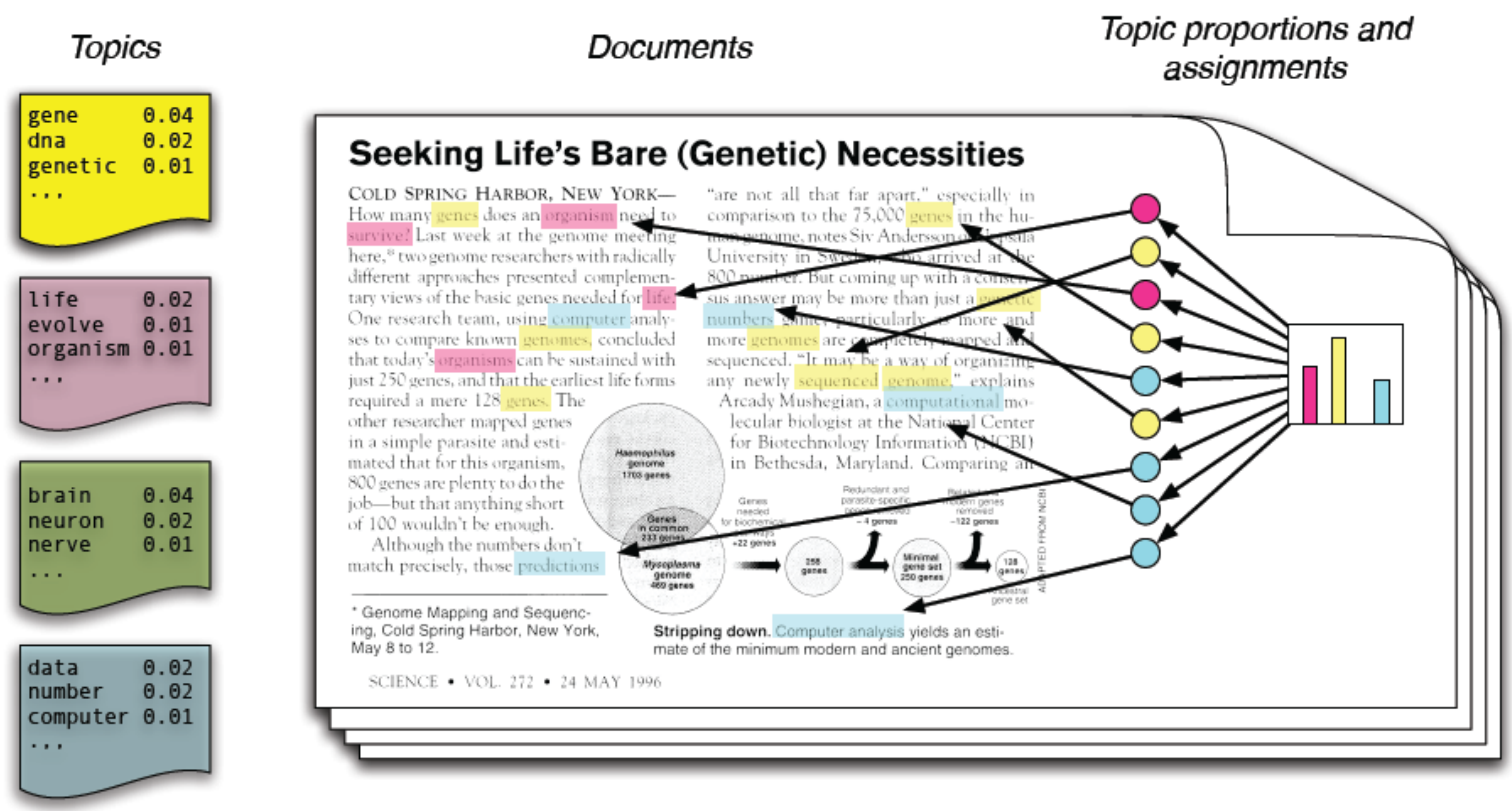


What we are talking about

Topic Selection

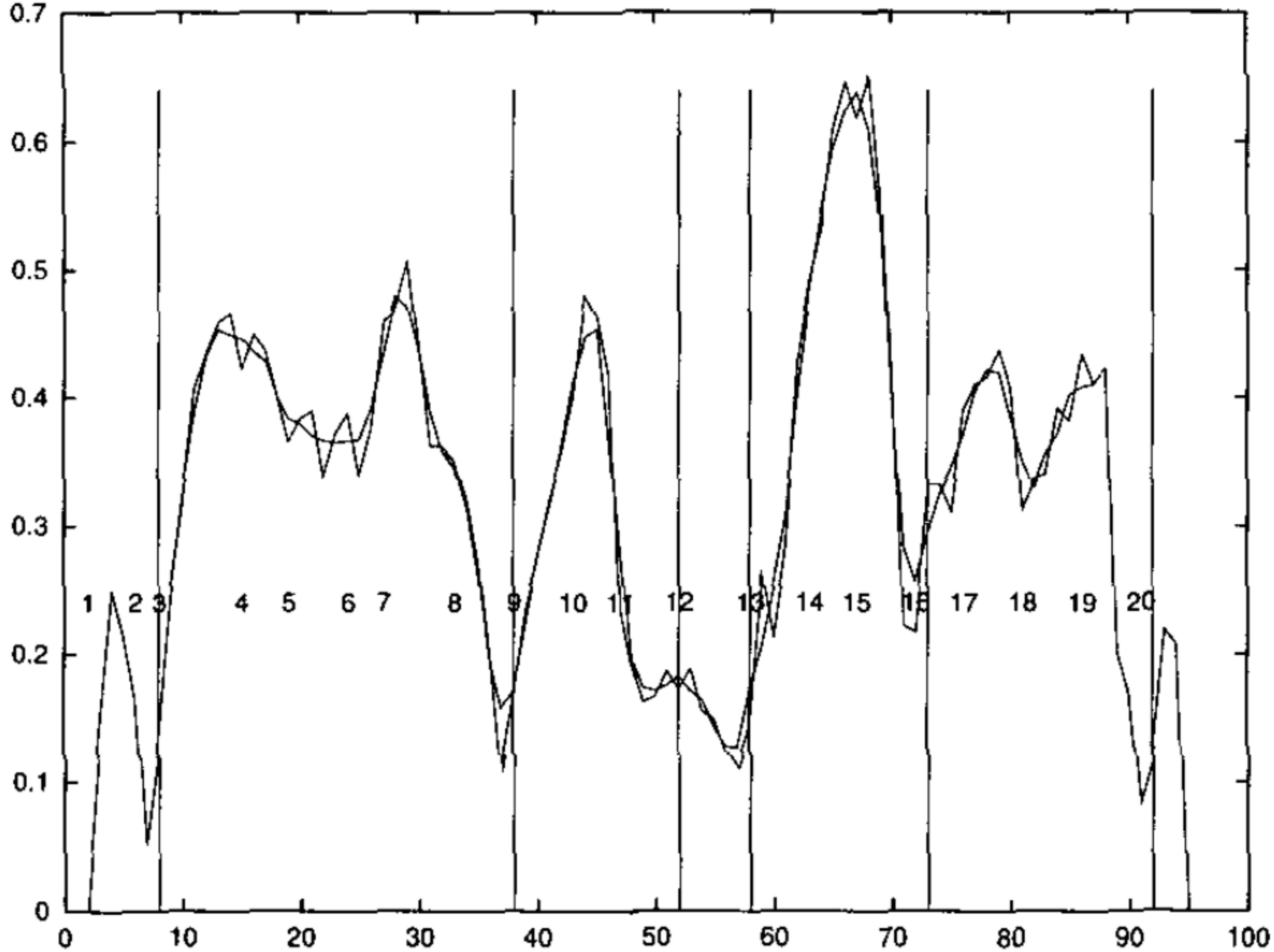
(Grice, 1975; Hardin & Higgins, 1996; Hearst, 1997; Passonneau & Litman, 1997; Drew & Holt, 1998; Galley et al., 2003; Blei, Ng & Jordan, 2003; Nguyen et al., 2014)

Topic Detection



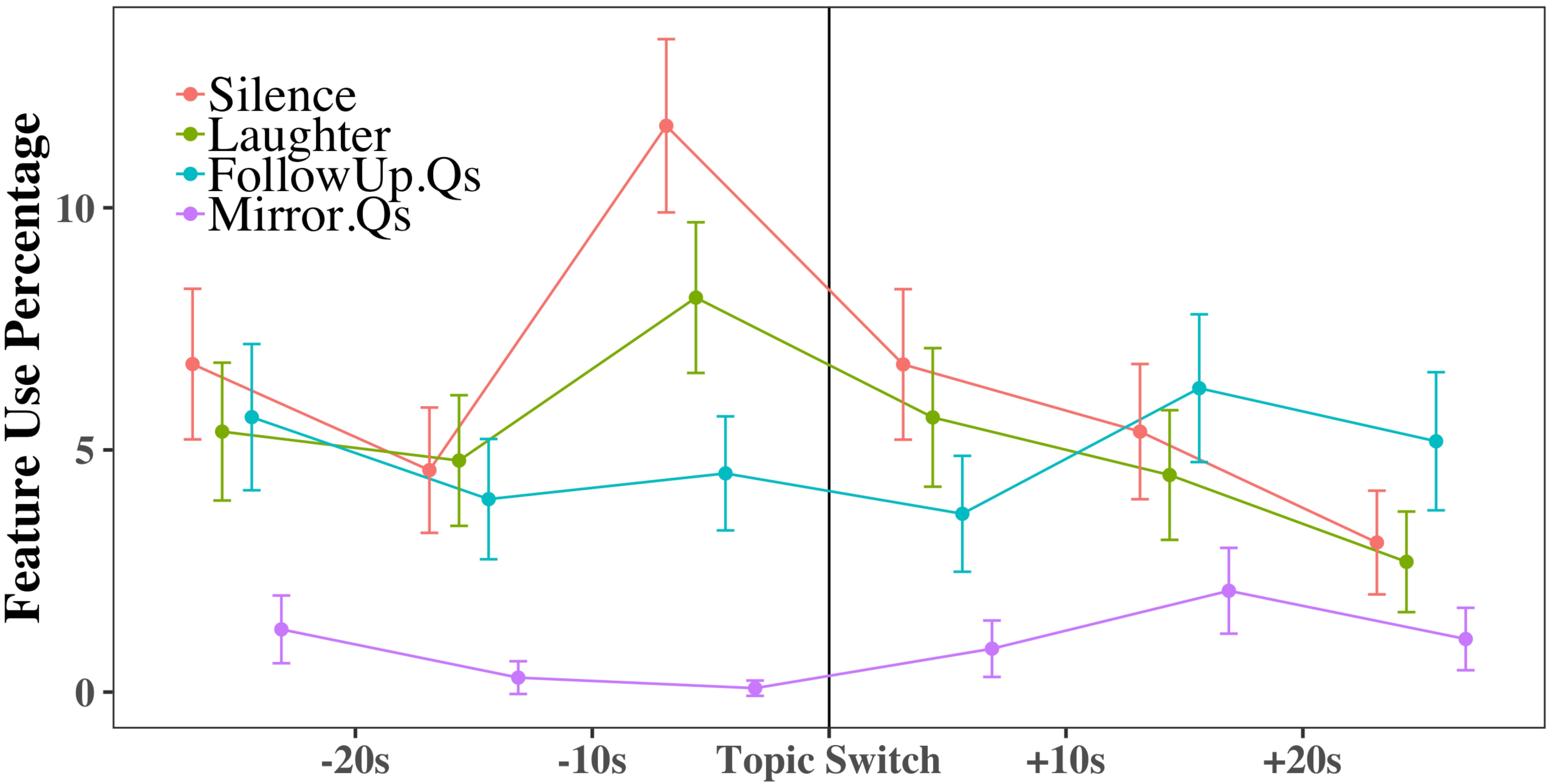
(Blei, Ng & Jordan, 2003; Blei & Lafferty, 2007)

Topic Segmentation



(**Hearst, 1997;** Passonneau & Litman, 1997; Drew & Holt, 1998; Galley et al., 2003; Purver, 2011; Nguyen et al., 2014)

Topic Transitions



(Yeomans & Brooks, 2024)

Open vs Structured Domain

Open domain

- Speed date
- Social media
- Dinner with friends
- Work meeting

Open vs Structured Domain

Open domain

Speed date

Social media

Dinner with friends

Work meeting

Structured domain

Open vs Structured Domain

Open domain

- Speed date
- Social media
- Dinner with friends
- Work meeting

Structured domain

- Booking agents
- Sales calls

Open vs Structured Domain

Open domain

Speed date

Social media

Dinner with friends

Work meeting

Structured domain

Booking agents

Sales calls

Open vs Structured Domain

Open domain

Speed date
Social media
Dinner with friends
Work meeting

Structured domain

Booking agents
Sales calls
Customer service
Crisis hotlines

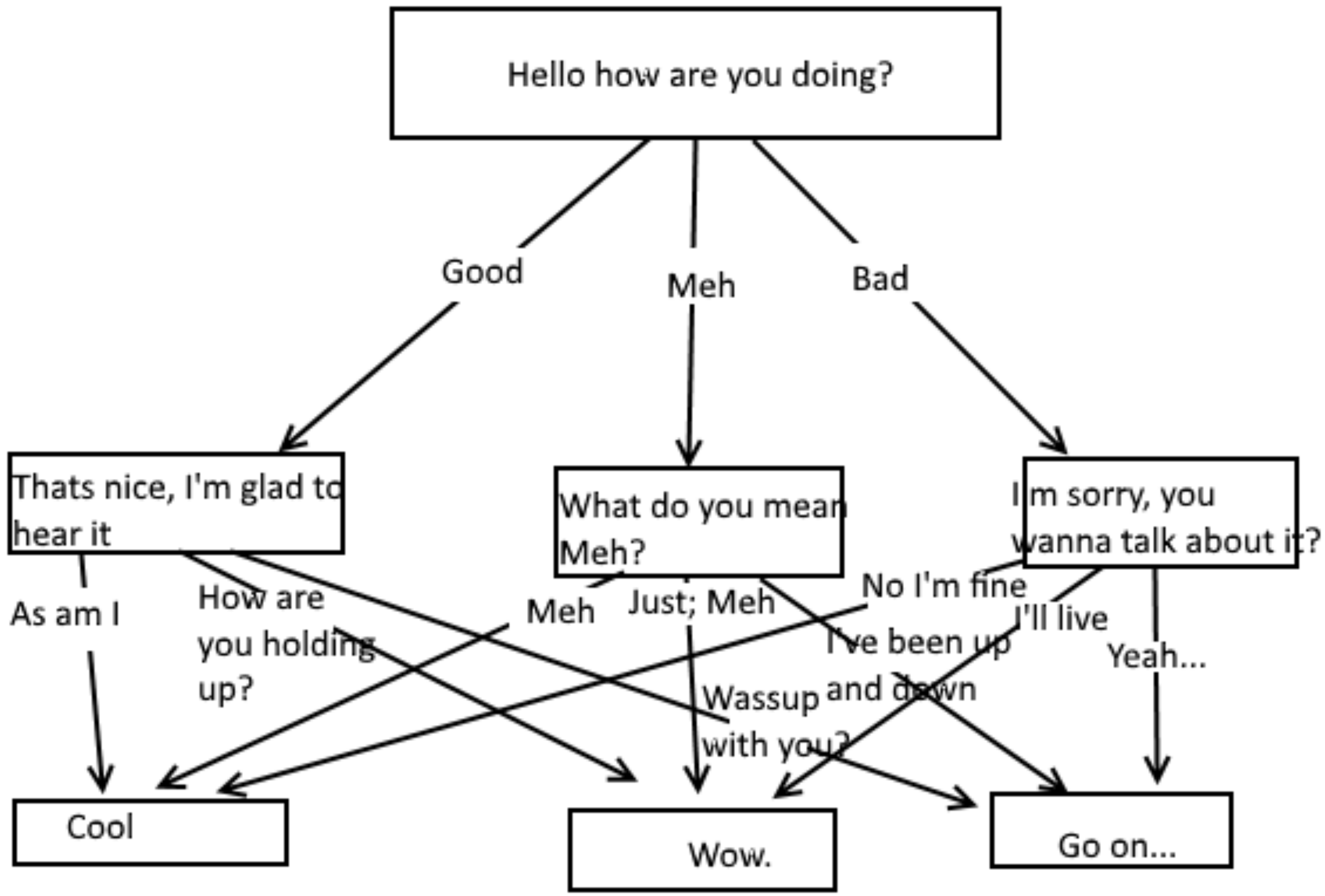
Open vs Structured Domain

Open domain

- Speed date
- Social media
- Dinner with friends
- Work meeting

Structured domain

- Booking agents
- Sales calls
- Customer service
- Crisis hotlines



Open vs Structured Domain

Open domain

Speed date

Social media

Dinner with friends

Work meeting

Structured domain

Booking agents

Sales calls

Customer service

Crisis hotlines

Interviews?

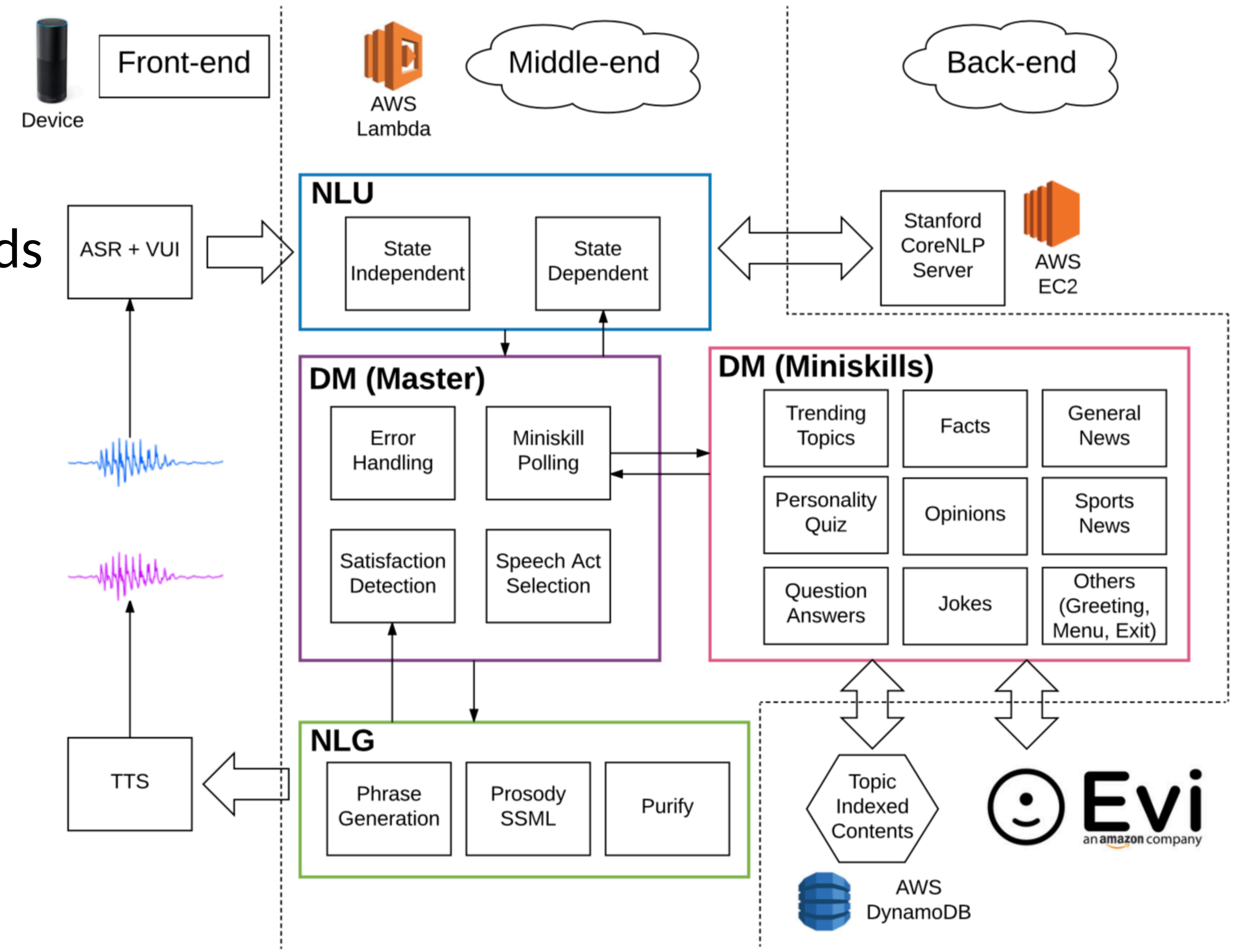
Negotiations?

Alexa?

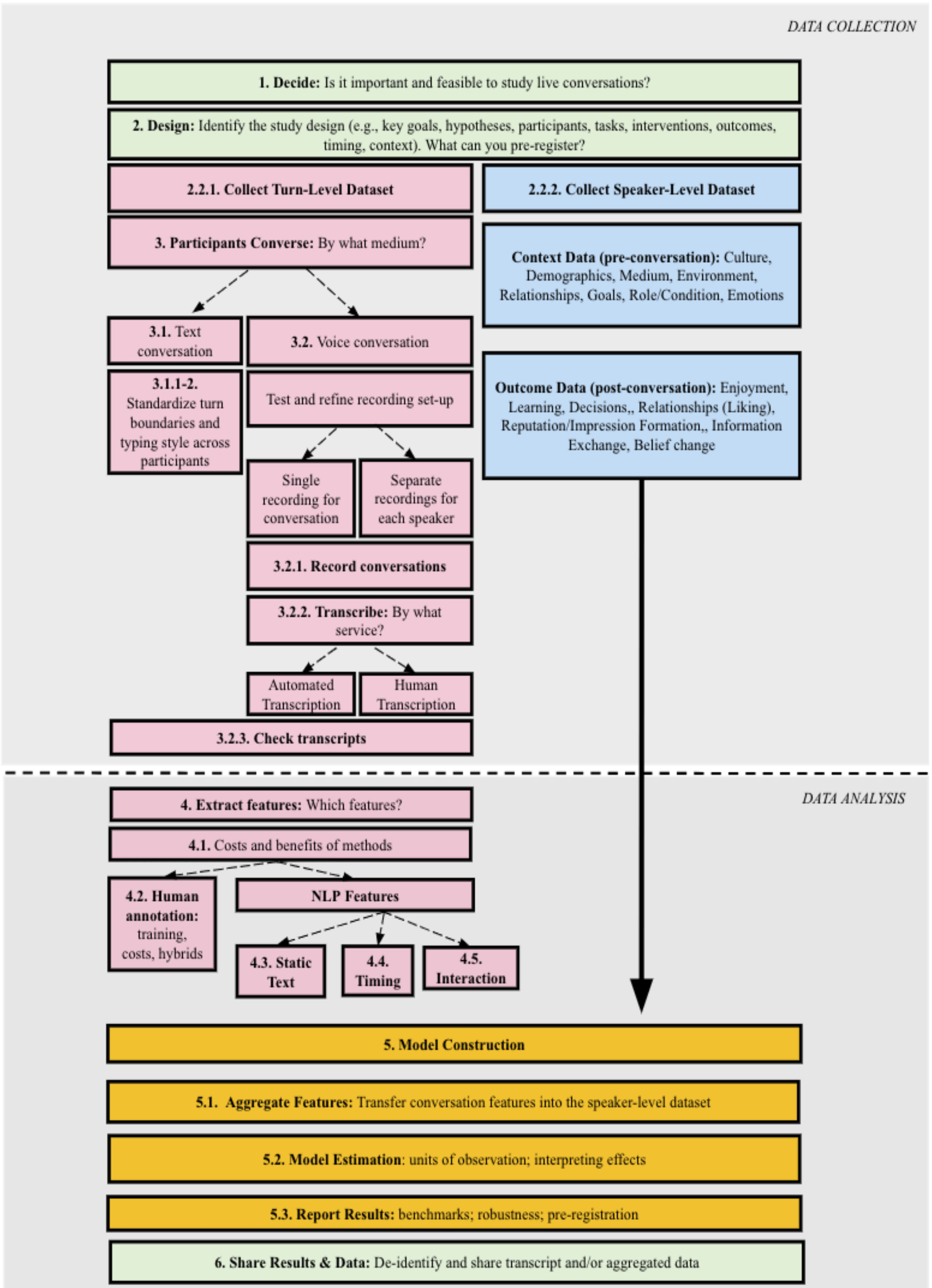
Open vs Structured Domain

Open domain
- Speed date
- Social media
- Dinner with friends
- Work meeting

Structured domain
- Booking agents
- Sales calls
- Customer service
- Crisis hotlines
- Interviews?
- Negotiations?
- Alexa?



The Whole Paper in one Figure



A Practical Guide To Conversation Research

(Yeomans, Boland, Collins, Abi-Esber & Brooks, 2023)

Conversations are Everywhere

Relationships As Conversations Over Time

Let's keep the conversation going