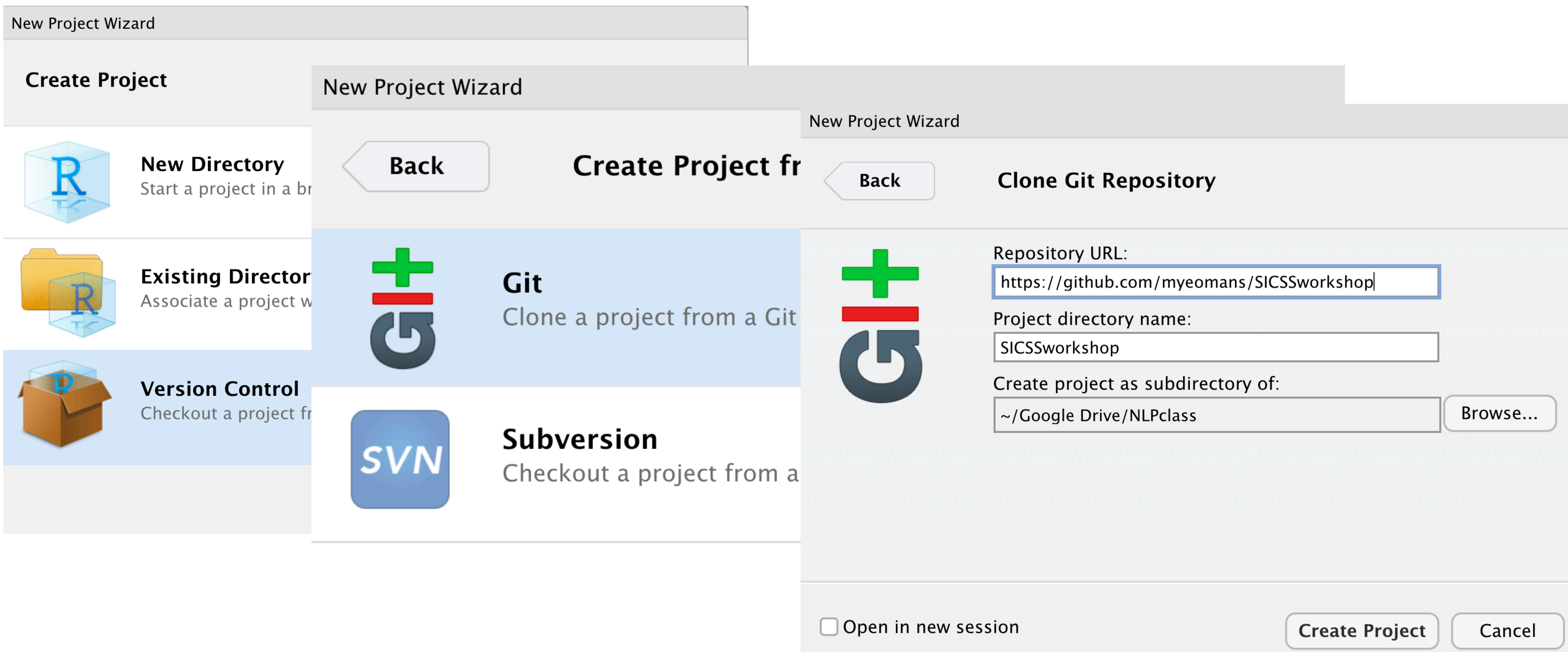
A Quick Pointer to Example Data and Code

If you want to code alongside me (or practice afterwards), you can find it all at:

https://github.com/myeomans/convoResearch

Download everything by **creating an RStudio Project**:



One Question for this Talk

How do we turn words into numbers?

One Question for this Talk

How do we turn words into numbers?

Measurement Validity

(Cronbach & Meehl, 1955; John & Benet- Martinez, 2000;
Flake, Pek & Hehman, 2017; Fried & Flake, 2018)

One Question for this Talk

How do we turn words into numbers?

Measurement Validity

(Cronbach & Meehl, 1955; John & Benet- Martinez, 2000;
Flake, Pek & Hehman, 2017; Fried & Flake, 2018)

Why?

"Prediction in Service of Estimation"

(Mullainathan & Speiss, 2017)

Measurement Validity

Traditionally: scale-based/structured

Measurement Validity

Traditionally: scale-based/structured

How much do you like this product?
1(not at all) to 7 (extremely)

Measurement Validity

<u>Traditionally: scale-based/structured</u>

How much do you like this product?
1(not at all) to 7 (extremely)

Concerns

- are researchers asking the right questions?
- how do people interpret the questions?
- how do people interpret the endpoints?
- how/when do we ask the question?
  Intrusiveness
  Realism

Measurement Validity

More Recently: behavior-based/unstructured

Measurement Validity

<u>More Recently: behavior-based/unstructured</u>

How much do you like this product?
(Open text box)

Measurement Validity

<u>More Recently: behavior-based/unstructured</u>

How much do you like this product?
(Open text box)

Concerns

- will people put effort into their answers?
- how/when do we collect the text?
  Context specificity
  Sample selection bias
- how do we process the data?
- how do we reduce to low dimensions?

Measurement Validity

<u>More Recently: behavior-based/unstructured</u>

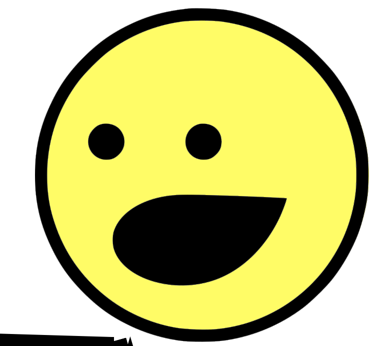
How much do you like this product?
(Open text box)

Concerns

- will people put effort into their answers?
- how/when do we collect the text?
  Context specificity
  Sample selection bias
- how do we process the data?
- **how do we reduce to low dimensions?**

Measurement in Language

Measurement in Language



I think he needs to do a better job of caring about the work environment he is in. If he loses his job which he is on the verge of, he would be in trouble

I am going to give them critical feedback on the work they have done so far. This is a big project and each of us has important work to contribute. I don't feel as if they are doing their best. I want to steer them in the right direction.

I plan to tell him that his failure to politely interact with the customers has resulted in lower sales and is currently putting his job at risk.

I plan on telling her that she needs to work on her hygiene because she hasn't been smelling very good at the office.

from Blunden, Green & Gino, 2018

Measurement in Language

-1.9 — I think he needs to do a better job of caring about the work environment he is in. If he loses his job which he is on the verge of, he would be in trouble

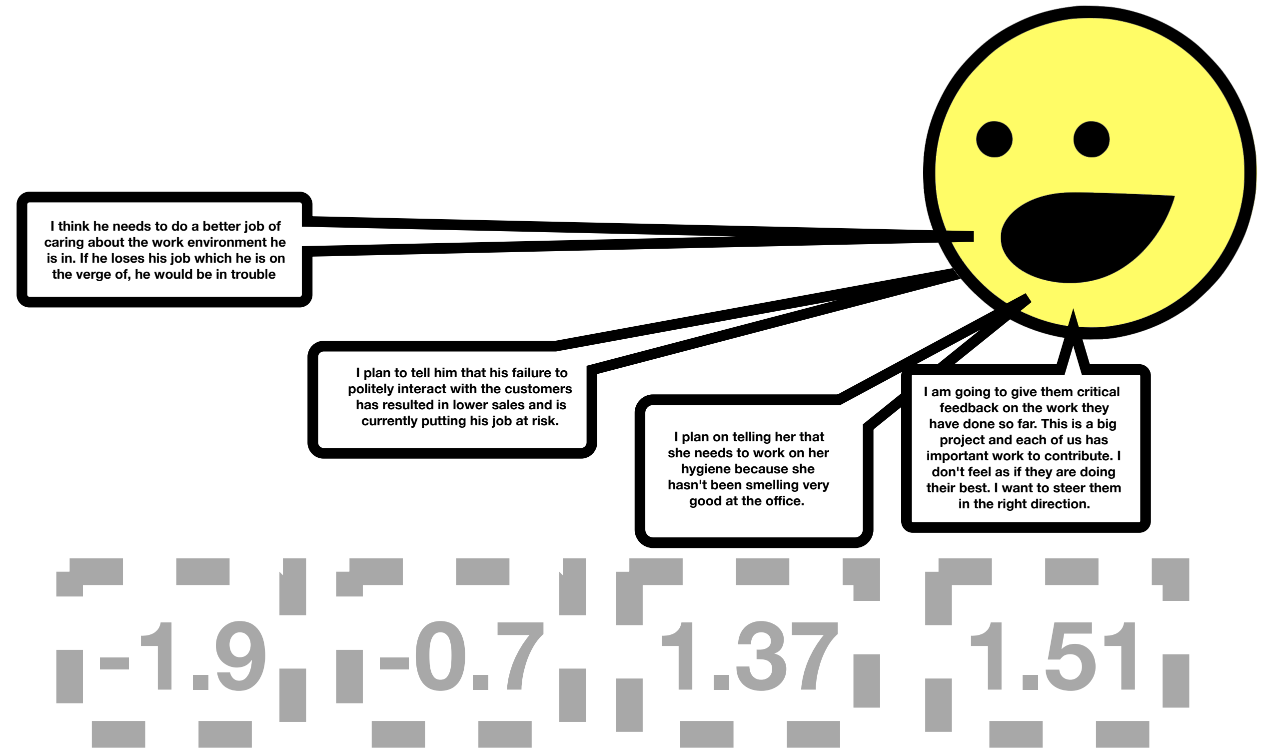
-0.7 — I am going to give them critical feedback on the work they have done so far. This is a big project and each of us has important work to contribute. I don't feel as if they are doing their best. I want to steer them in the right direction.

1.37 — I plan to tell him that his failure to politely interact with the customers has resulted in lower sales and is currently putting his job at risk.

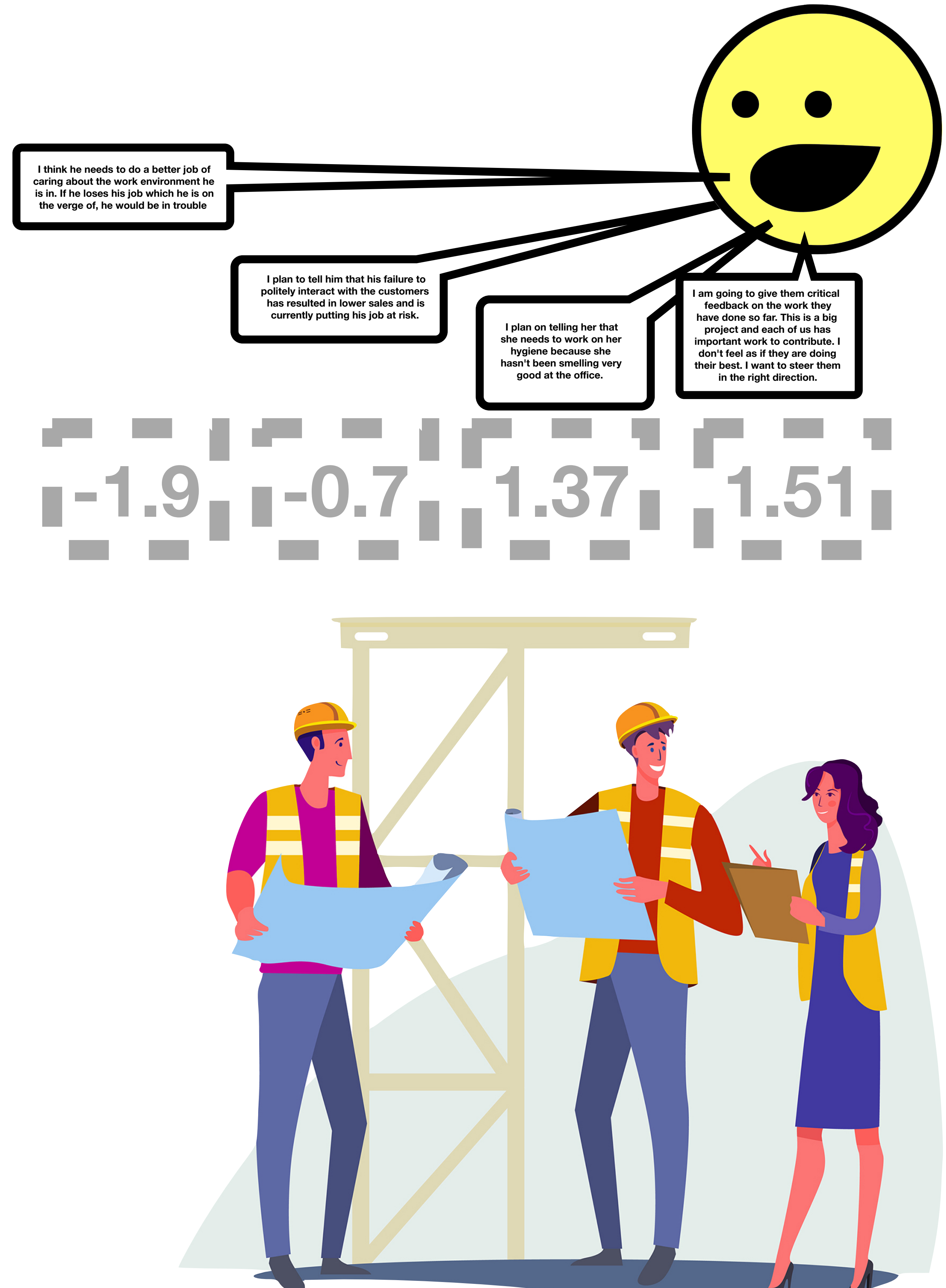
1.51 — I plan on telling her that she needs to work on her hygiene because she hasn't been smelling very good at the office.

from Blunden, Green & Gino, 2018

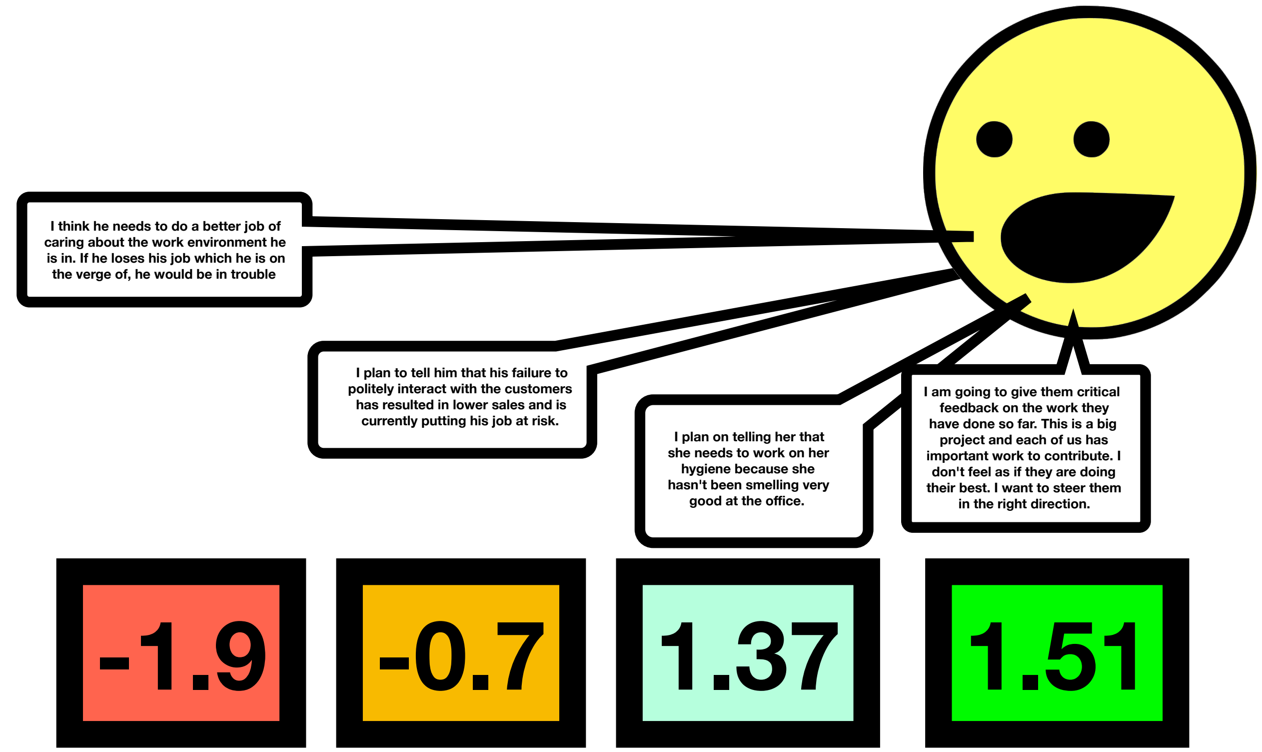
Measurement in Language



Measurement in Language



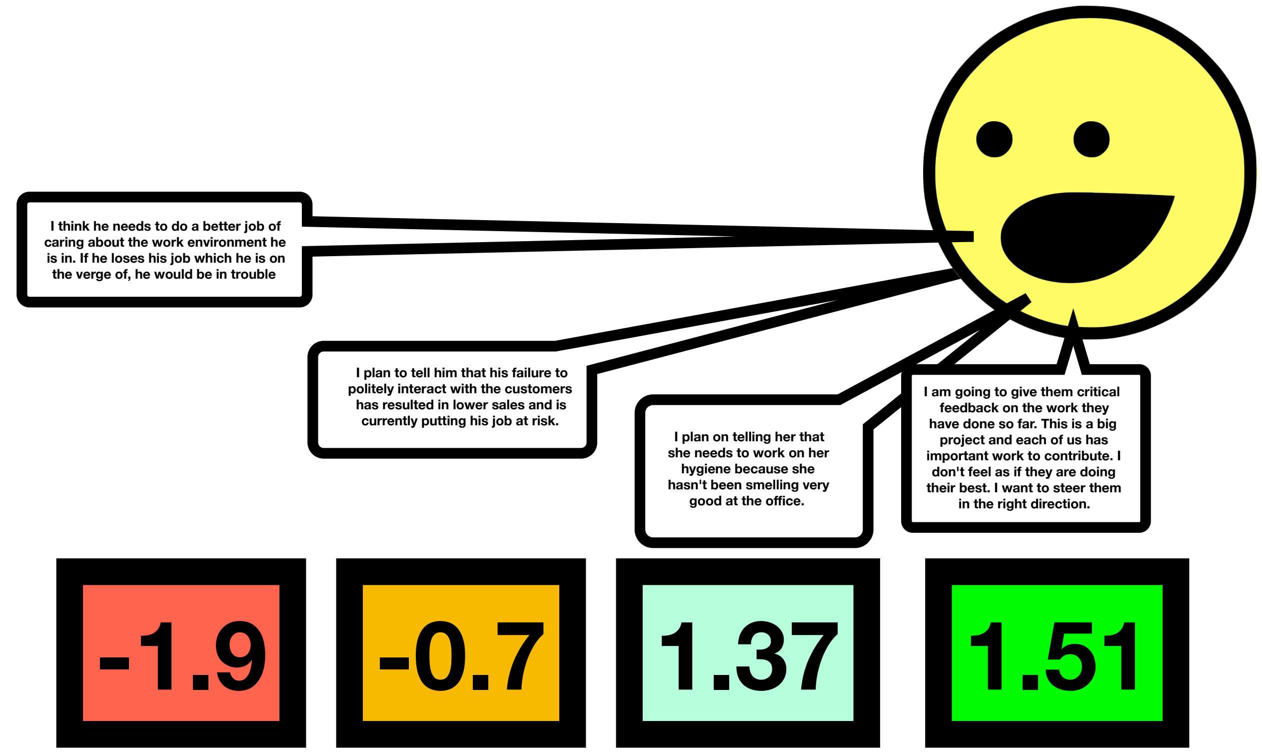
Measurement in Language



Measurement in Language

Humans: Plusses

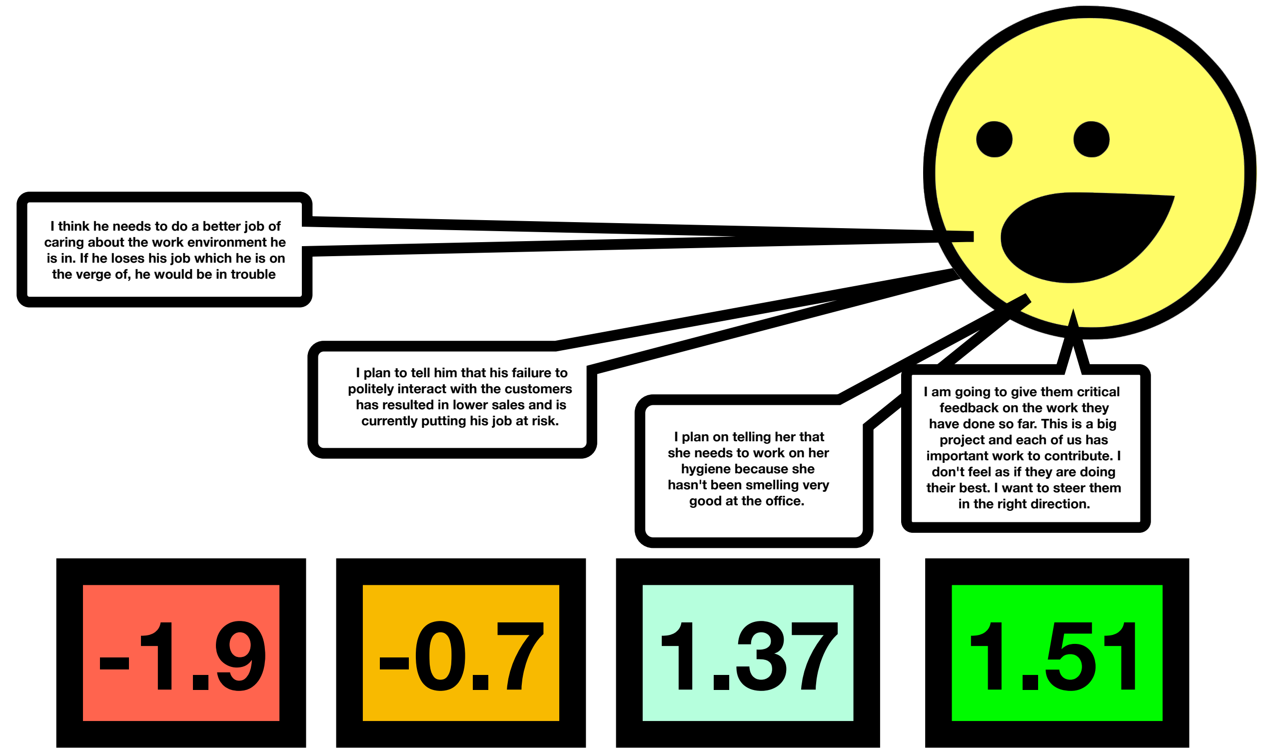
What we've always been doing



Measurement in Language

Humans: Plusses

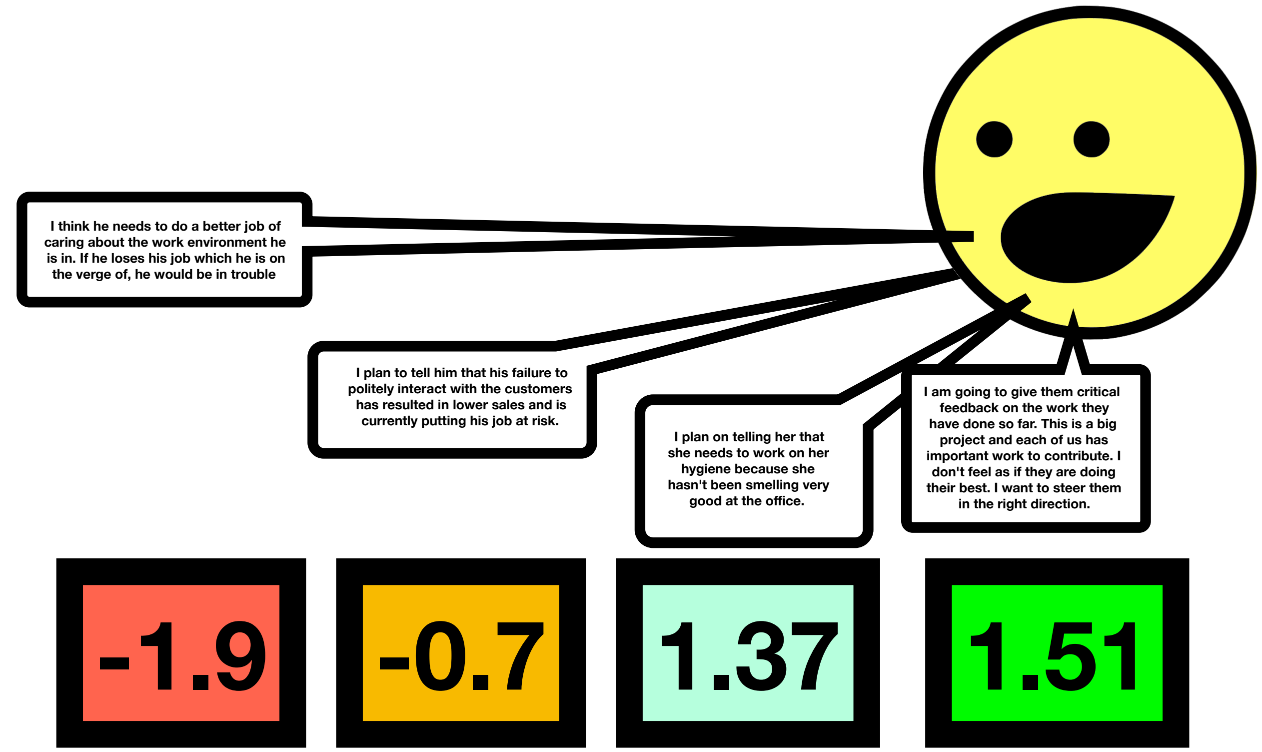
- What we've always been doing
- More accurate than algorithms for complex tasks



Measurement in Language

Humans: Plusses

- What we've always been doing
- More accurate than algorithms
  - for complex tasks
- Can understand context, nuance



Measurement in Language

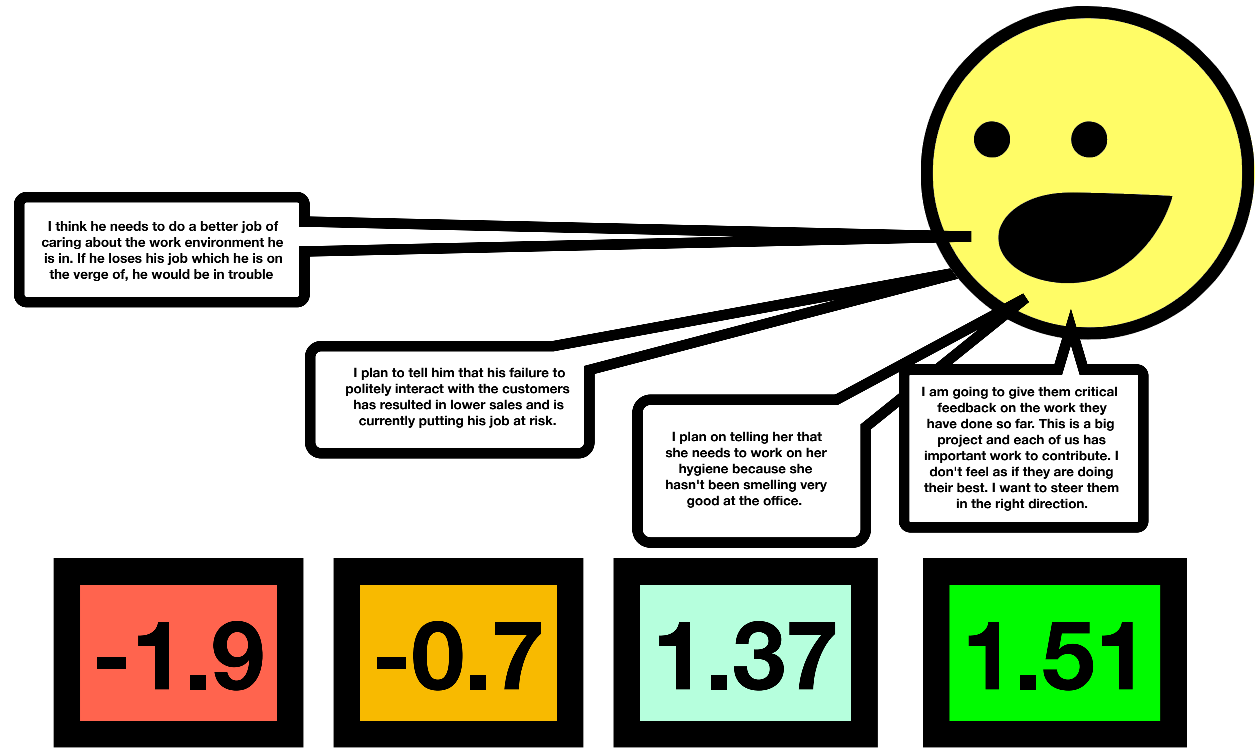
Humans: Plusses

What we've always been doing

More accurate than algorithms

for complex tasks

Can understand context, nuance



Humans: Minuses

High marginal cost of labor

Measurement in Language

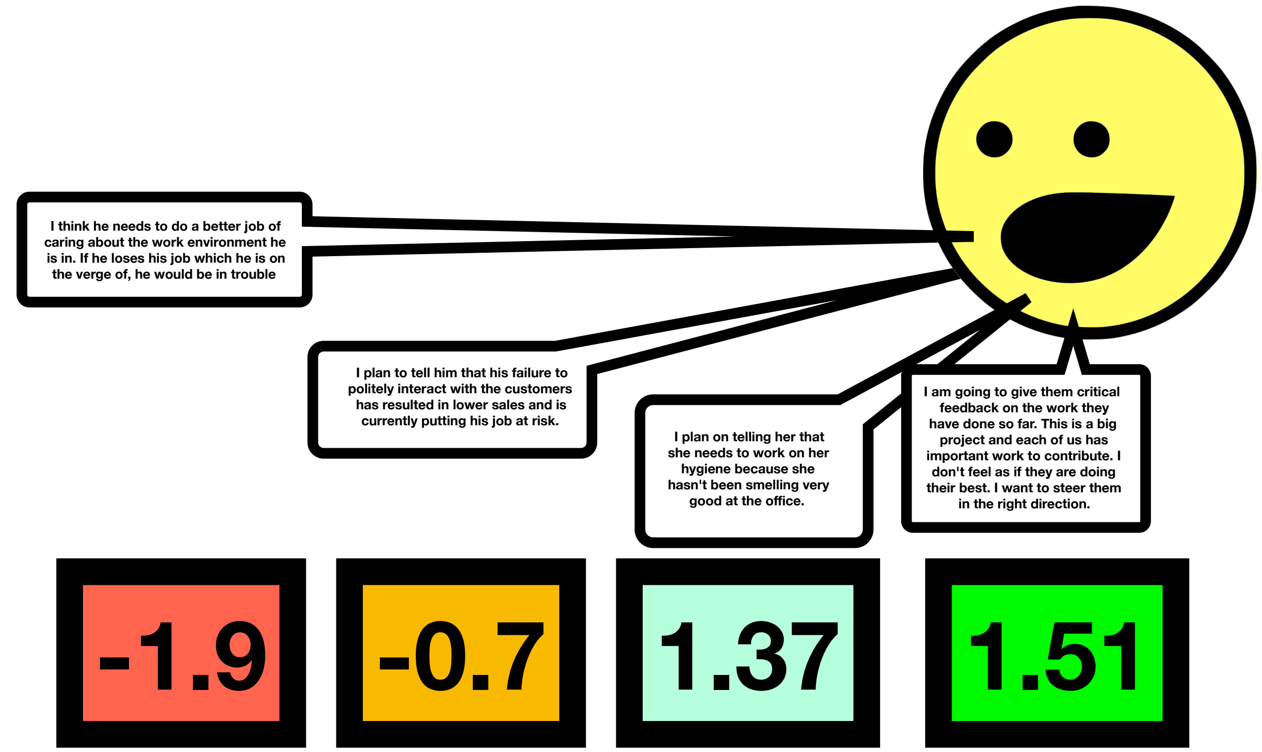
Humans: Plusses

What we've always been doing

More accurate than algorithms

for complex tasks

Can understand context, nuance



Humans: Minuses

High marginal cost of labor

Not reliable

Measurement in Language

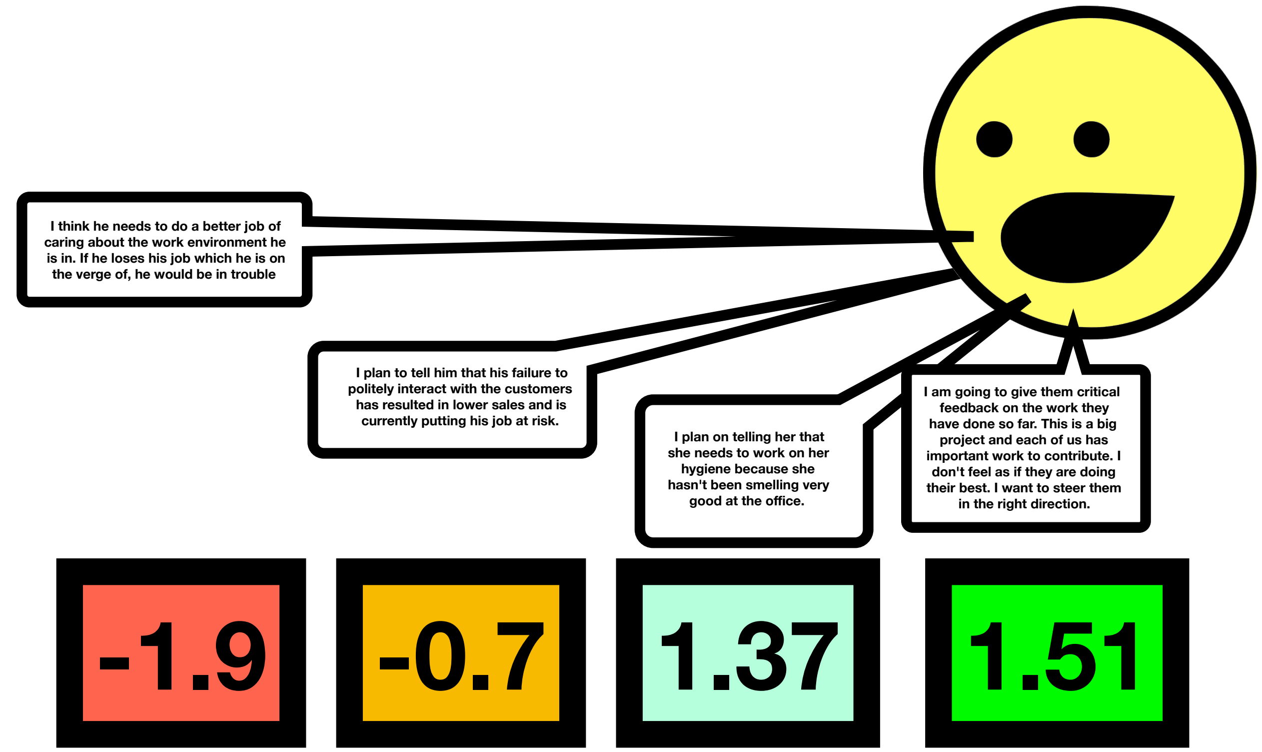
Humans: Plusses

What we've always been doing

More accurate than algorithms

for complex tasks

Can understand context, nuance



Humans: Minuses

High marginal cost of labor

Not reliable

Not transparent

Goals for NLP in Social Science

Goals for NLP in Social Science

Approximate good things about humans

- Validate measures in existing theoretical/empirical framework
- Measure with roughly the same accuracy as trained RAs
- Account for context-specificity

Goals for NLP in Social Science

Approximate good things about humans

Validate measures in existing theoretical/empirical framework

Measure with roughly the same accuracy as trained RAs

Account for context-specificity

Improve on bad things about humans

Decrease marginal cost of measurement in new texts

Increase replicability/reliability of measures

Increase interpretability of resulting models

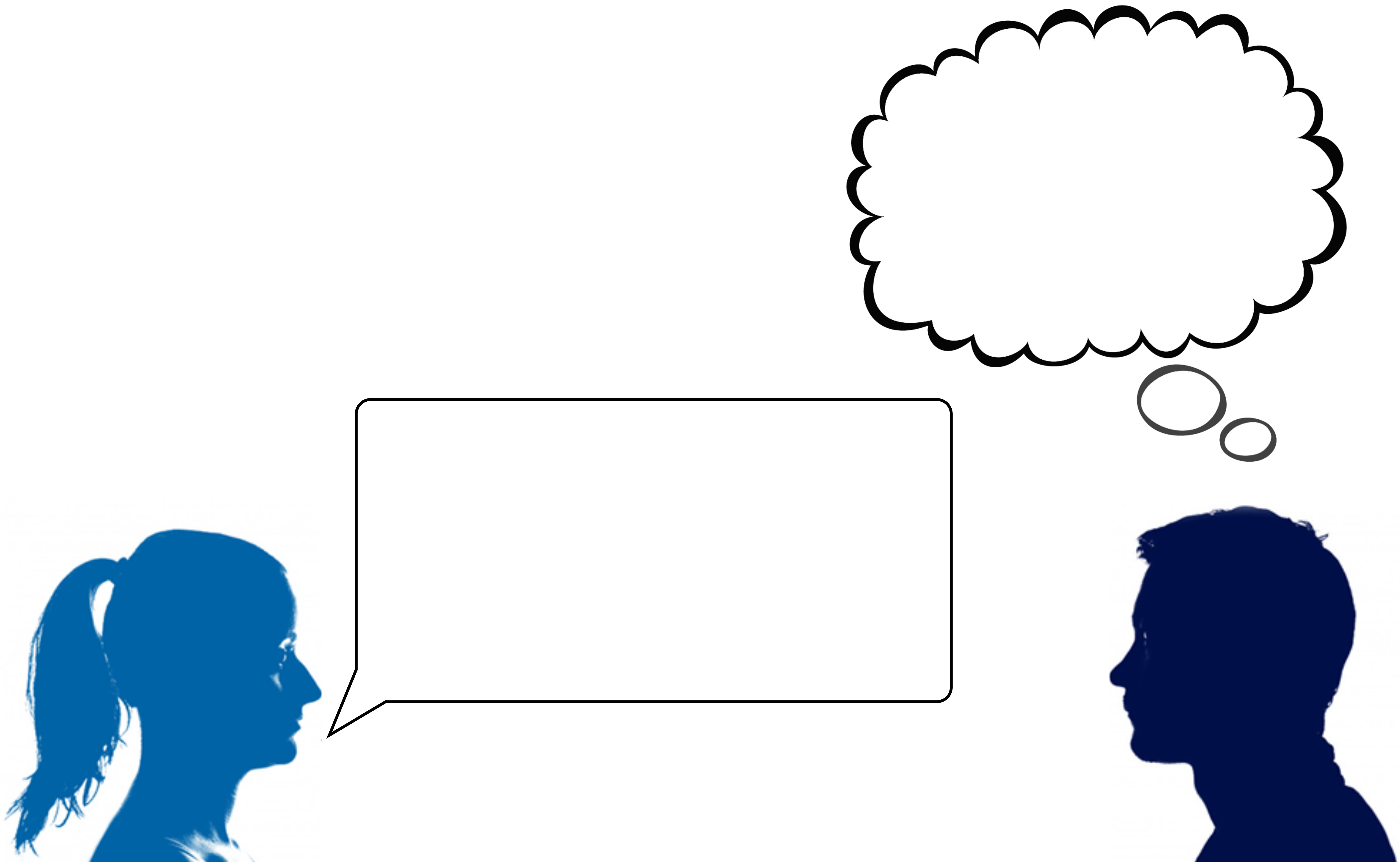
How is an Idea Communicated?

How is an Idea Communicated?

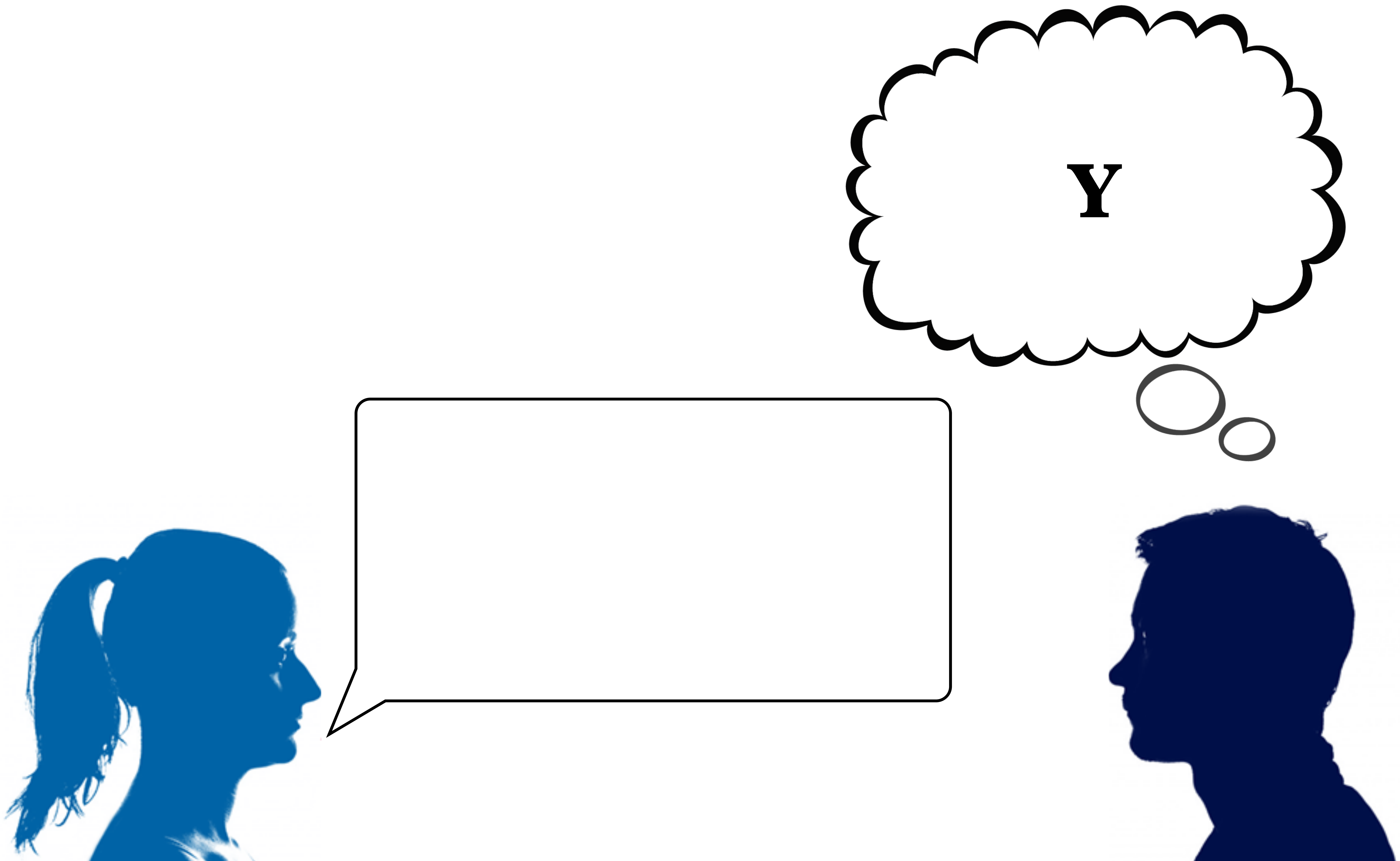
How is an Idea Communicated?

How is an Idea Communicated?

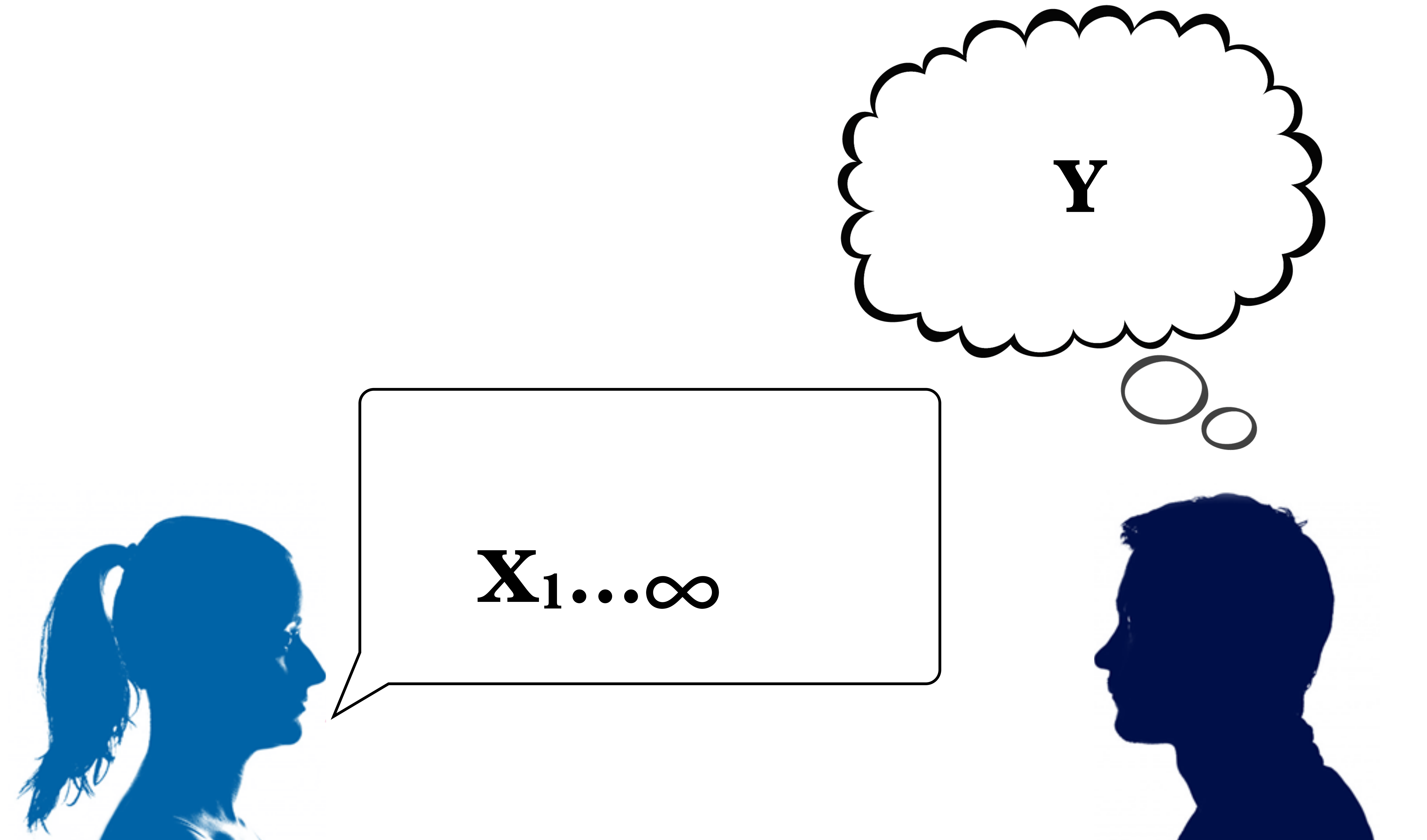
How is an Idea Communicated?



How is an Idea Communicated?



How is an Idea Communicated?



A Linguistic Model

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

A Linguistic Model

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

Feature Extraction

Select set of observables ***x***

Usually theory-based

Sometimes estimated (e.g. unsupervised learning)

A Linguistic Model

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

Feature Extraction

Select set of observables $\boldsymbol{x}$

Usually theory-based

Sometimes estimated (e.g. unsupervised learning)

Feature Estimation

Determine **β** weights

A Linguistic Model

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

Feature Extraction

Select set of observables $\boldsymbol{x}$

Usually theory-based

Sometimes estimated (e.g. unsupervised learning)

Feature Estimation

Determine $\boldsymbol{\beta}$ weights

Usually empirically determined

A Linguistic Model

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

Feature Extraction

Select set of observables ***x***

Usually theory-based

Sometimes estimated (e.g. unsupervised learning)

Feature Estimation

Determine **β** weights

Usually empirically determined

Sometimes guesstimated (e.g. equal weights)

A Linguistic Model

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

Feature Extraction

Select set of observables ***x***

Usually theory-based

Sometimes estimated (e.g. unsupervised learning)

Feature Estimation

Determine **β** weights

Usually empirically determined ← **DO THIS**

Sometimes guesstimated (e.g. equal weights)

Feature Extraction Methods

Feature Extraction Methods

Bag-of-words approaches

Structural approaches

Feature Extraction Methods

Bag-of-words approaches

Ngrams
Dictionaries
Topic Models

Structural approaches

Embedding Models
Politeness
Dialogue Acts

The Bag of Words

I love this movie! It's sweet, but with satirical humor. The dialogue is great and the adventure scenes are fun... It manages to be whimsical and romantic while laughing at the conventions of the fairy tale genre. I would recommend it to just about anyone. I've seen it several times, and I'm always happy to see it again whenever I have a friend who hasn't seen it yet!

it	6
I	5
the	4
to	3
and	3
seen	2
yet	1
would	1
whimsical	1
times	1
sweet	1
satirical	1
adventure	1
genre	1
fairy	1
humor	1
have	1
great	1

Pre-Processing the Bag

Pre-Processing the Bag

Word Stemming

Pre-Processing the Bag

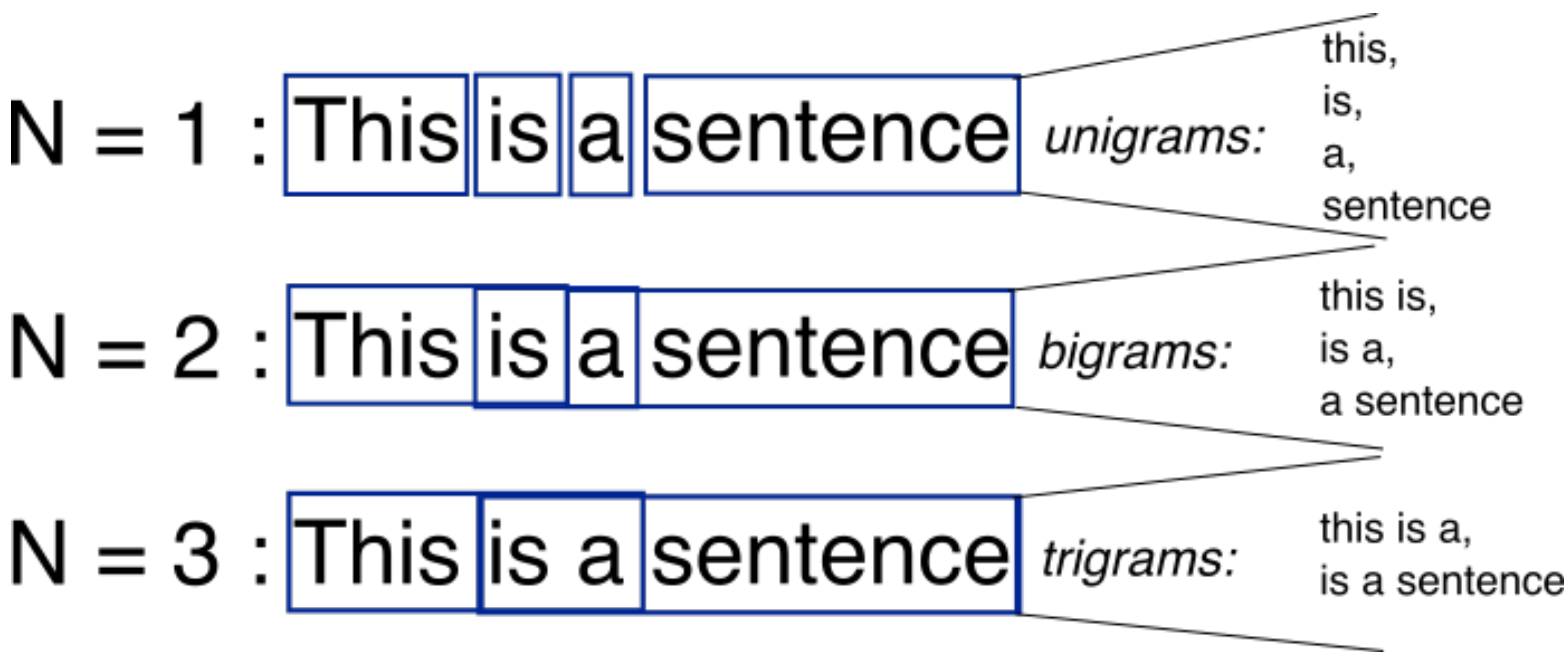
Word Stemming

changing	*stemming* →	chang
changed		chang
change		chang

studying	*stemming* →	studi
studies		studi
study		studi

Pre-Processing the Bag

Constructing phrases - “n-grams”



Pre-Processing the Bag

Dropping “stop words”

i	down	is	should've	she's	more	because	haven	yourself	why	a	didn't
me	in	are	now	her	most	as	haven't	yourselves	how	an	doesn
my	out	was	d	hers	other	until	isn	he	all	the	doesn't
myself	on	were	ll	herself	some	while	isn't	him	any	and	hadn
we	off	be	m	it	such	of	ma	his	both	but	hadn't
our	over	been	o	it's	no	at	mightn	himself	each	if	hasn
ours	under	being	re	its	nor	by	mightn't	she	few	or	hasn't
ourselves	again	have	ve	itself	not	for	mustn	who	s	before	wasn
you	further	has	y	they	only	with	mustn't	whom	t	after	wasn't
you're	then	had	ain	them	own	about	needn	this	can	above	weren
you've	once	having	aren	their	same	against	needn't	that	will	below	weren't
you'll	here	do	aren't	theirs	so	between	shan	that'll	just	to	won
you'd	there	does	couldn	themselves	than	into	shan't	these	don	from	won't
your	when	did	couldn't	what	too	through	shouldn	those	don't	up	wouldn
yours	where	doing	didn	which	very	during	shouldn't	am	should		wouldn't

Pre-Processing the Bag

Dropping “stop words”

i	down	is	should've	she's	more	because	haven	yourself	why	a	didn't
me	in	are	now	her	most	as	haven't	yourselves	how	an	doesn
my	out	was	d	hers	other	until	isn	he	all	the	doesn't
myself	on	were	ll	herself	some	while	isn't	him	any	and	hadn
we	off	be	m	it	such	of	ma	his	both	but	hadn't
our	over	been	o	it's	no	at	mightn	himself	each	if	hasn
ours	under	being	re	its	nor	by	mightn't	she	few	or	hasn't
ourselves	again	have	ve	itself	not	for	mustn	who	s	before	wasn
you	further	has	y	they	only	with	mustn't	whom	t	after	wasn't
you're	then	had	ain	them	own	about	needn	this	can	above	weren
you've	once	having	aren	their	same	against	needn't	that	will	below	weren't
you'll	here	do	aren't	theirs	so	between	shan	that'll	just	to	won
you'd	there	does	couldn	themselves	than	into	shan't	these	don	from	won't
your	when	did	couldn't	what	too	through	shouldn	those	don't	up	wouldn
yours	where	doing	didn	which	very	during	shouldn't	am	should		wouldn't

Pre-Processing the Bag

Dropping “stop words”

i	down	is	should've	she's	more	because	haven	yourself	why	a	didn't
me	in	are	now	her	most	as	haven't	yourselves	how	an	doesn
my	out	was	d	hers	other	until	isn	he	all	the	doesn't
myself	on	were	ll	herself	some	while	isn't	him	any	and	hadn
we	off	be	m	it	such	of	ma	his	both	but	hadn't
our	over	been	o	it's	no	at	mightn	himself	each	if	hasn
ours	under	being	re	its	nor	by	mightn't	she	few	or	hasn't
ourselves	again	have	ve	itself	not	for	mustn	who	s	before	wasn
you	further	has	y	they	only	with	mustn't	whom	t	after	wasn't
you're	then	had	ain	them	own	about	needn	this	can	above	weren
you've	once	having	aren	their	same	against	needn't	that	will	below	weren't
you'll	here	do	aren't	theirs	so	between	shan	that'll	just	to	won
you'd	there	does	couldn	themselves	than	into	shan't	these	don	from	won't
your	when	did	couldn't	what	too	through	shouldn	those	don't	up	wouldn
yours	where	doing	didn	which	very	during	shouldn't	am	should		wouldn't

Pre-Processing the Bag

Dropping "stop words"

i	down	is	should've	she's	more	because	haven	yourself	why	a	didn't
me	in	are	now	her	most	as	haven't	yourselves	how	an	doesn
my	out	was	d	hers	other	until	isn	he	all	the	doesn't
myself	on	were	ll	herself	some	while	isn't	him	any	and	hadn
we	off	be	m	it	such	of	ma	his	both	but	hadn't
our	over	been	o	it's	no	at	mightn	himself	each	if	hasn
ours	under	being	re	its	nor	by	mightn't	she	few	or	hasn't
ourselves	again	have	ve	itself	not	for	mustn	who	s	before	wasn
you	further	has	y	they	only	with	mustn't	whom	t	after	wasn't
you're	then	had	ain	them	own	about	needn	this	can	above	weren
you've	once	having	aren	their	same	against	needn't	that	will	below	weren't
you'll	here	do	aren't	theirs	so	between	shan	that'll	just	to	won
you'd	there	does	couldn	themselves	than	into	shan't	these	don	from	won't
your	when	did	couldn't	what	too	through	shouldn	those	don't	up	wouldn
yours	where	doing	didn	which	very	during	shouldn't	am	should		wouldn't

Dictionaries

Dictionaries

Every word has its own score (from raters)

Dictionaries

Every word has its own score (from raters)

	Word	Score	Word	Scor
Valence	love	1.000	toxic	0.008
	happy	1.000	nightmare	0.005
Arousal	elated	0.960	mellow	0.069
	frenzy	0.965	napping	0.046
Dominanc e	powerful	0.991	weak	0.045
	leadership	0.983	empty	0.081

(Osgood et al., 1957; Mohammad, 2018)

Dictionaries

Every word has its own score (from raters)

	Word	Score	Word	Scor
Valence	love	1.000	toxic	0.008
	happy	1.000	nightmare	0.005
Arousal	elated	0.960	mellow	0.069
	frenzy	0.965	napping	0.046
Dominanc e	powerful	0.991	weak	0.045
	leadership	0.983	empty	0.081

(Osgood et al., 1957; Mohammad, 2018)

Every word in a category has the same score

Dictionaries

Every word has its own score (from raters)

	Word	Score	Word	Scor
Valence	love	1.000	toxic	0.008
	happy	1.000	nightmare	0.005
Arousal	elated	0.960	mellow	0.069
	frenzy	0.965	napping	0.046
Dominance	powerful	0.991	weak	0.045
	leadership	0.983	empty	0.081

(Osgood et al., 1957; Mohammad, 2018)

Every word in a category has the same score

LIWC

(Tausczik & Pennebaker, 2010)

“a+”	"abound"	"abounds"	"achievement"
"abundance"	"abundant"	"accessable"	“acumen"
"accessible"	"acclaim"	"acclaimed"	"adequate"
"acclamation"	"accolade"	"accolades"	"achievements
"accommodative"	"accomodative'	"accomplish"	"adaptable"
"accomplished"	“accomplishmen	ccomplishment	"adjustable"

Dictionaries

Dictionaries

Advantages

Easy to implement

Dictionaries

Advantages

Easy to implement

Some dictionaries have extensive validation

Dictionaries

Advantages

Easy to implement

Some dictionaries have extensive validation

Limitations

No use of word order   e.g. “bad” vs. “not bad”

Dictionaries

Advantages

Easy to implement

Some dictionaries have extensive validation

Limitations

No use of word order   e.g. “bad” vs. “not bad”

Easy to create -> many dictionaries have no validation

Dictionaries

Advantages

Easy to implement

Some dictionaries have extensive validation

Limitations

No use of word order   e.g. “bad” vs. “not bad”

Easy to create -> many dictionaries have no validation

Hard to interpret result e.g. “authenticity” dictionary

Dictionaries

Advantages

Easy to implement

Some dictionaries have extensive validation

Limitations

No use of word order e.g. “bad” vs. “not bad”

Easy to create -> many dictionaries have no validation

Hard to interpret result e.g. “authenticity” dictionary

Assumes context universality

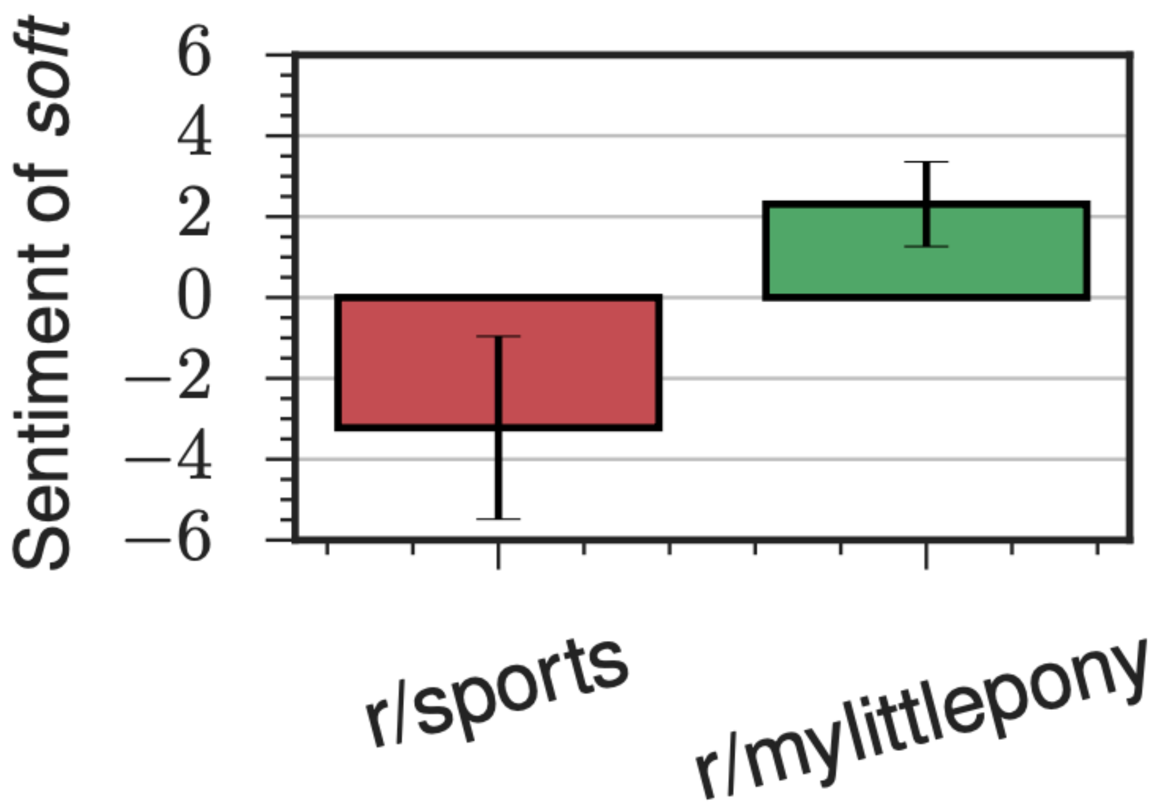
Dictionaries

Context Specificity

"Inducing Domain-Specific Sentiment Lexicons from Unlabeled Corpora"
(Hamilton et al., 2016)

Dictionaries

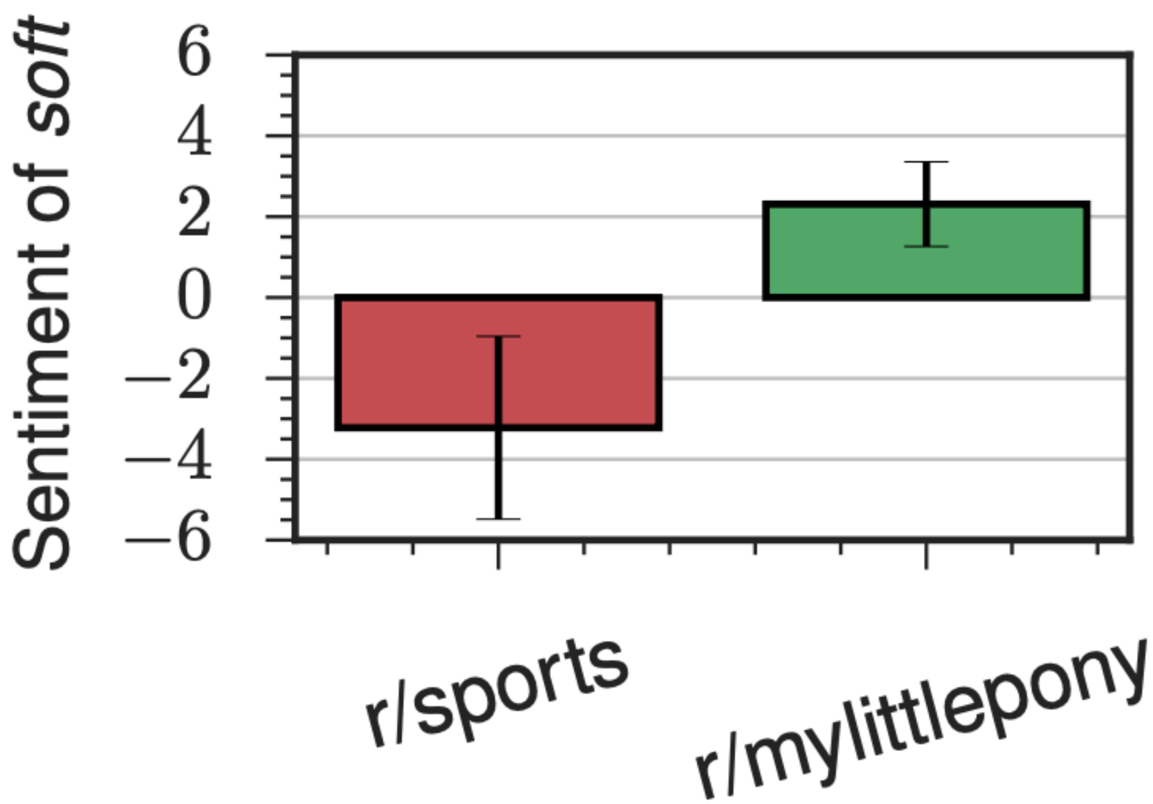
Context Specificity

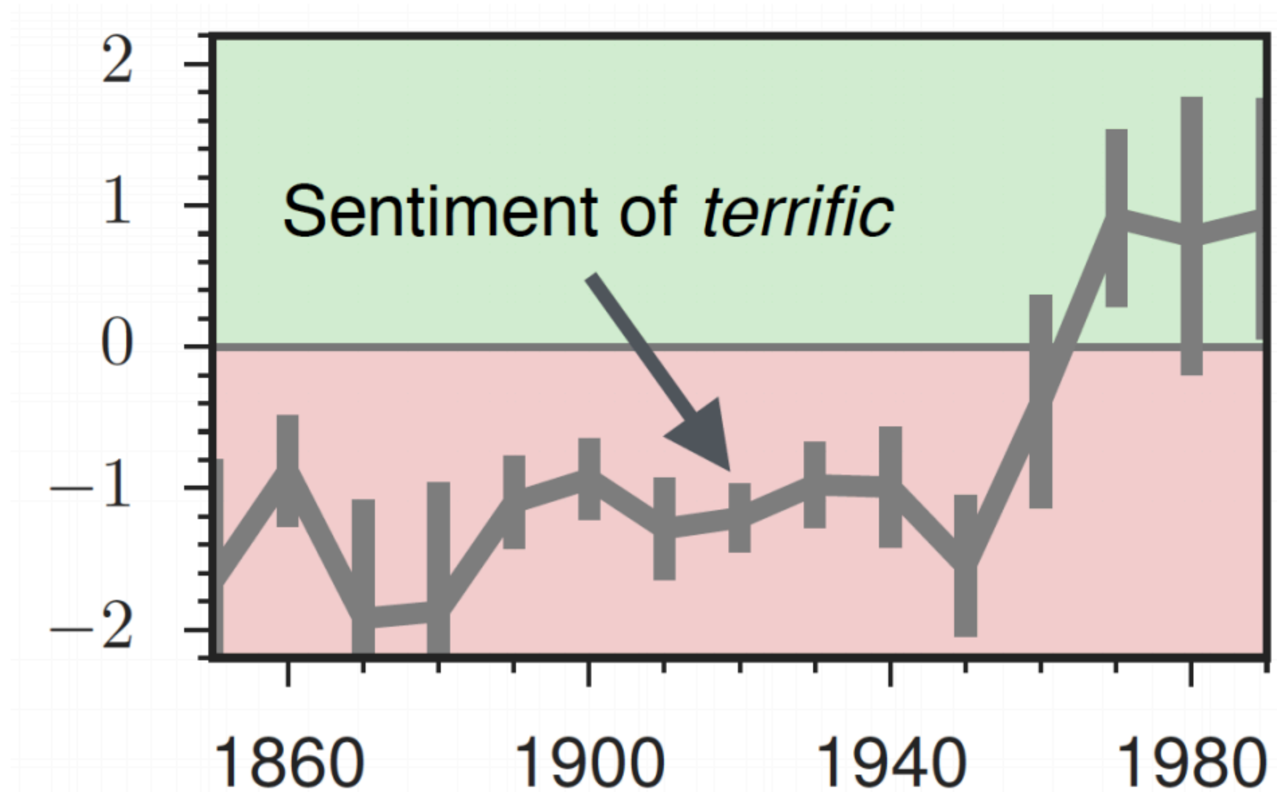


"Inducing Domain-Specific Sentiment Lexicons from Unlabeled Corpora"
(Hamilton et al., 2016)

Dictionaries

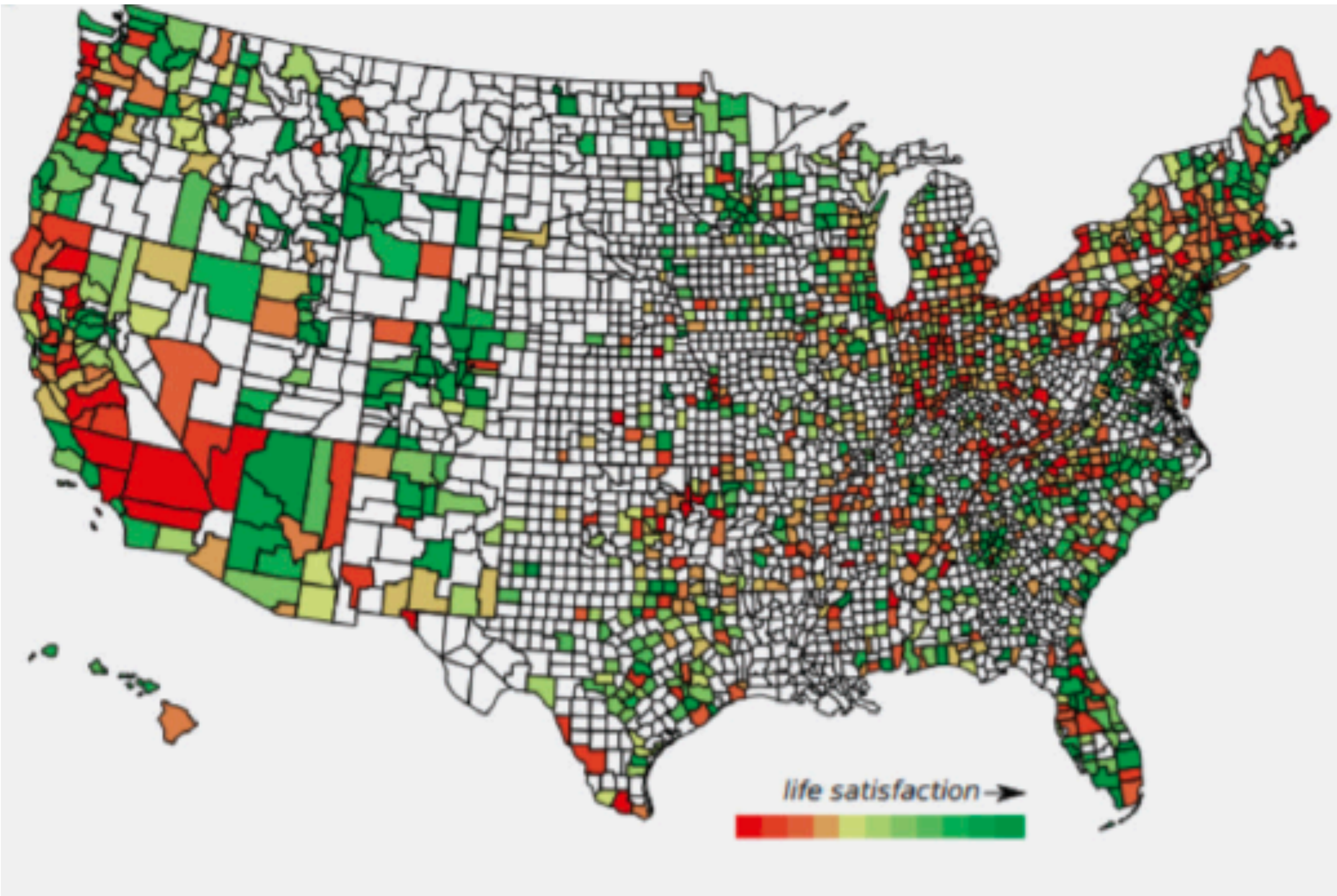
Context Specificity





"Inducing Domain-Specific Sentiment Lexicons from Unlabeled Corpora"
(Hamilton et al., 2016)

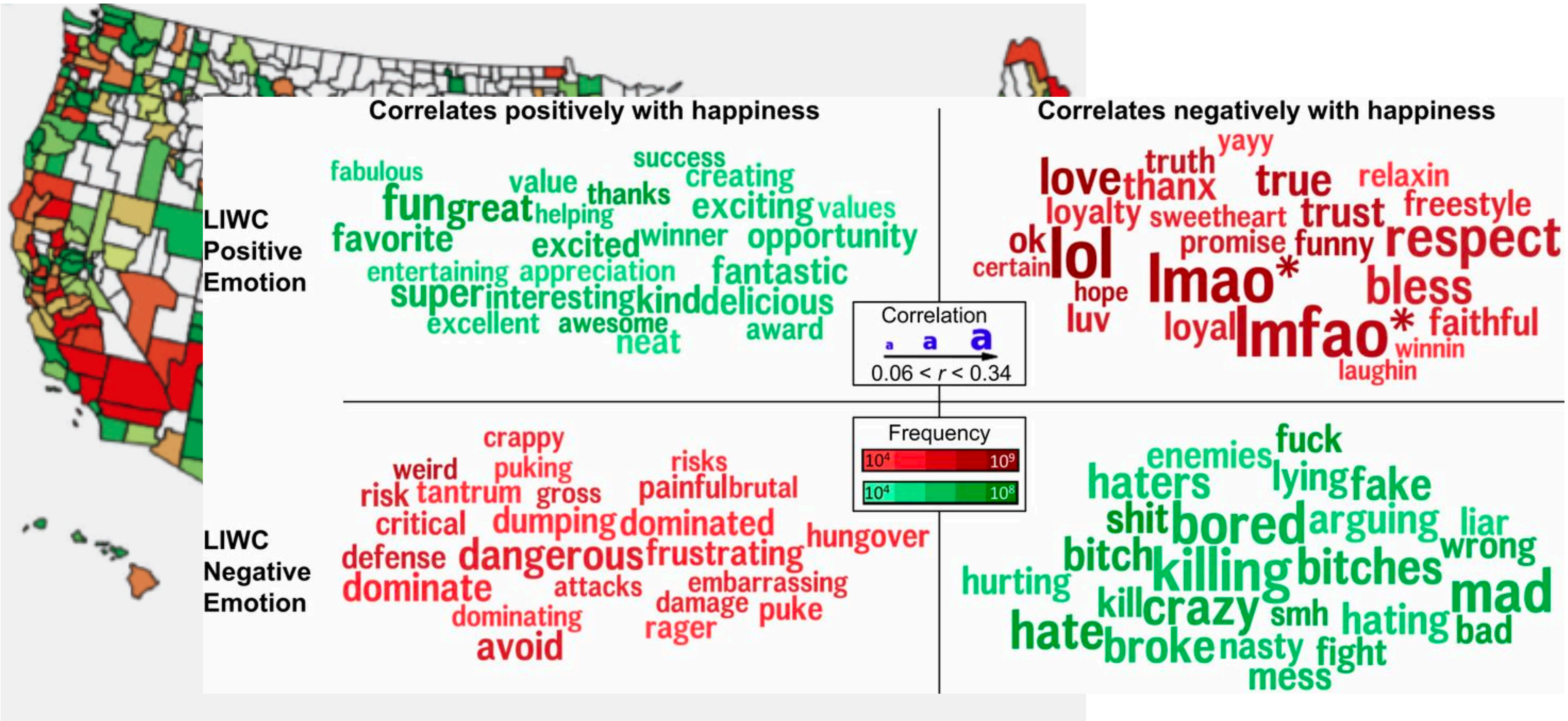
Dictionaries



"Estimating geographic subjective well-being from Twitter"

(Jaidka et al., 2020)

Dictionaries

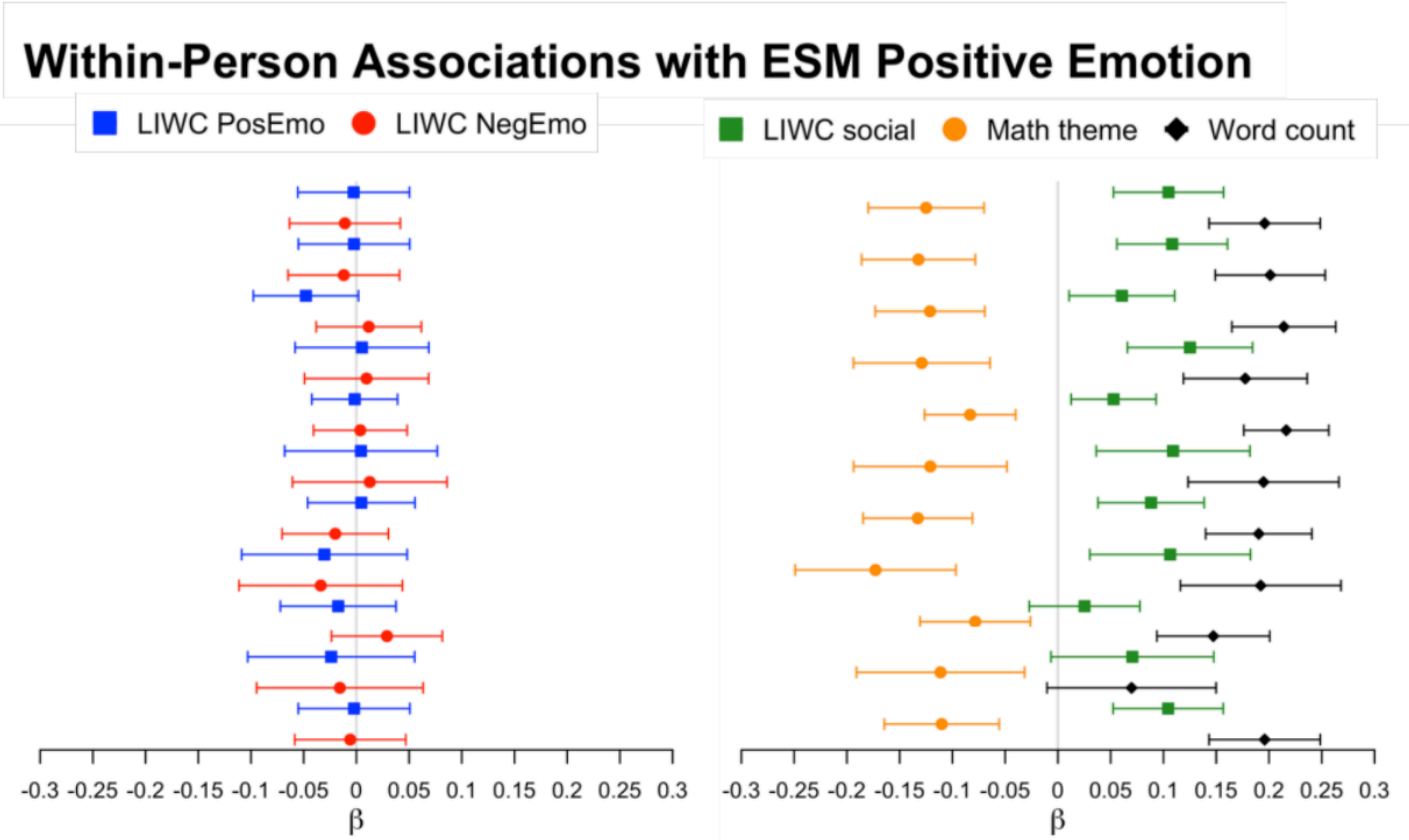


“Estimating geographic subjective well-being from Twitter”

(Jaidka et al., 2020)

Dictionaries

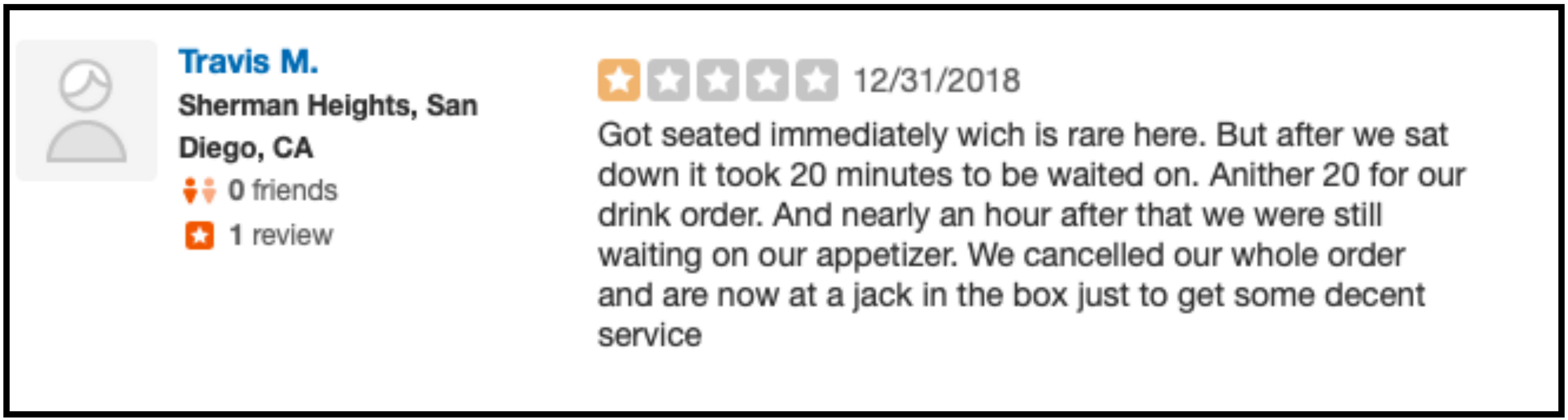
Specification	N_{people}	N_{obs}
Final analysis	185	1,579
No time covariate	185	1,579
Min. 3 words per file	194	1,734
Min. 10 words per file	159	1,255
Min. 1 file per 3-hour block	230	2,586
Min. 5 files per 3-hour block	116	844
Min. 3 observations	231	1,740
Min. 10 observations	58	723
Target hour (Min. 1 file)	187	1,601
Target hour (Min. 2 files)	106	757
Weighted by word count	185	1,579



“Emotion Fluctuations and Everyday Speech”
(Sun et al., 2019)

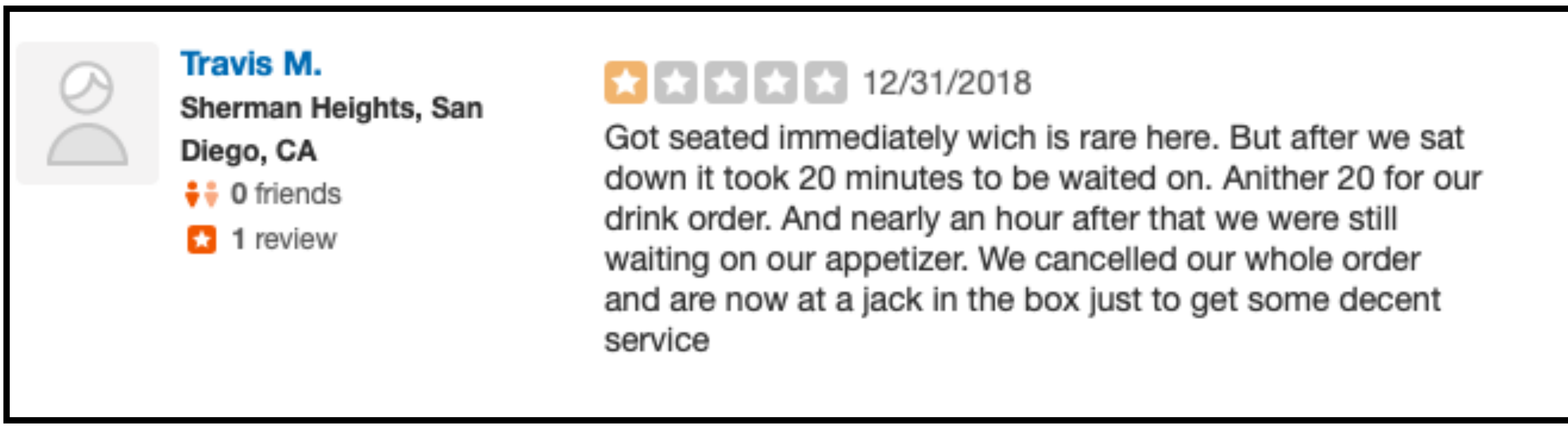
Dictionaries

Best case scenario: descriptions of things



Dictionaries

Best case scenario: descriptions of things



Even then… a domain-specific model will outperform

Supervised Machine Learning (briefly)

Supervised Machine Learning (briefly)

Extract (too) many features

All 1,2,3-word phrases (incl. stop words)

Supervised Machine Learning (briefly)

Extract (too) many features

All 1,2,3-word phrases (incl. stop words)

Build a Model to Predict Human Labels

- key insight: "regularization"
- slightly fancier than a multiple regression

(see Tibshirani, 1996; Friedman, Hastie & Tibshirani, 2010)

Supervised Machine Learning (briefly)

Extract (too) many features

All 1,2,3-word phrases (incl. stop words)

Build a Model to Predict Human Labels

- key insight: "regularization"
- slightly fancier than a multiple regression

(see Tibshirani, 1996; Friedman, Hastie & Tibshirani, 2010)

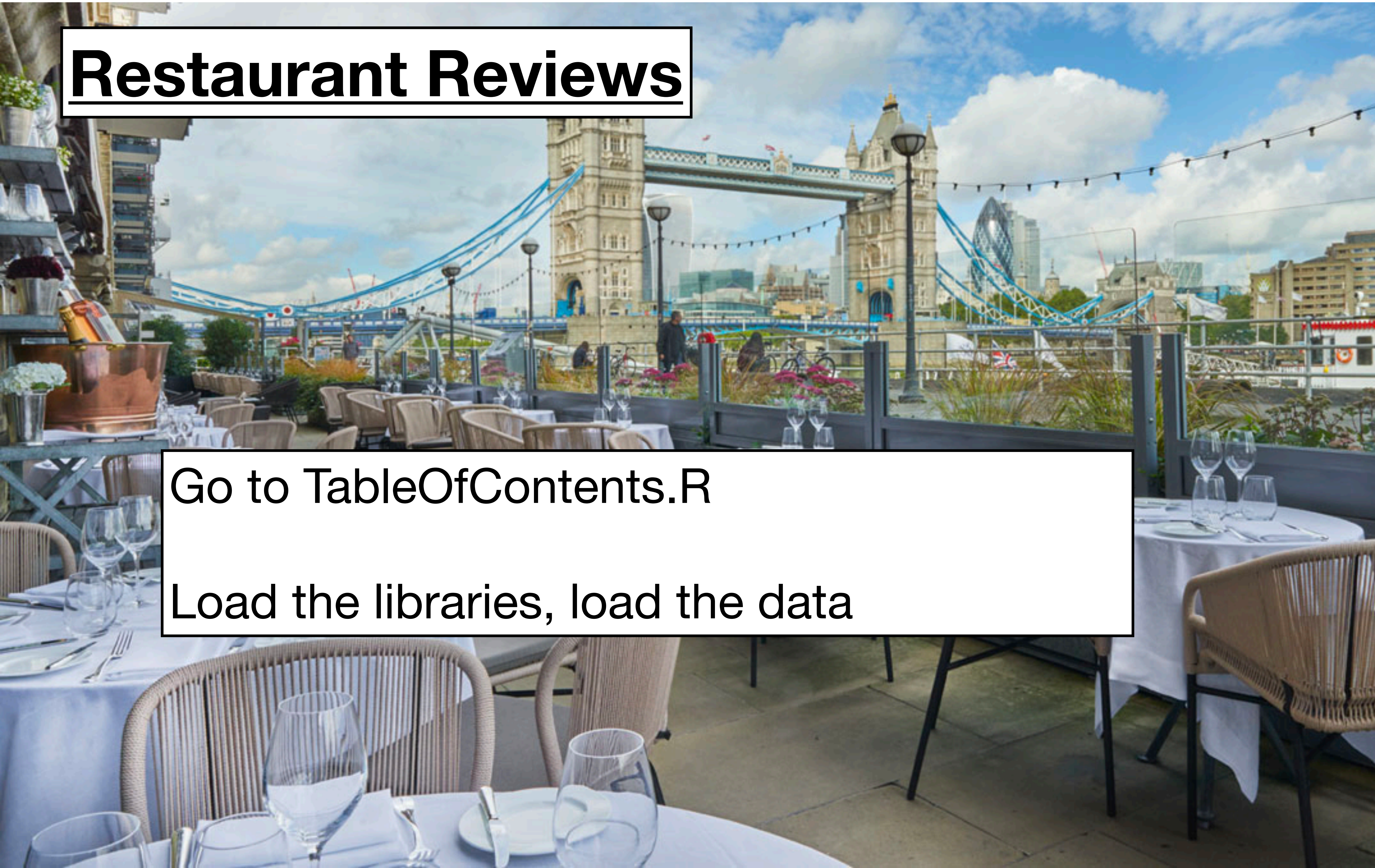
Evaluate Model Accuracy

- key insight: "out of sample prediction"
- nested cross-validation

(see Stone, 1974; Varma & Simon, 2006)

Application in Your Code

Application in Your Code

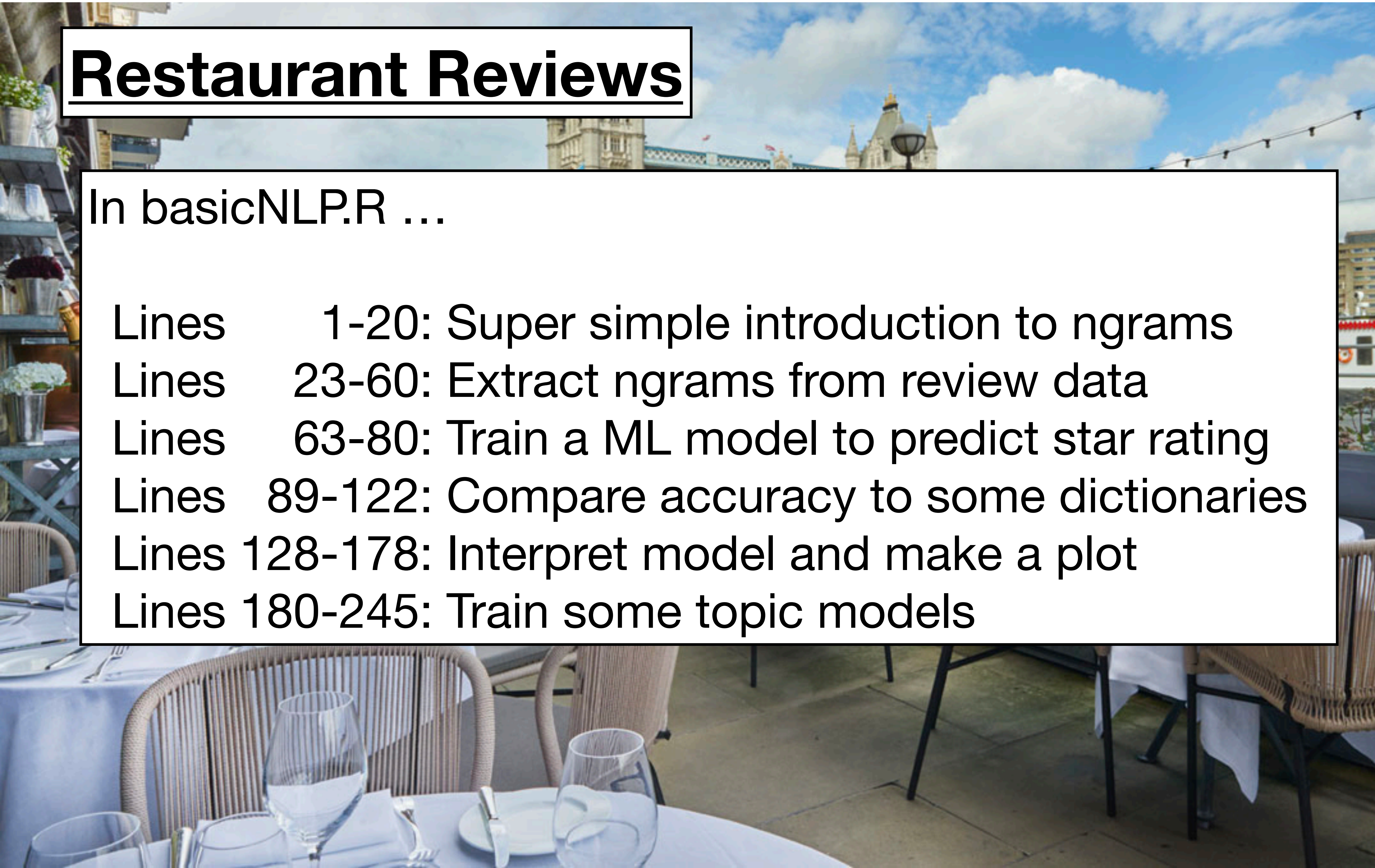


Restaurant Reviews

Go to TableOfContents.R

Load the libraries, load the data

Application in Your Code



Restaurant Reviews

In basicNLP.R …

- Lines 1-20: Super simple introduction to ngrams
- Lines 23-60: Extract ngrams from review data
- Lines 63-80: Train a ML model to predict star rating
- Lines 89-122: Compare accuracy to some dictionaries
- Lines 128-178: Interpret model and make a plot
- Lines 180-245: Train some topic models

Application in Your Code

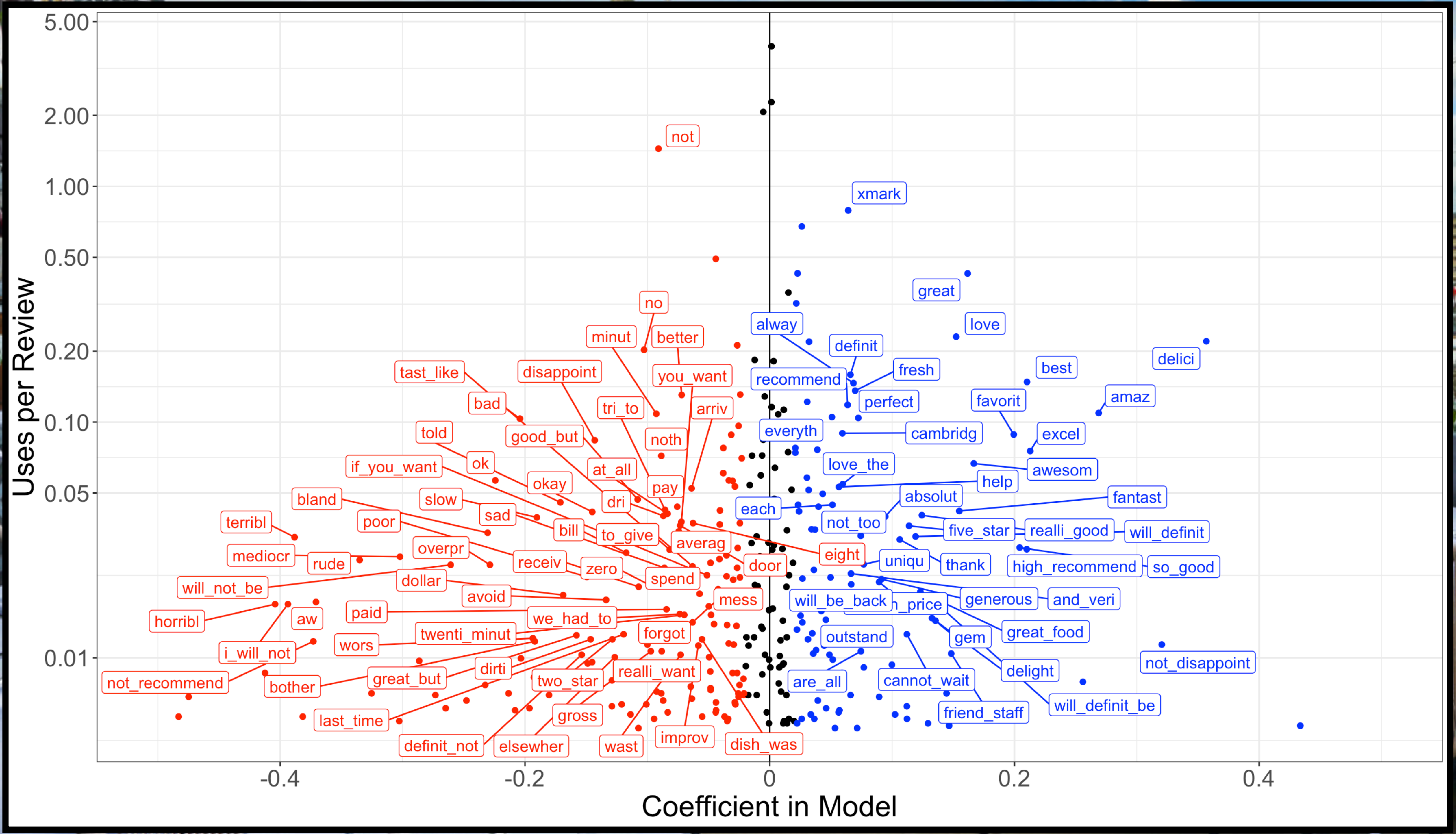
<u>Building a sentiment model</u>

Domain-general mTurk-annotated dictionary: r = 0.31

Model trained on movie reviews: r = 0.50

Model trained in our data: r = 0.64

Application in Your Code



A Concrete Mega-Analysis

12 measures of concreteness
17 datasets across four domains
9,780 documents

(Yeomans, 2021)

A Concrete Mega-Analysis

12 measures of concreteness
17 datasets across four domains
9,780 documents

Zero = < .03	Medium = .2 - .4
Very Low = .03 - .1	High = .4 - .6
Low = .1 - .2	Very High = >.6

Word-Level

Categorical

(Yeomans, 2021)

Name of Measure	Measurement Validity			
	Advice	**Plan Distance**	**Plan Specificity**	**Describing**
mTurk Ratings				
Original MRC				
Bootstrap MRC				
Immediacy				
Larrimore- LIWC				
Pan-LIWC				
Part-of- Speech LCM				
Syntax LCM				
DICTION				

A Concrete Mega-Analysis

12 measures of concreteness
17 datasets across four domains
9,780 documents

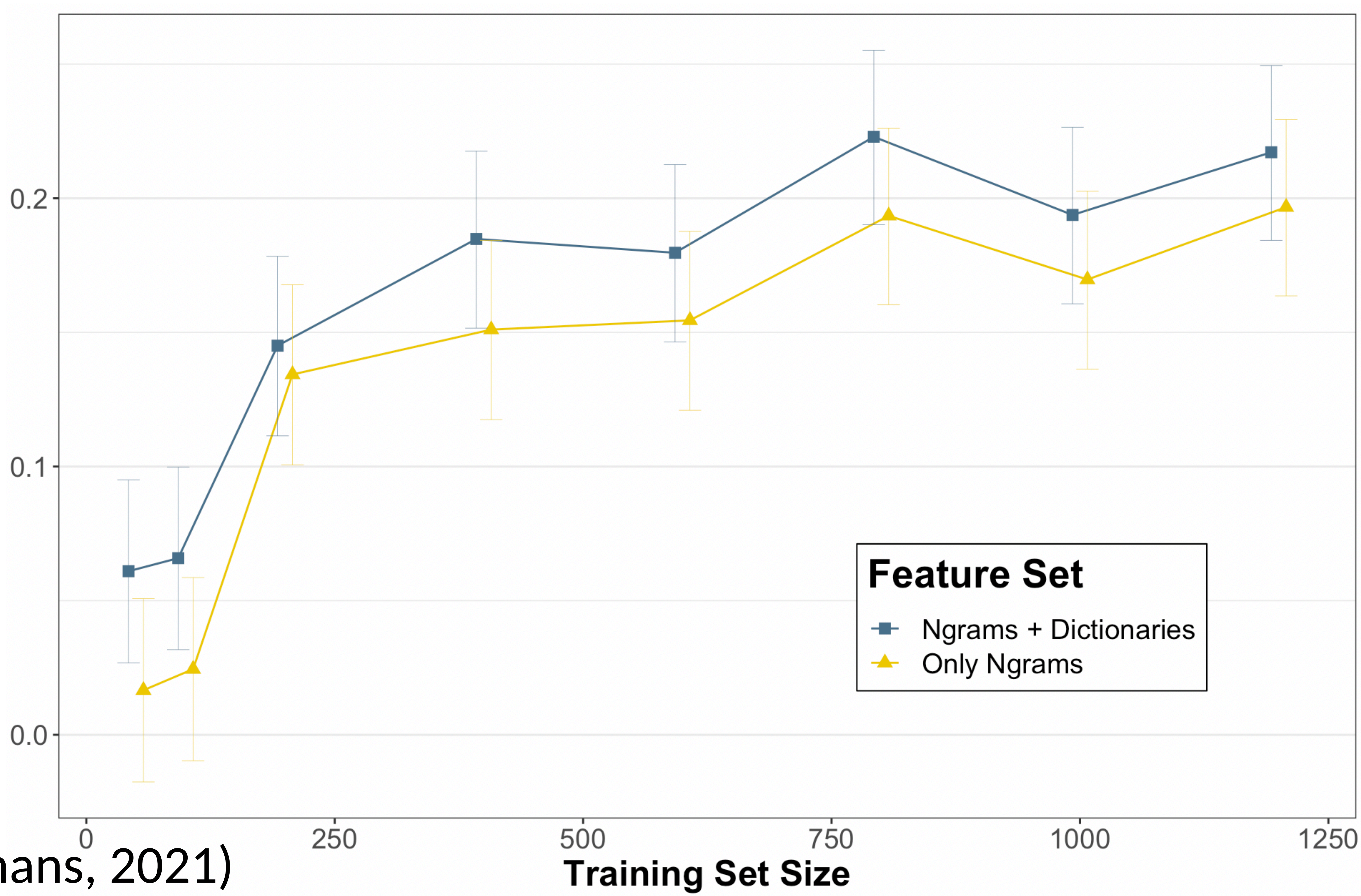
Zero = < .03	Medium = .2 - .4
Very Low = .03 - .1	High = .4 - .6
Low = .1 - .2	Very High = >.6

	Name of Measure	Measurement Validity			
		Advice	Plan Distance	Plan Specificity	Describing
Word-Level	mTurk Ratings	Low	Low	Low	Low
	Original MRC	Low	Low	Very Low	Medium
	Bootstrap MRC	Low	Low	Low	Low
Categorical	Immediacy	Zero	Very Low	Zero	Medium
	Larrimore- LIWC	Very Low	Very Low	Very Low	Zero
	Pan-LIWC	Zero	Very Low	Very Low	Zero
	Part-of- Speech LCM	Zero	Very Low	Zero	Medium
	Syntax LCM	Zero	Zero	Very Low	Low
	DICTION	Very Low	Zero	Zero	Very Low

(Yeomans, 2021)

How Many Annotations?

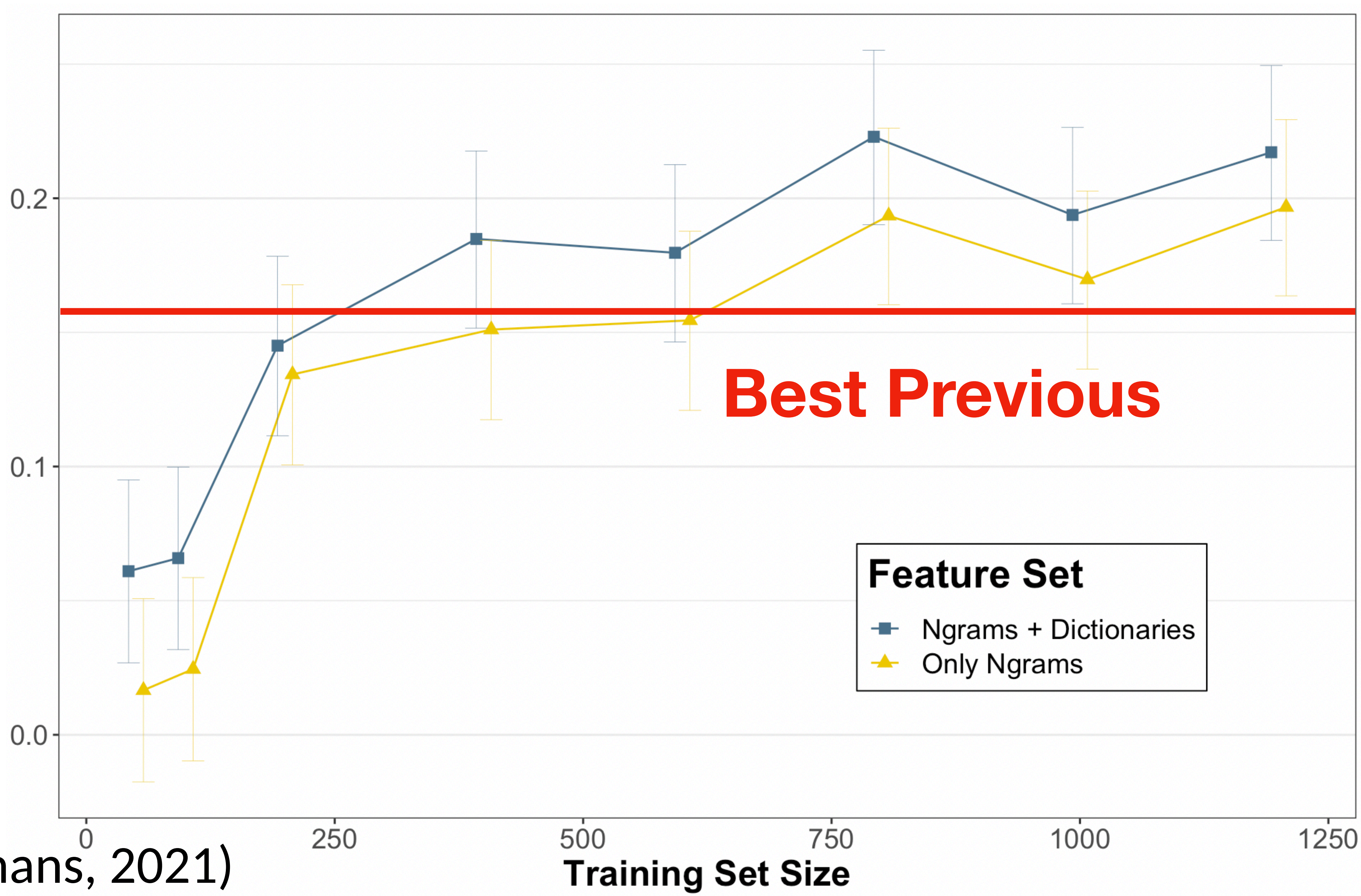
Advice Specificity



(Yeomans, 2021)

How Many Annotations?

Advice Specificity



(Yeomans, 2021)

How Many Annotations?

(Yeomans, 2021)

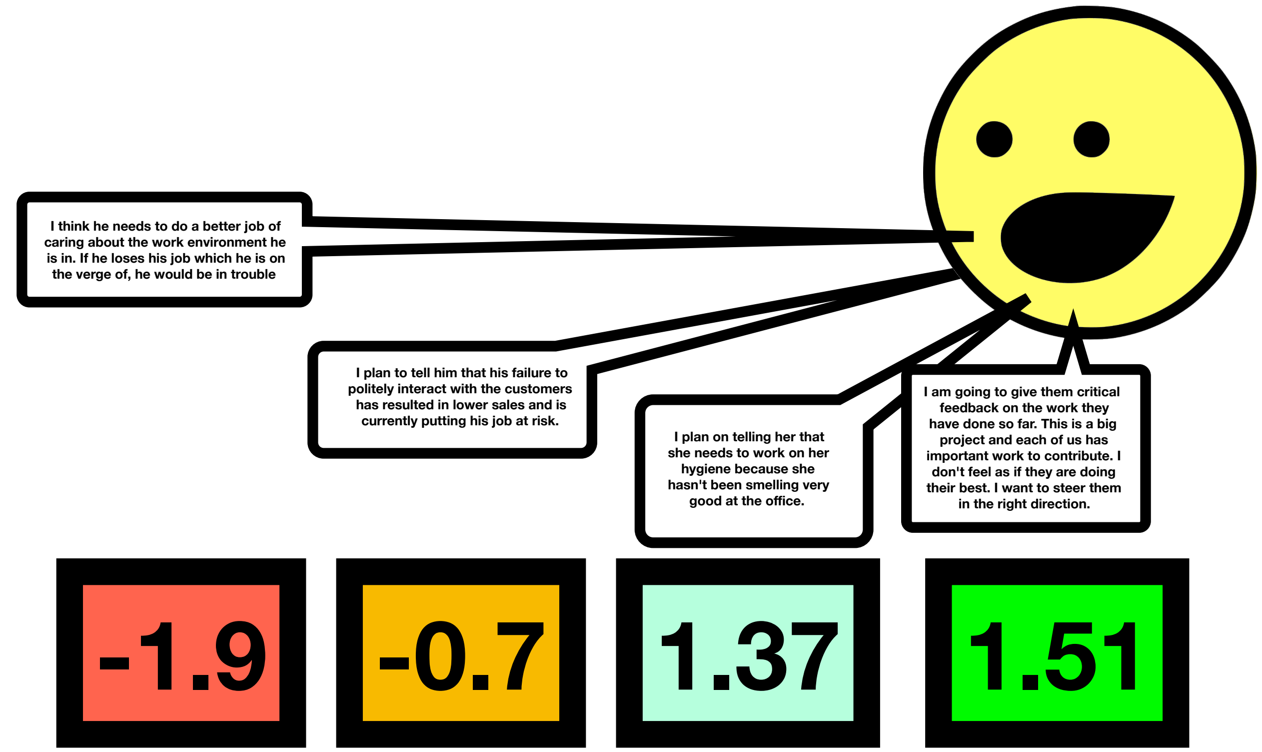
Measurement in Language

Humans: Plusses

What we've always been doing

More accurate than algorithms
for complex tasks

Can understand context, nuance



Humans: Minuses

High marginal cost of labor

Not reliable

Not transparent

Measurement in Language

Humans: Plusses

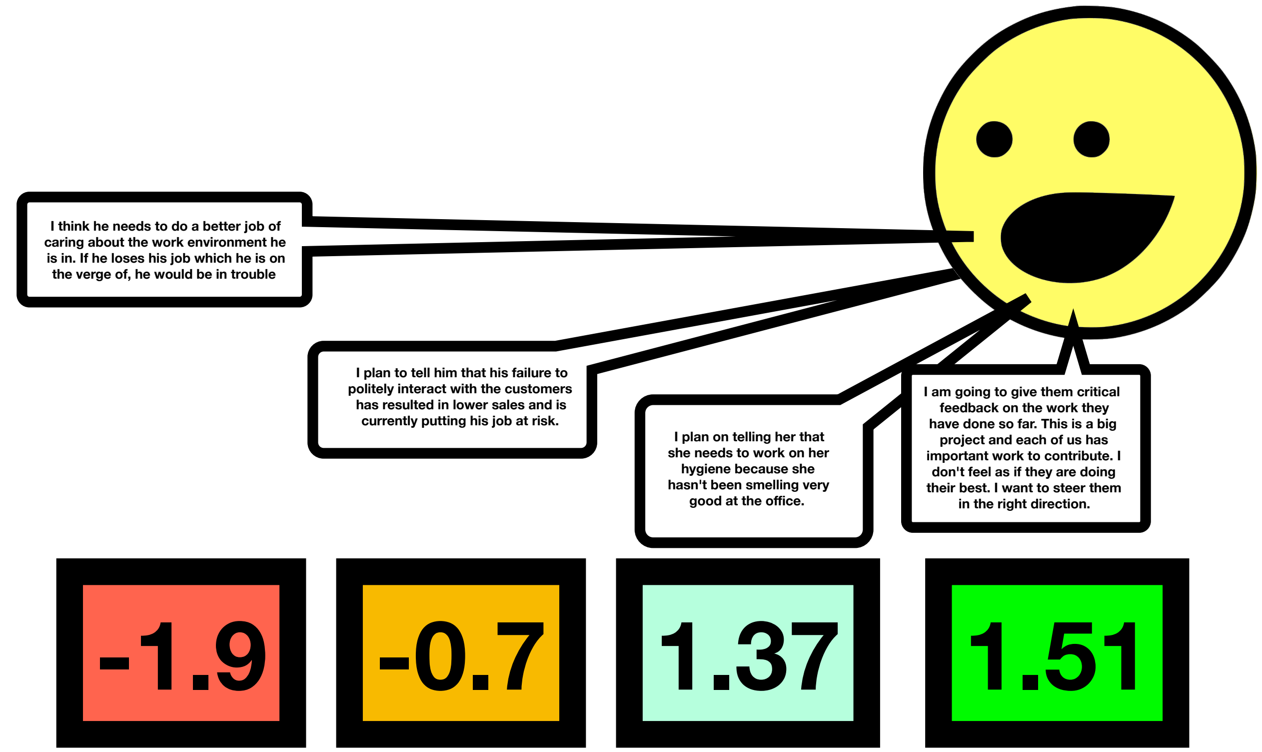
What we've always been doing

More accurate than algorithms

for complex tasks

Can understand context, nuance

<u>Give scores to train new algorithms</u>



Humans: Minuses

High marginal cost of labor

Not reliable

Not transparent

Why is Conversation Different?

Why is Conversation Different?

Three kinds of Dimensionality - words * time * goals!

Why is Conversation Different?

Three kinds of Dimensionality - words * time * goals!

Language choices within each turn

Why is Conversation Different?

Three kinds of Dimensionality - words * time * goals!

Language choices within each turn

Time

Why is Conversation Different?

Three kinds of Dimensionality - words * time * goals!

Language choices within each turn

Time

Emergent properties across multiple turns

Why is Conversation Different?

Three kinds of Dimensionality - words * time * goals!

Language choices within each turn

Time

Emergent properties across multiple turns

How do speakers establish/maintain common ground?

Why is Conversation Different?

Three kinds of Dimensionality - words * time * goals!

Language choices within each turn

Time

Emergent properties across multiple turns

How do speakers establish/maintain common ground?

Who drives agenda/topic choices?

Why is Conversation Different?

Three kinds of Dimensionality - words * time * goals!

Language choices within each turn

Time

Emergent properties across multiple turns

How do speakers establish/maintain common ground?

Who drives agenda/topic choices?

Goals

What are the speakers' individual goals?

Why is Conversation Different?

Three kinds of Dimensionality - words * time * goals!

Language choices within each turn

Time

Emergent properties across multiple turns
How do speakers establish/maintain common ground?
Who drives agenda/topic choices?

Goals

What are the speakers' individual goals?

- information exchange
- co-ordinating behaviour
- leaving a good impression
- enjoyment

Why is Conversation Different?

Three kinds of Dimensionality - words * time * goals!

Language choices within each turn

Time

Emergent properties across multiple turns

How do speakers establish/maintain common ground?

Who drives agenda/topic choices?

Goals

What are the speakers' individual goals?

- information exchange
- co-ordinating behaviour
- leaving a good impression
- enjoyment

Are those goals shared, or in conflict?

Type of Conversations

Open domain

- Speed date
- Social media
- Friendly chat
- Work meeting

Type of Conversations

Open domain

Speed date
Social media
Friendly chat
Work meeting

Structured domain

Type of Conversations

Open domain

Speed date
Social media
Friendly chat
Work meeting

Structured domain

Booking agents
Sales calls

Type of Conversations

Open domain

Speed date
Social media
Friendly chat
Work meeting

Structured domain

Booking agents
Sales calls

Type of Conversations

Open domain

Speed date
Social media
Friendly chat
Work meeting

Structured domain

Booking agents
Sales calls
Customer service
Crisis hotlines

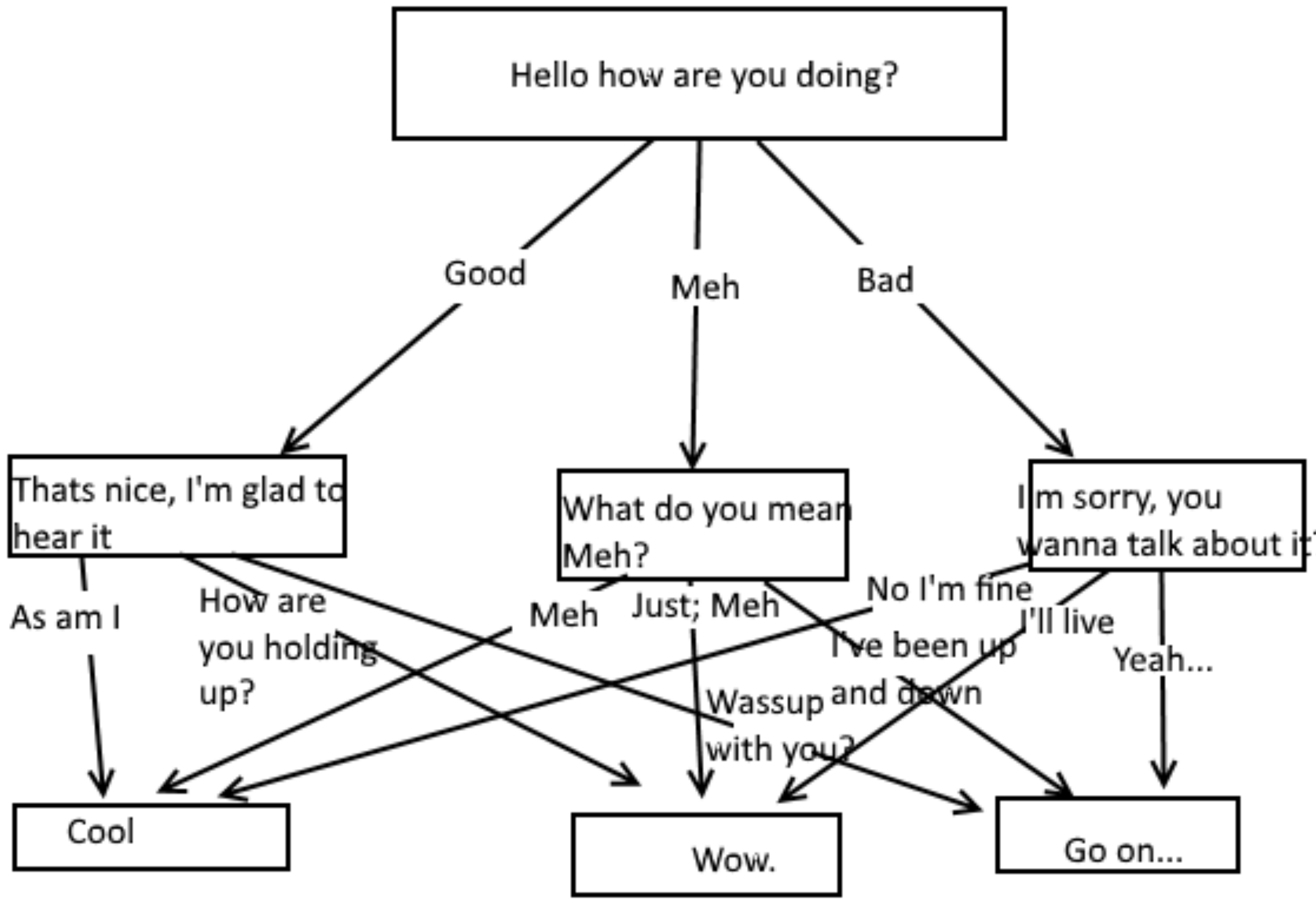
Type of Conversations

Open domain

- Speed date
- Social media
- Friendly chat
- Work meeting

Structured domain

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



Type of Conversations

Open domain

Speed date
Social media
Friendly chat
Work meeting

Structured domain

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

Type of Conversations

Open domain

Speed date
Social media
Friendly chat
Work meeting

Structured domain

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

<- Much easier to analyse!
if you take advantage of structure

Dialogue Acts

Fragment of a labeled conversation (from the Switchboard corpus).

Speaker	Dialogue Act	Utterance
A	YES-NO-QUESTION	So do you go to college right now?
A	ABANDONED	Are yo-,
B	YES-ANSWER	*Yeah,*
B	STATEMENT	*it's my last year [laughter].*
A	DECLARATIVE-QUESTION	You're a, so you're a senior now.
B	YES-ANSWER	*Yeah,*
B	STATEMENT	*I'm working on my projects trying to graduate [laughter].*
A	APPRECIATION	Oh, good for you.
B	BACKCHANNEL	*Yeah.*
A	APPRECIATION	That's great,
A	YES-NO-QUESTION	um, is, is N C University is that, uh, State,
B	STATEMENT	*N C State.*
A	SIGNAL-NON-UNDERSTANDING	What did you say?
B	STATEMENT	*N C State.*

(Schutze et al., 2000)

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%

(Schutze et al., 2000)

Politeness Features

(Yeomans et al., 2018)

Politeness Features

Common social words & phrases

Second Person, First Person Plural, First Person Single, Goodbye, Hello, Impersonal Pronoun, Can You, Could You, For Me, For You, Ask Agency, Give Agency, Please, By The Way, Let Me Know

(Yeomans et al., 2018)

Politeness Features

Common social words & phrases

Second Person, First Person Plural, First Person Single, Goodbye, Hello, Impersonal Pronoun, Can You, Could You, For Me, For You, Ask Agency, Give Agency, Please, By The Way, Let Me Know

Dictionaries

Filler Pause, Reasoning, Reassurance, Negation, Informal Title, Formal Title, Hedges, Swearing

(Yeomans et al., 2018)

Politeness Features

Common social words & phrases

Second Person, First Person Plural, First Person Single, Goodbye, Hello, Impersonal Pronoun, Can You, Could You, For Me, For You, Ask Agency, Give Agency, Please, By The Way, Let Me Know

Dictionaries

Filler Pause, Reasoning, Reassurance, Negation, Informal Title, Formal Title, Hedges, Swearing

Sentence Structure

Conjunction Start, WH Questions, YesNo Questions, Affirmation, Adverb Limiter

(Yeomans et al., 2018)

Politeness Features

Common social words & phrases

Second Person, First Person Plural, First Person Single, Goodbye, Hello, Impersonal Pronoun, Can You, Could You, For Me, For You, Ask Agency, Give Agency, Please, By The Way, Let Me Know

Dictionaries

Filler Pause, Reasoning, Reassurance, Negation, Informal Title, Formal Title, Hedges, Swearing

Sentence Structure

Conjunction Start, WH Questions, YesNo Questions, Affirmation, Adverb Limiter

Dependency Parsing

Acknowledgement, Bare Command, Gratitude, Agreement, Disagreement, Subjectivity, Truth Intensifier, Apology, Negative Emotion, Positive Emotion

(Yeomans et al., 2018)

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

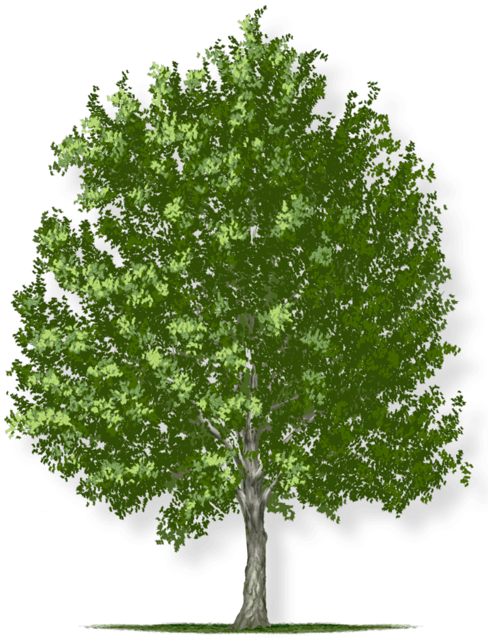
Apologies

"I apologize for my behavior"

"I apologise for my behaviour"

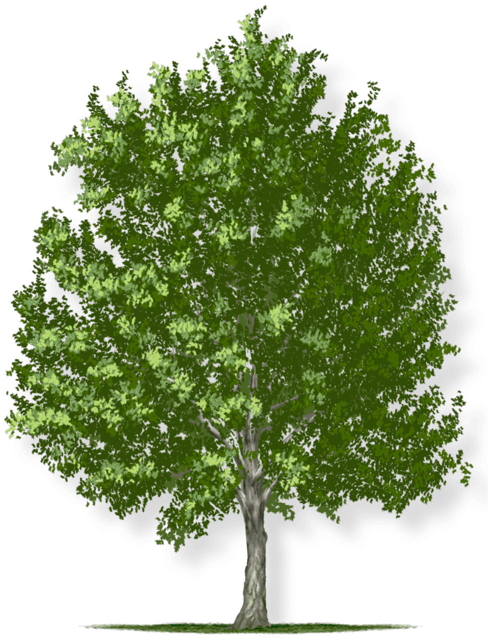
"Not until you apologise for your behaviour"

"I sincerely apologise for my behaviour"



Dependency Trees

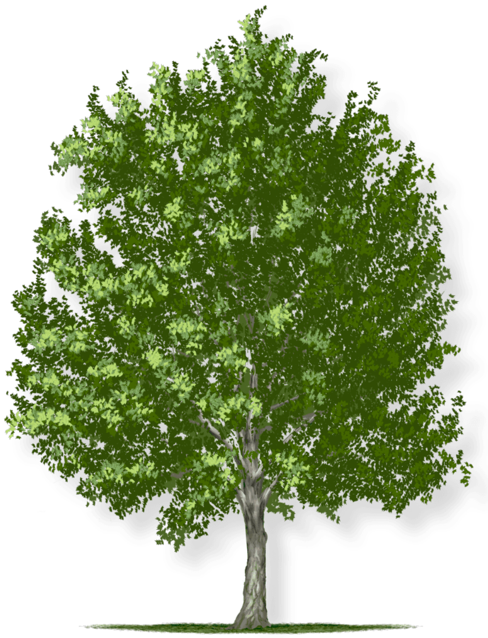
I sincerely apologise for my behaviour



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

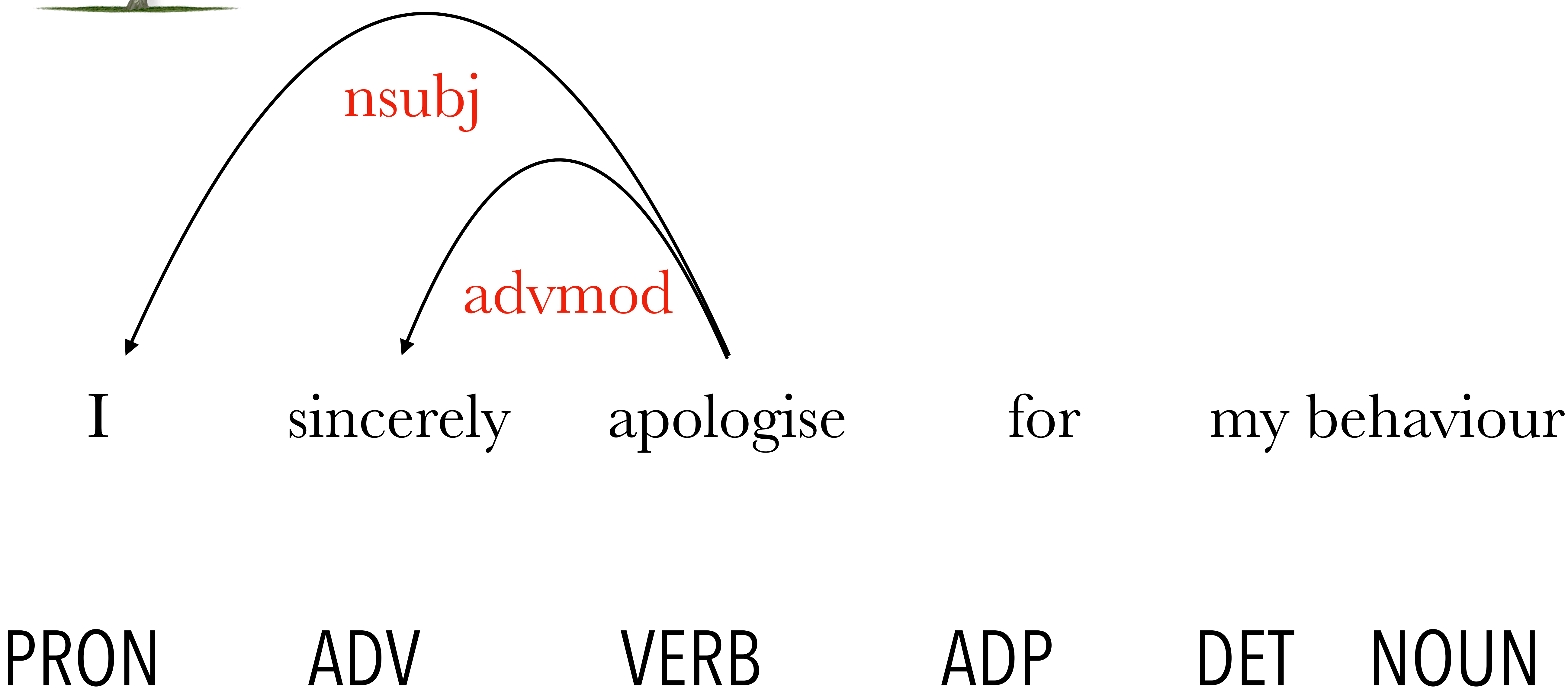
Dependency Trees

nsubj

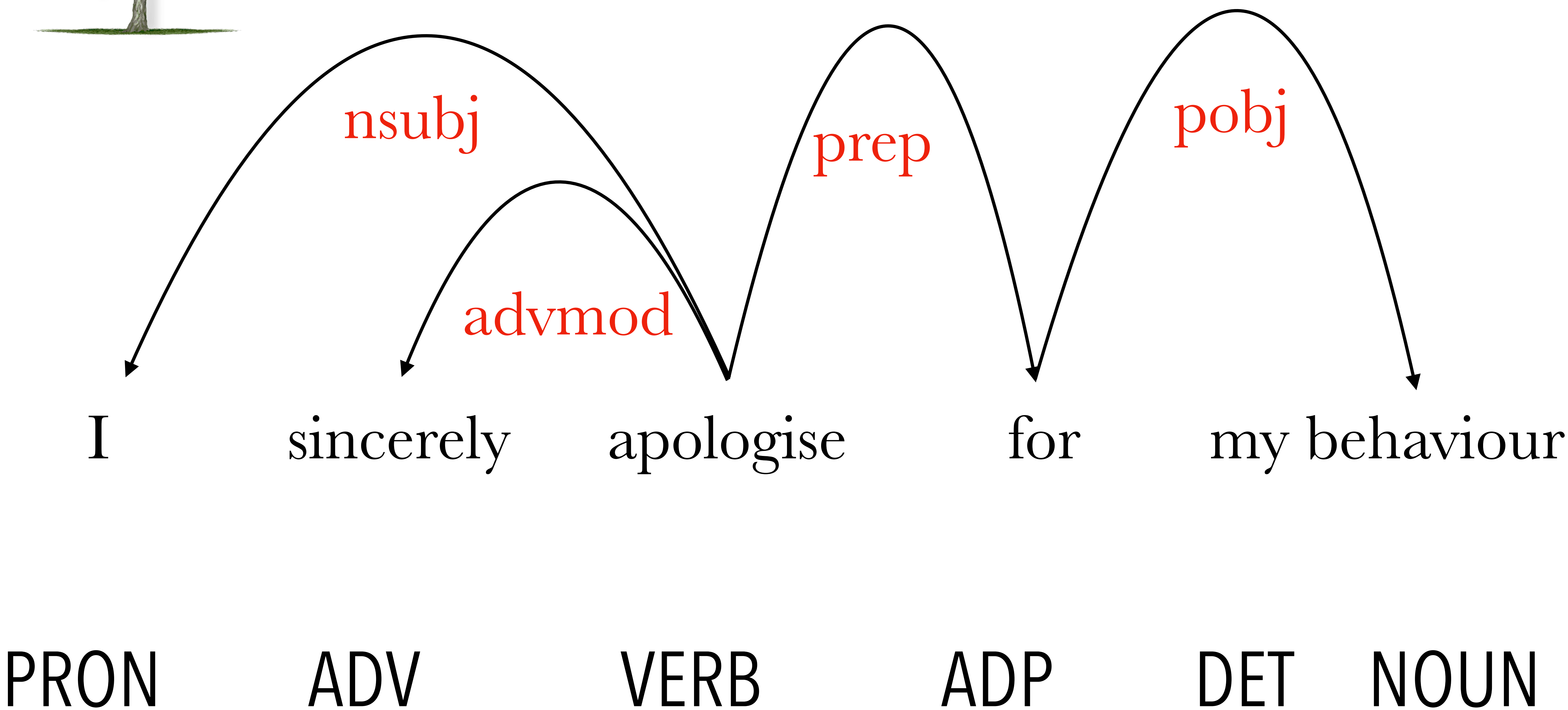
I sincerely apologise for my behaviour

PRON ADV VERB ADP DET NOUN

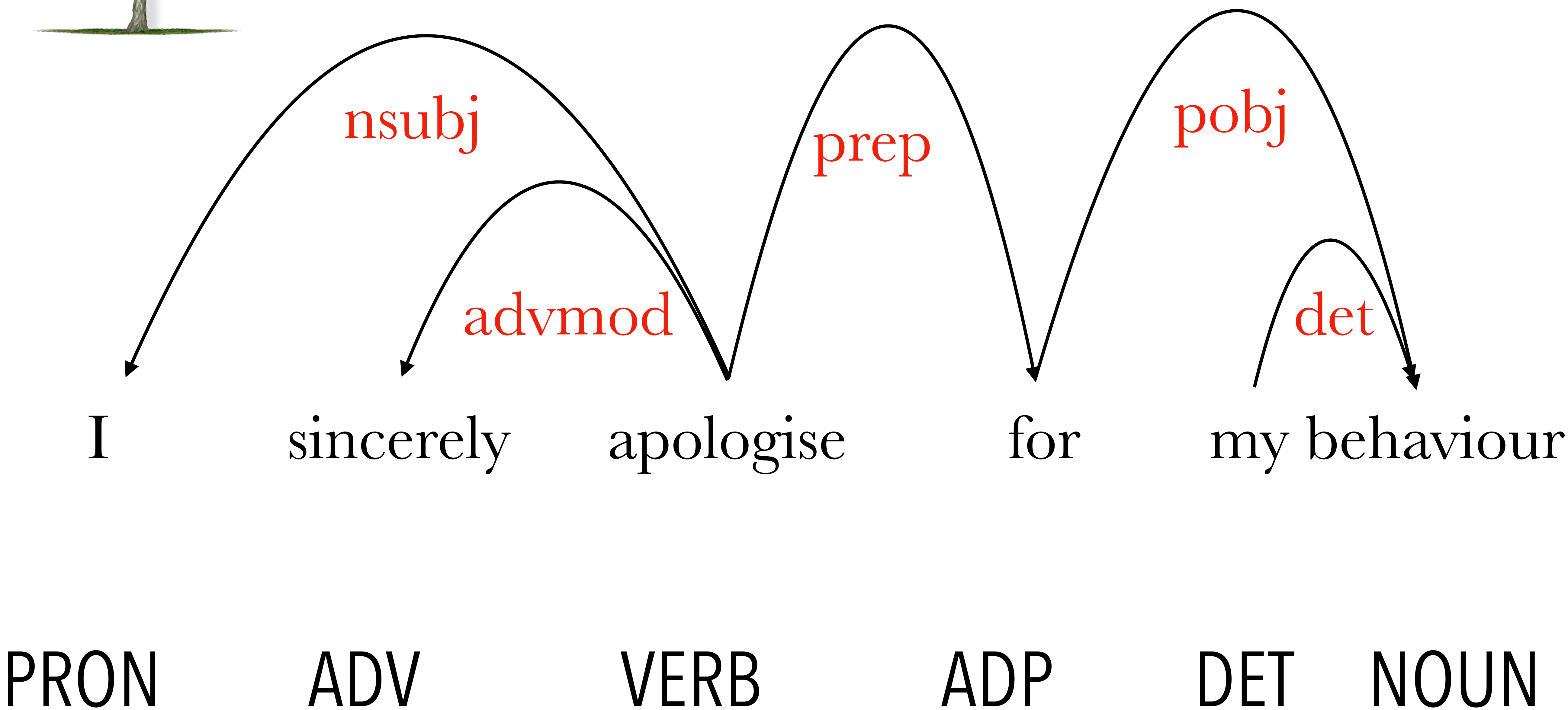
Dependency Trees



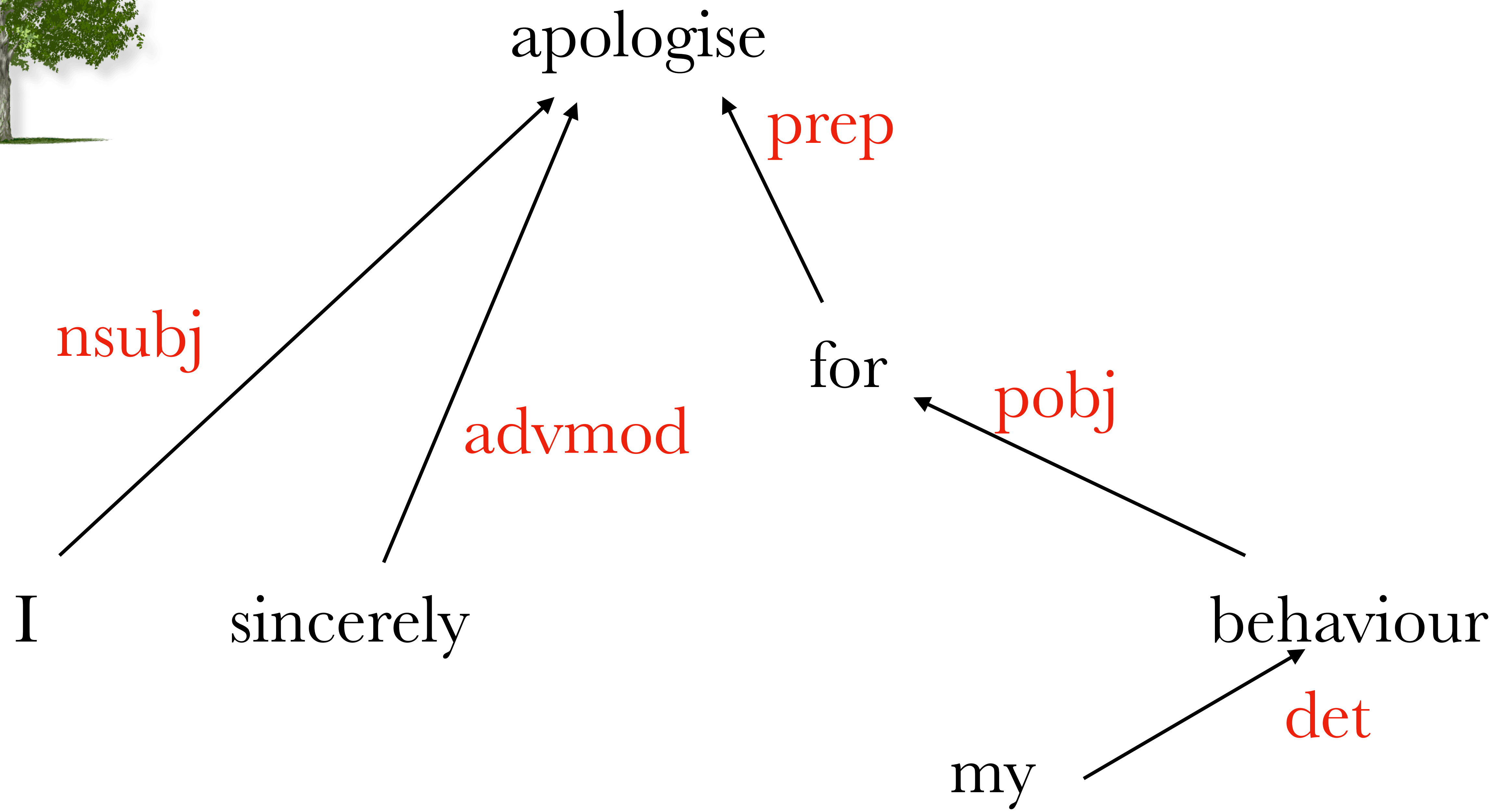
Dependency Trees

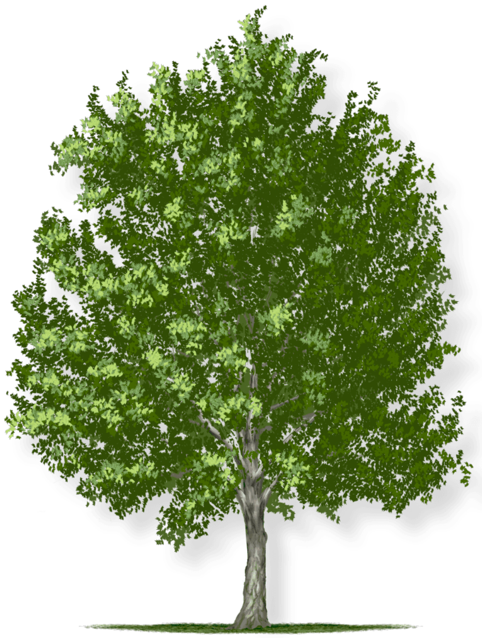


Dependency Trees



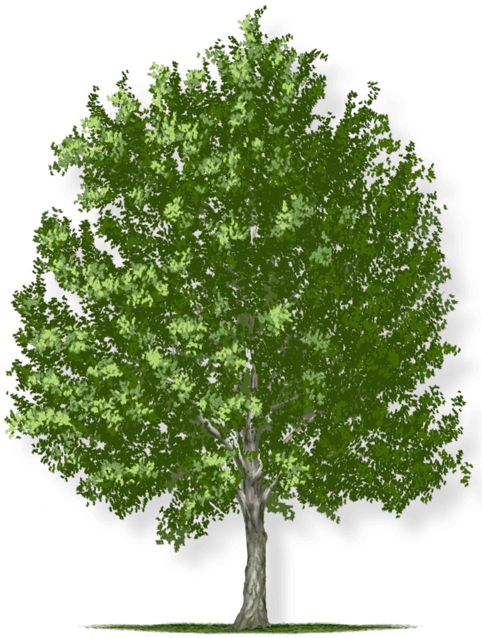
Dependency Trees





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

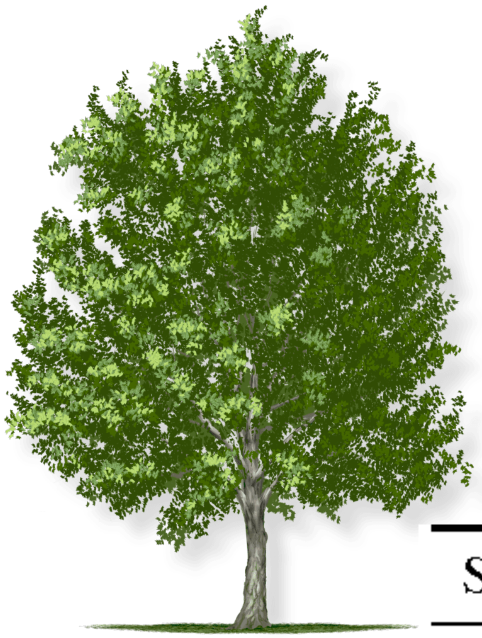
Dependency Trees

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)

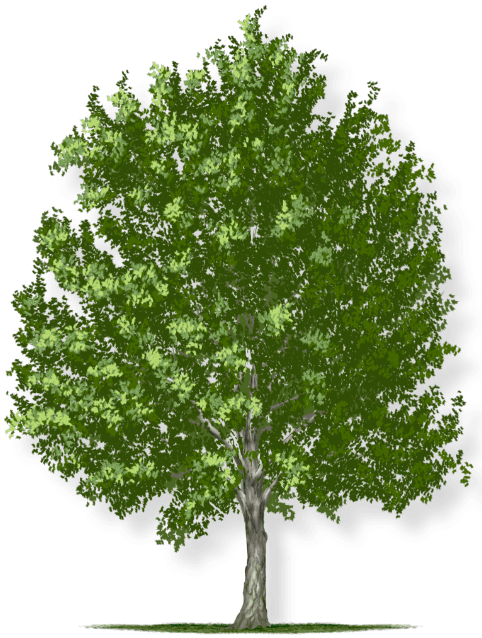
Dependency Trees

	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)

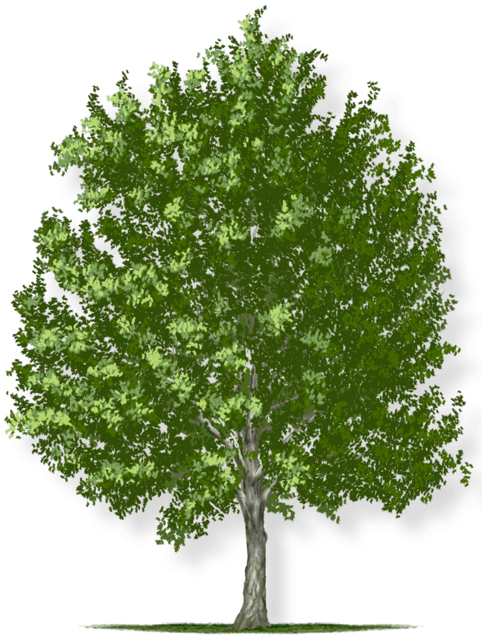


Dependency Trees

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			

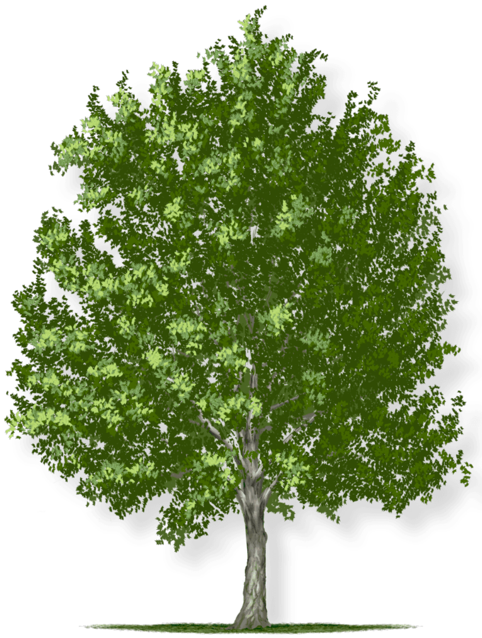


Dependency Trees



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.

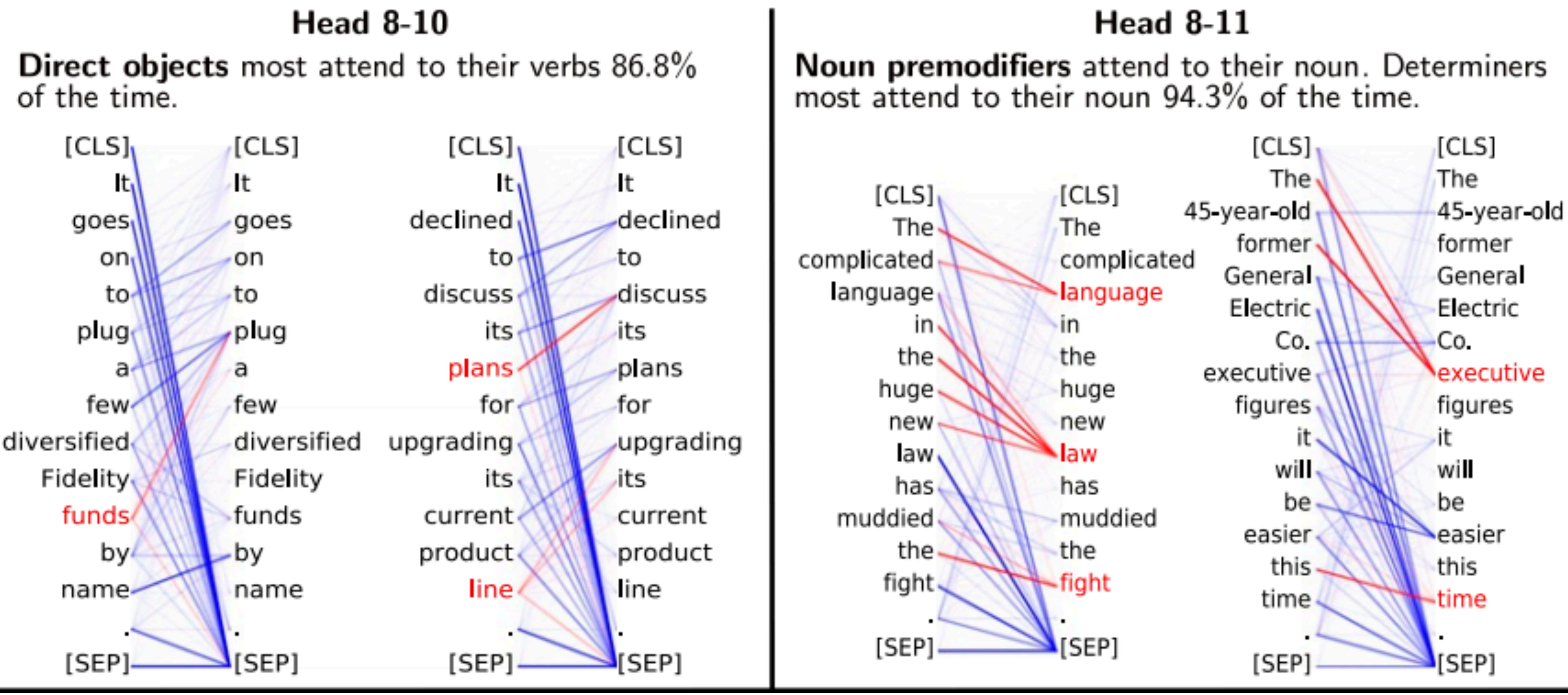
One might expect that a machine-learning model trained to predict the next word in a text will just be a giant associational learning machine, with lots of statistics on how often the word restaurant is followed by kitchen and perhaps some basic abstracted sequence knowledge such as knowing that adjectives 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

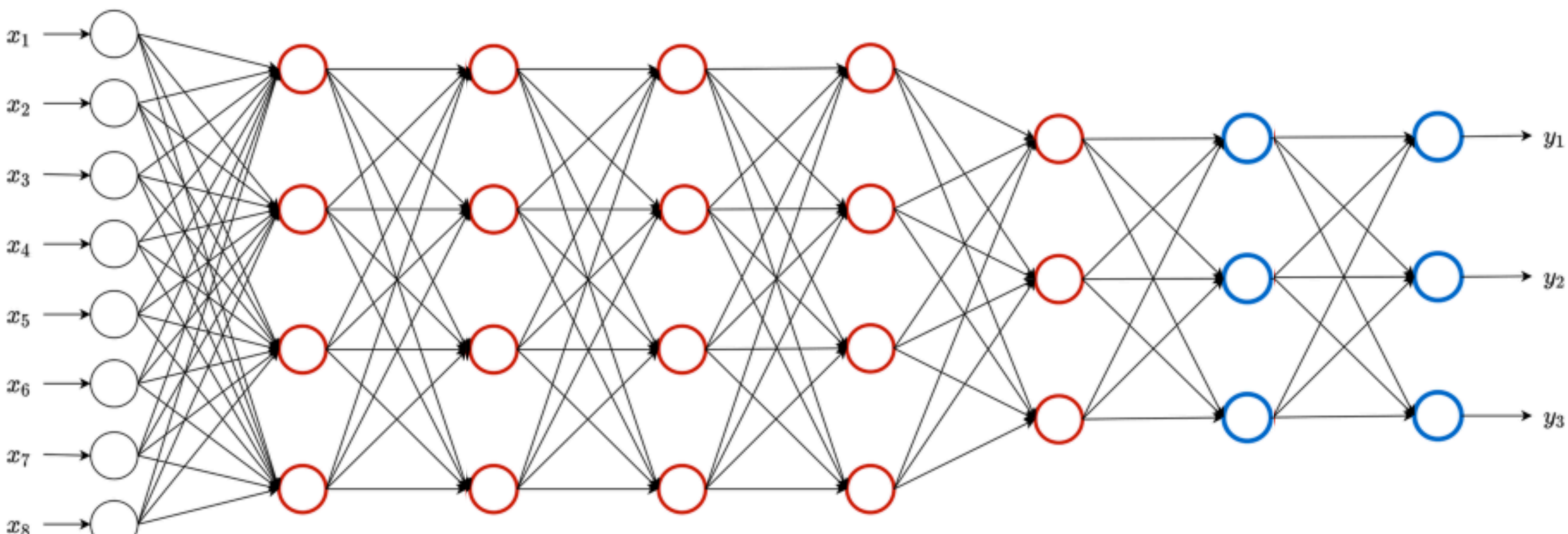
Christopher D. Manning[a,1], Kevin Clark[a], John Hewitt[a], Urvashi Khandelwal[a], and Omer Levy[b]

[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

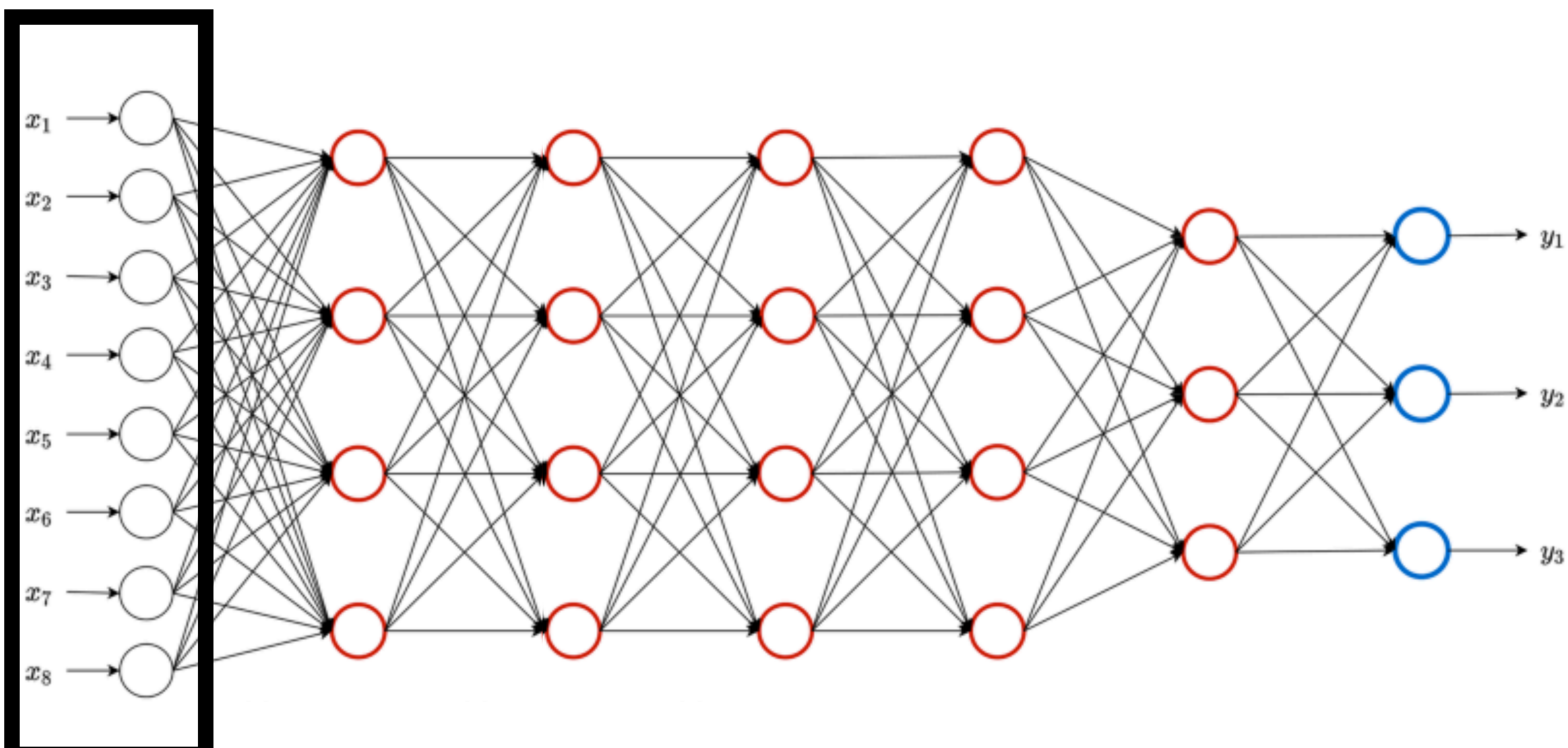


Pre-Trained Models



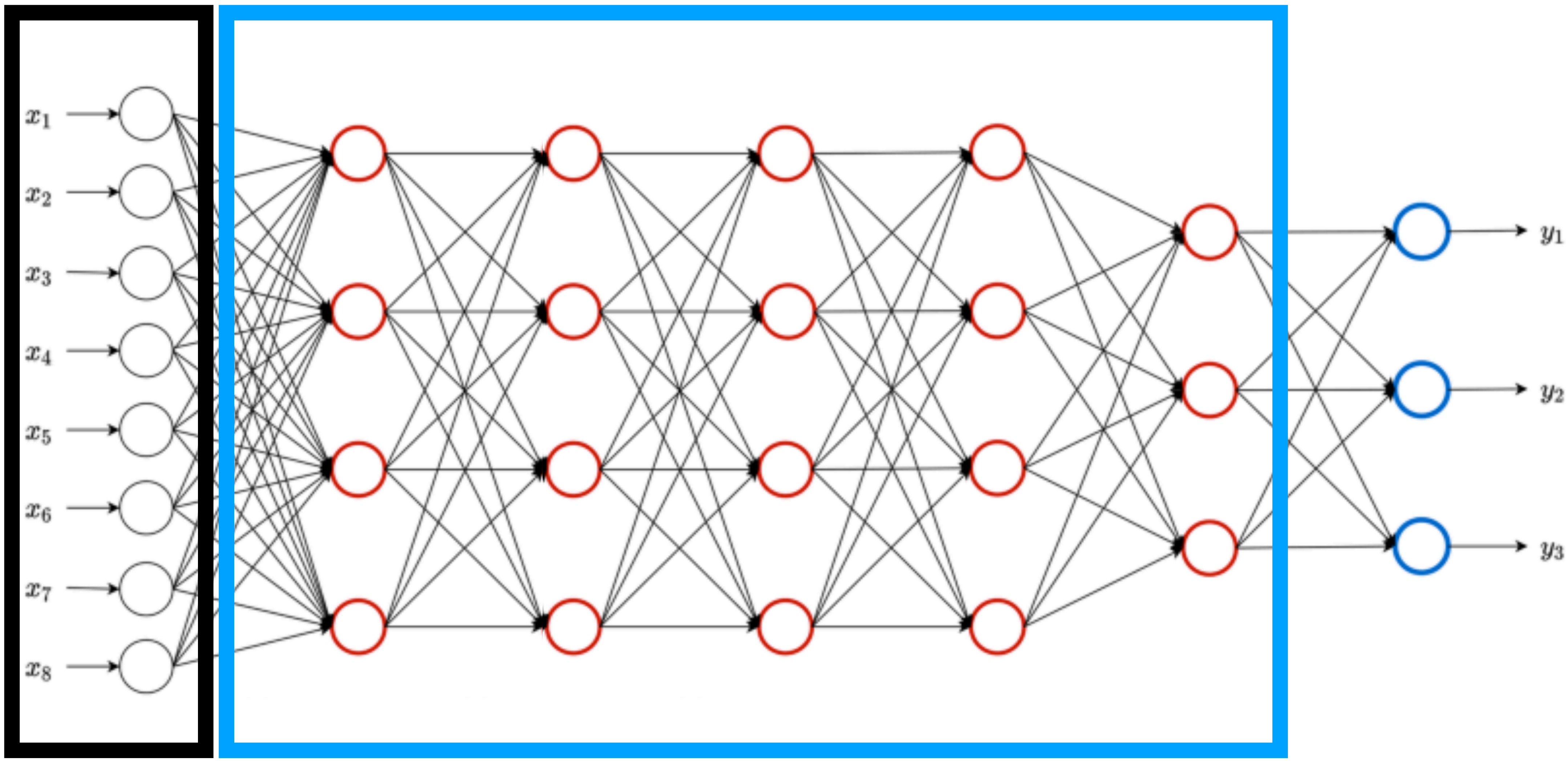
Pre-Trained Models

Words in a Document



Pre-Trained Models

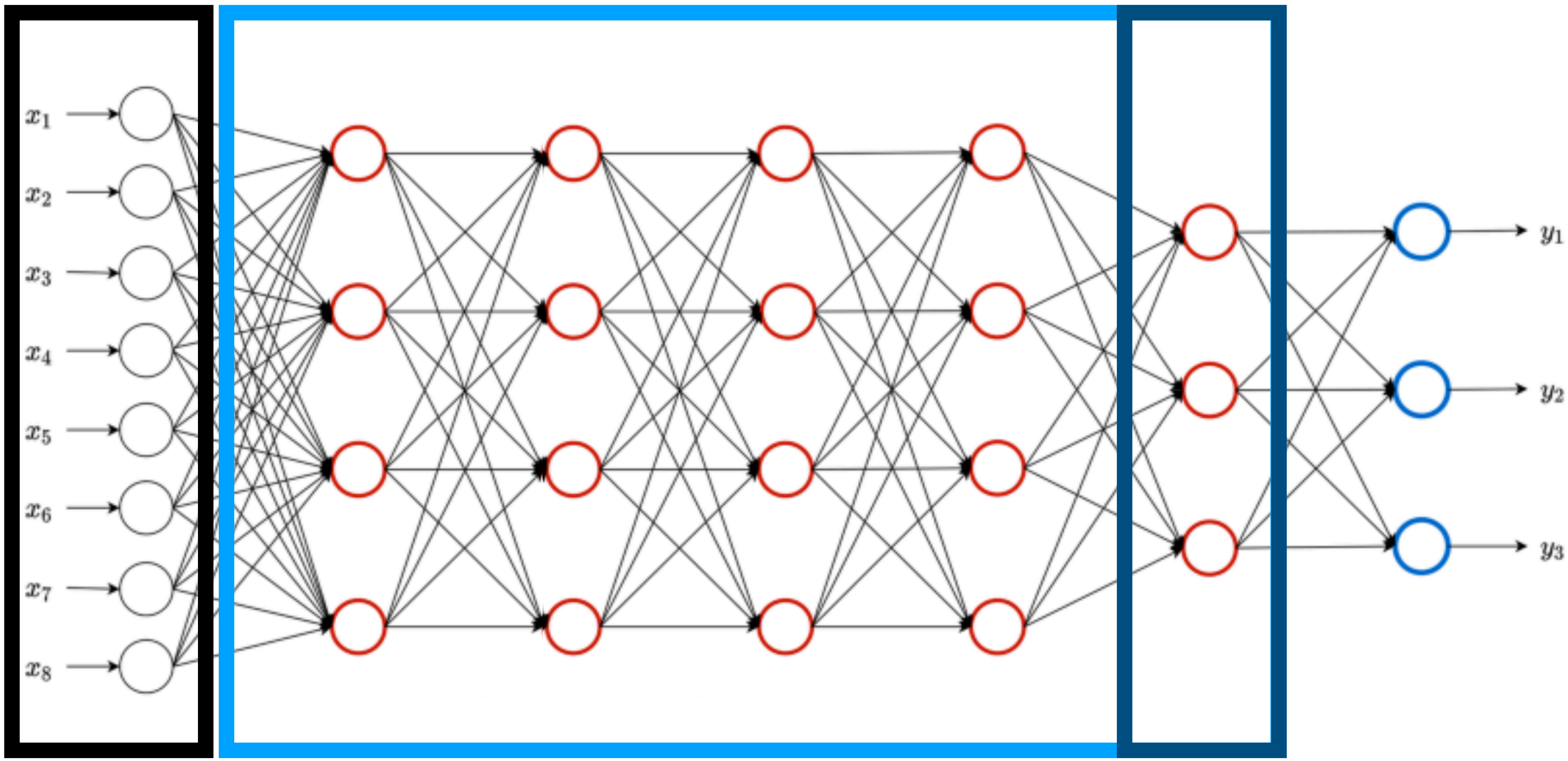
spaCy



Pre-Trained Models

spaCy

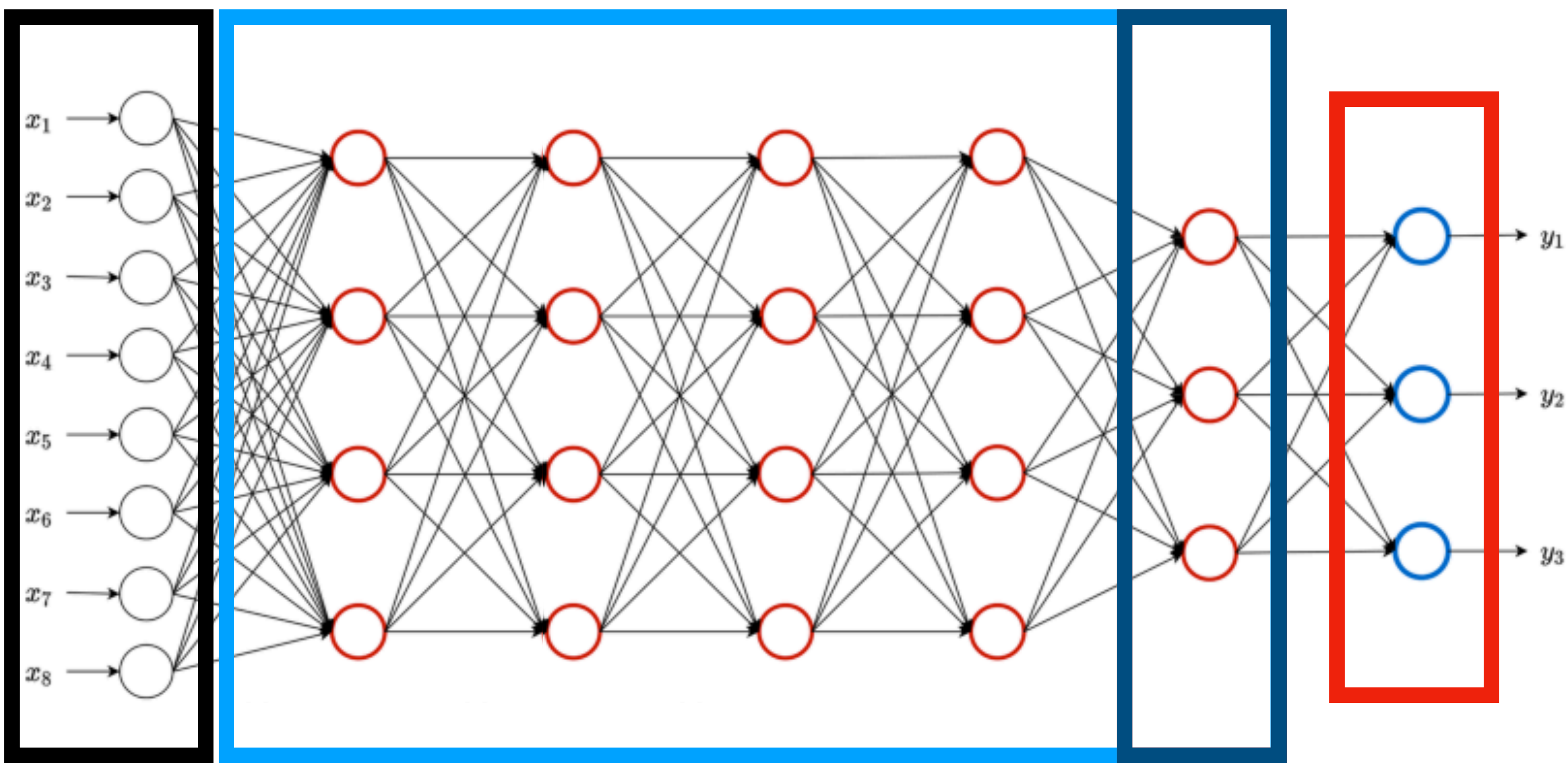
Grammar Labels



Pre-Trained Models

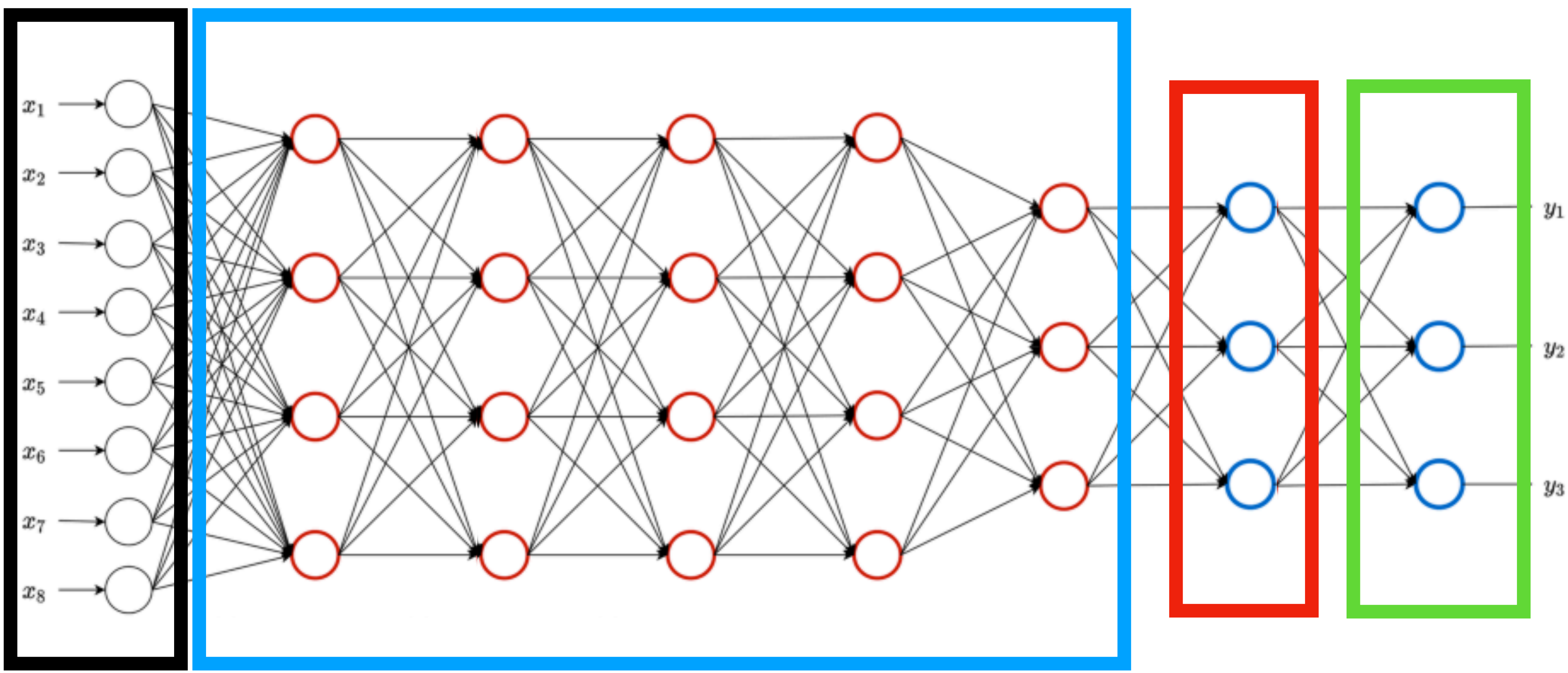
spaCy

Politeness Features



Pre-Trained Models

New quantity of interest



Communicating Warmth in Distributive Negotiations is Surprisingly Counter-Productive

(Jeong, Minson, Yeomans & Gino, 2018)

Communicating Warmth in Distributive Negotiations is Surprisingly Counter-Productive

(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 Distributive Negotiations is Surprisingly Counter-Productive

(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

Warm offers earn worse final deals

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

A Linguistic Model of Warmth

1. Collect negotiation transcripts

A Linguistic Model of Warmth

1. Collect negotiation transcripts

2. Randomly assign their negotiation strategy

A Linguistic Model of Warmth

1. Collect negotiation transcripts

2. Randomly assign their negotiation strategy

3. Train an algorithm to detect the strategy automatically

A Linguistic Model of Warmth

1. Collect negotiation transcripts

2. Randomly assign their negotiation strategy

3. Train an algorithm to detect the strategy automatically

We want to build an algorithm that is:

- valid

A Linguistic Model of Warmth

1. Collect negotiation transcripts

2. Randomly assign their negotiation strategy

3. Train an algorithm to detect the strategy automatically

We want to build an algorithm that is:

- valid
- scaleable

A Linguistic Model of Warmth

1. Collect negotiation transcripts

2. Randomly assign their negotiation strategy

3. Train an algorithm to detect the strategy automatically

We want to build an algorithm that is:

- valid
- scaleable
- interpretable

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 $\boldsymbol{x}$ 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:

- valid
- interpretable
- scaleable

Communicating Warmth in …

mTurkers can write phone offers! (n = 355)

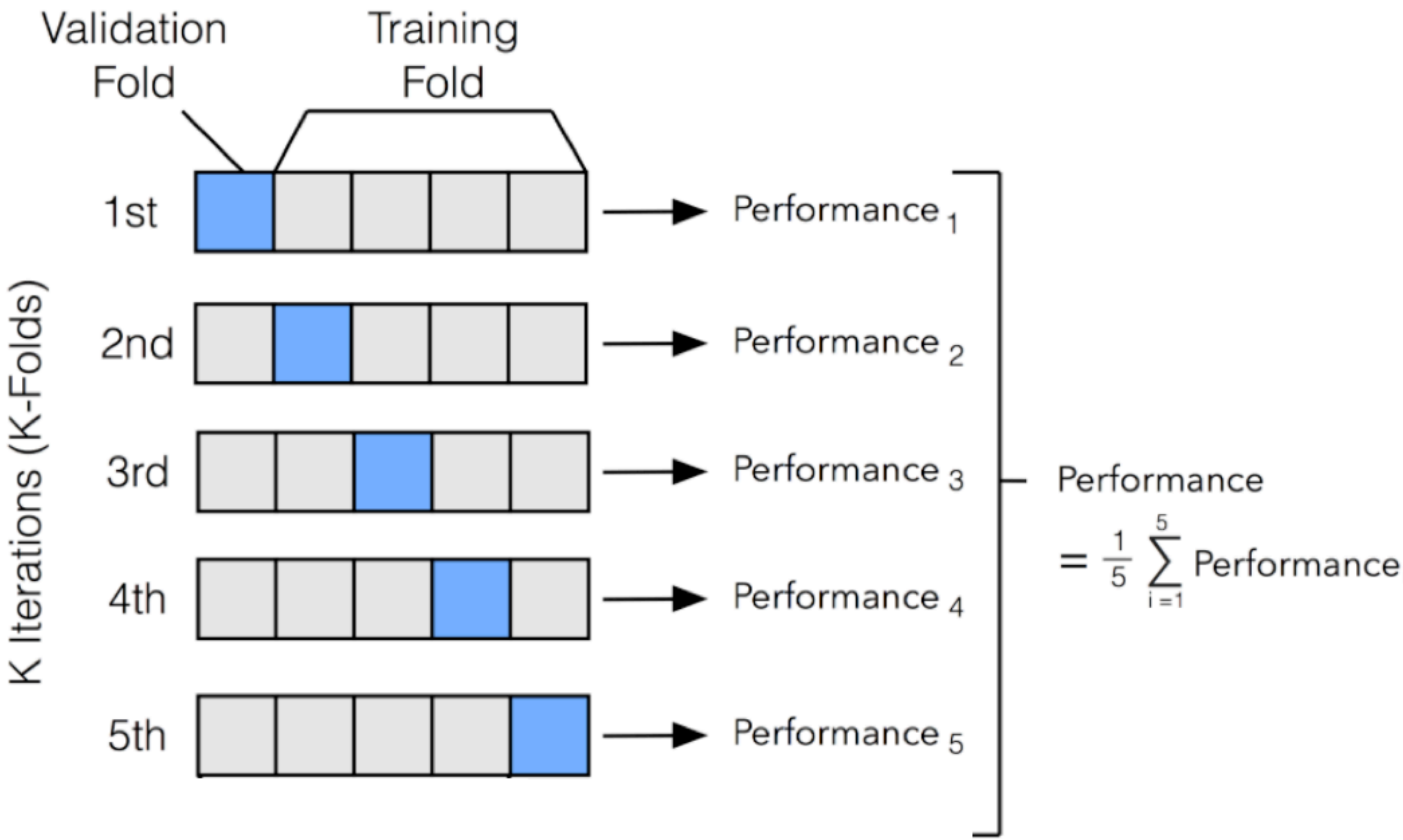
- encouragement design - tough vs. warm instructions
- “ground truth” is treatment effect

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

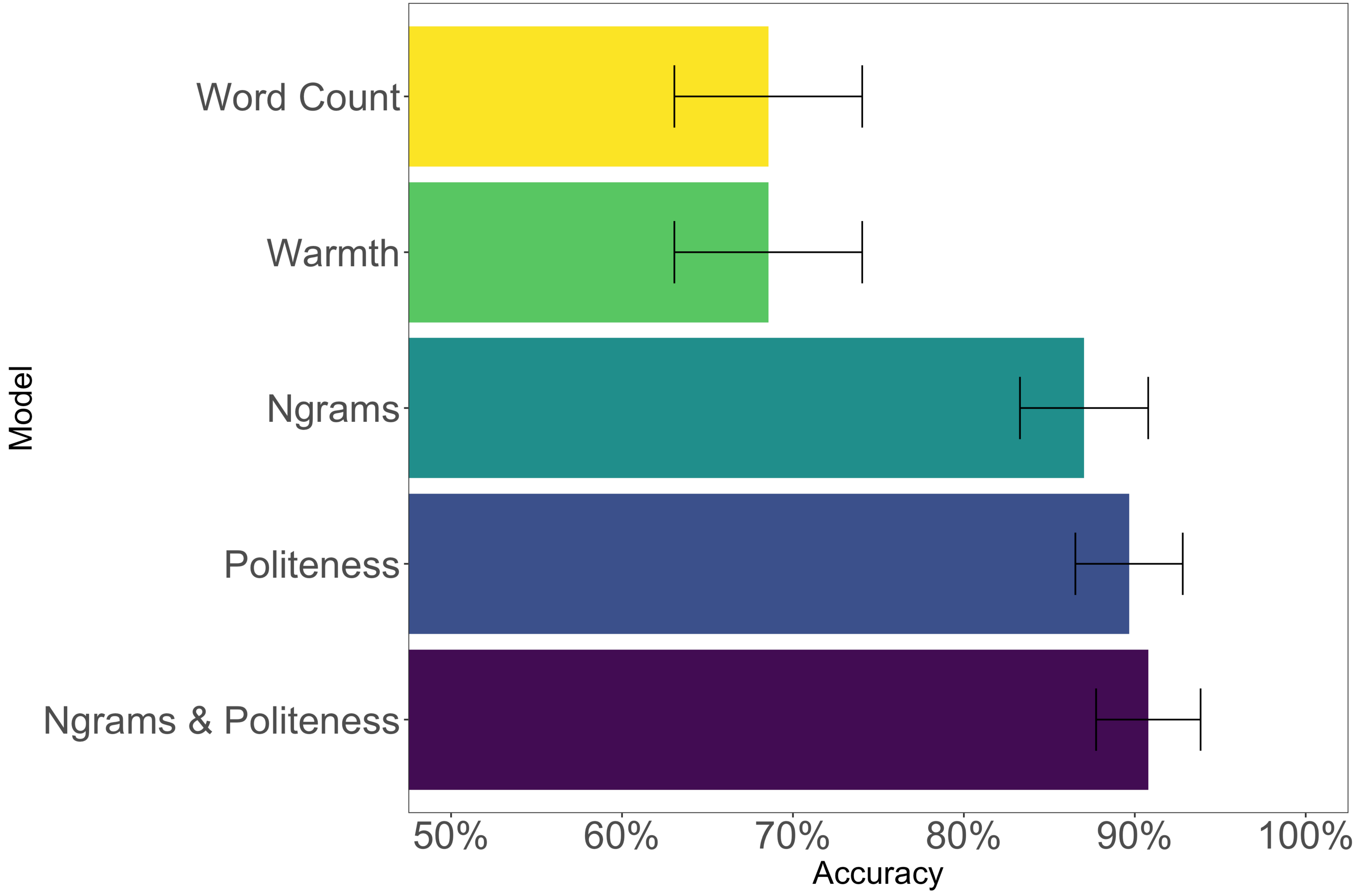
We want to build an algorithm that is:

- **valid**
- interpretable
- scaleable

Estimating Validity



Communicating Warmth in …



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:

- valid
- **interpretable**
- 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 ...

mTurkers can write phone offers! (n = 355)

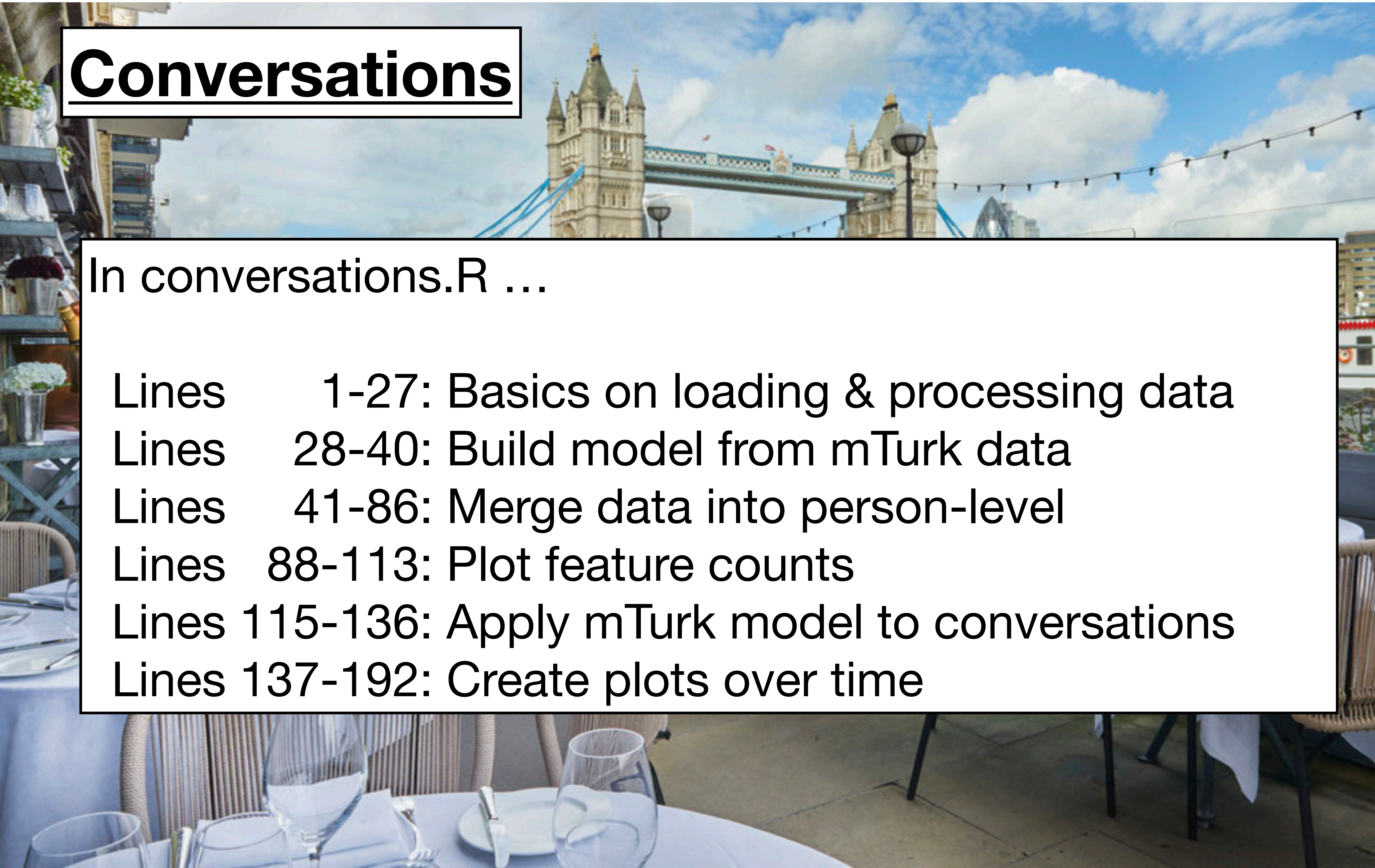
- encouragement design - tough vs. warm instructions
- "ground truth" is treatment effect

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:

- valid
- interpretable
- **scaleable**

Application in Your Code

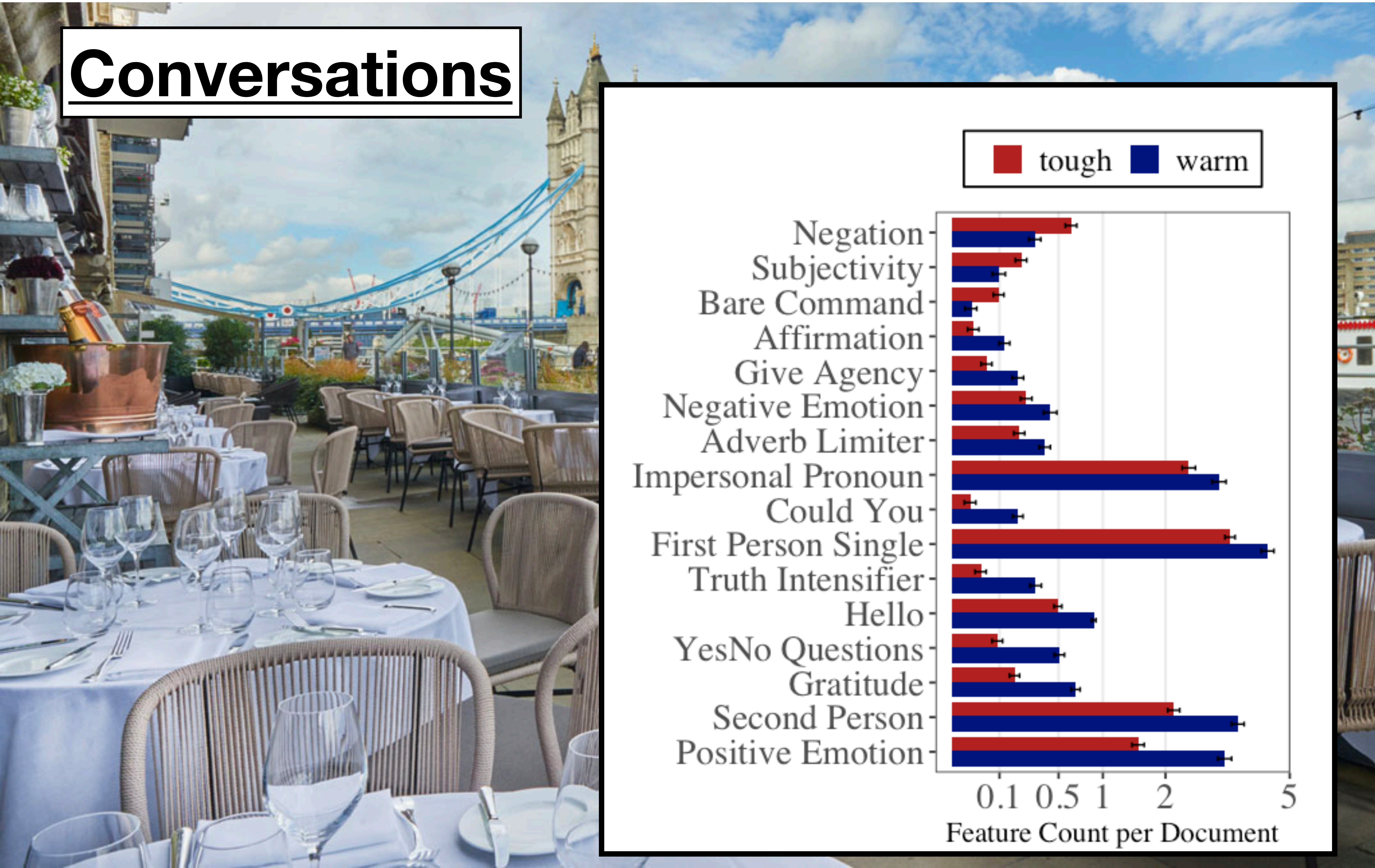


Conversations

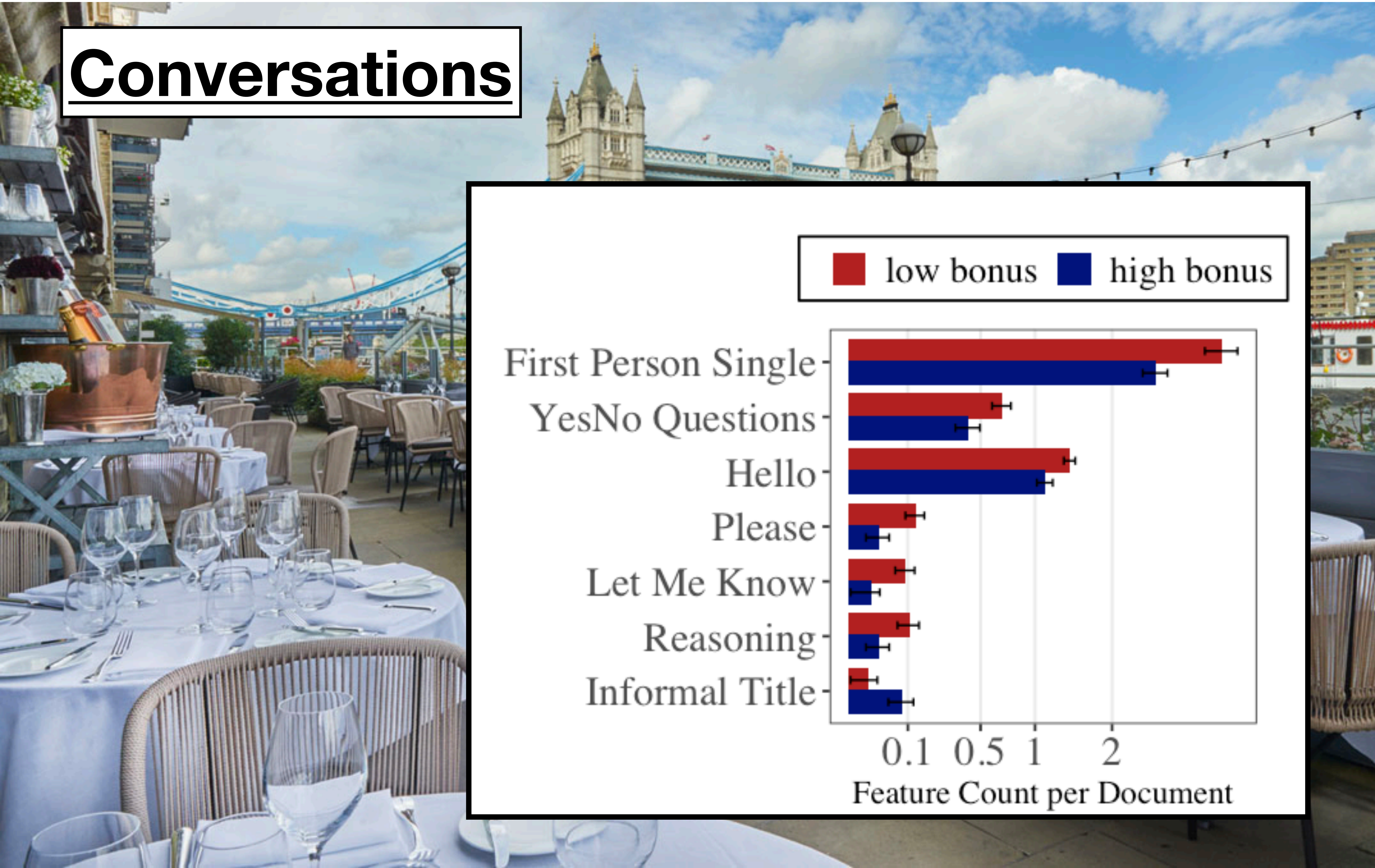
In conversations.R …

Lines 1-27: Basics on loading & processing data
Lines 28-40: Build model from mTurk data
Lines 41-86: Merge data into person-level
Lines 88-113: Plot feature counts
Lines 115-136: Apply mTurk model to conversations
Lines 137-192: Create plots over time

Application in Your Code

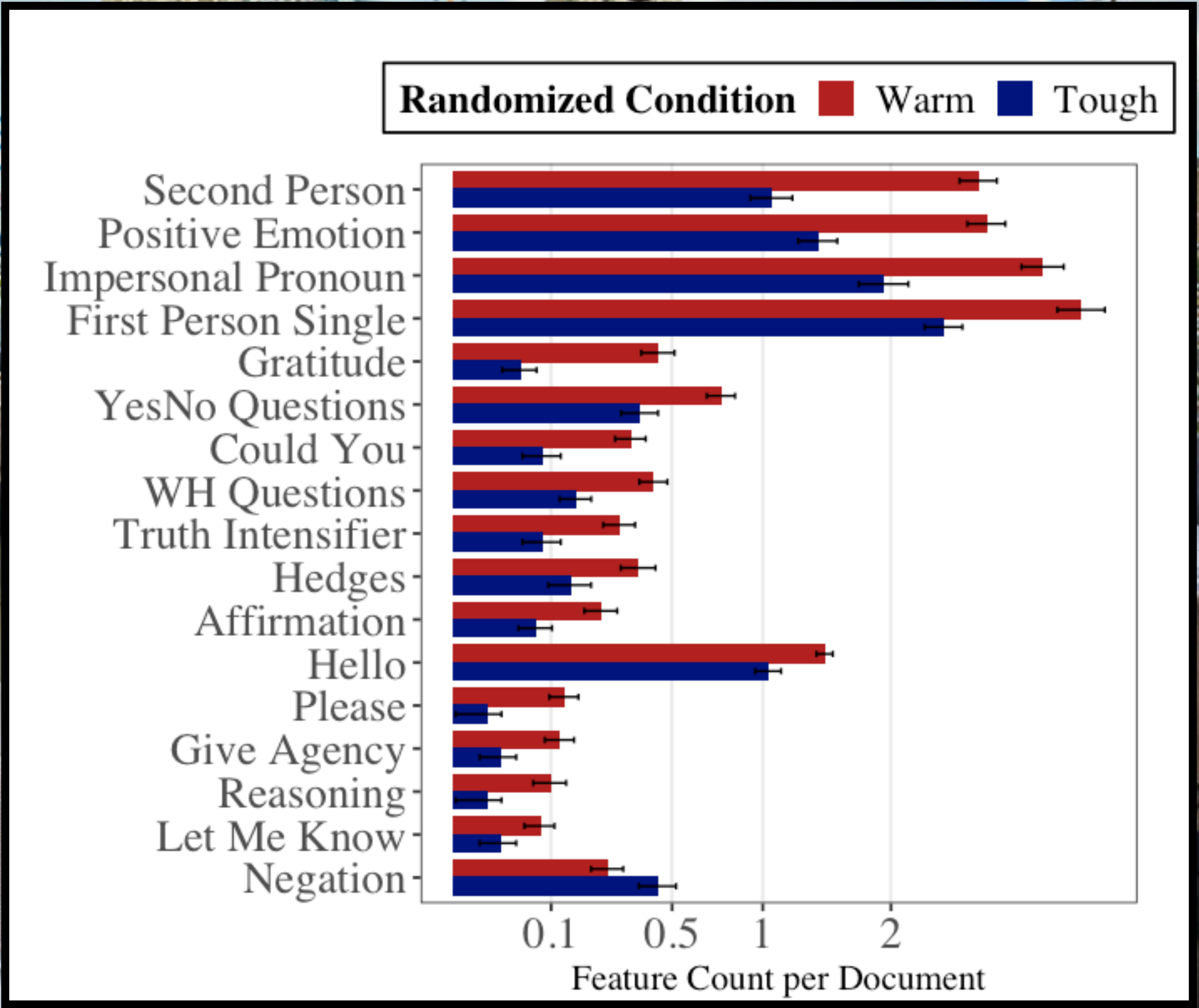


Application in Your Code



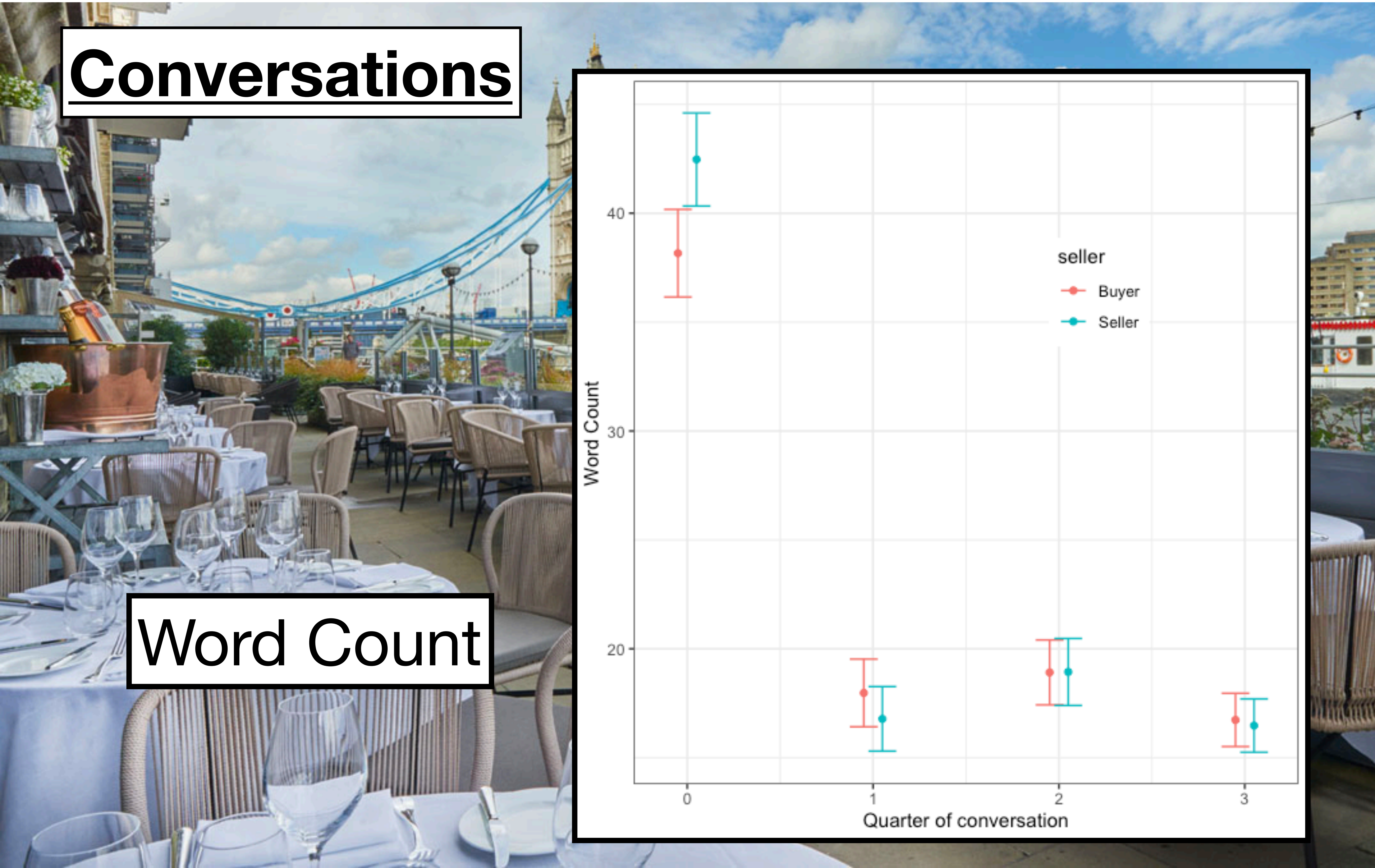
Application in Your Code

Conversations

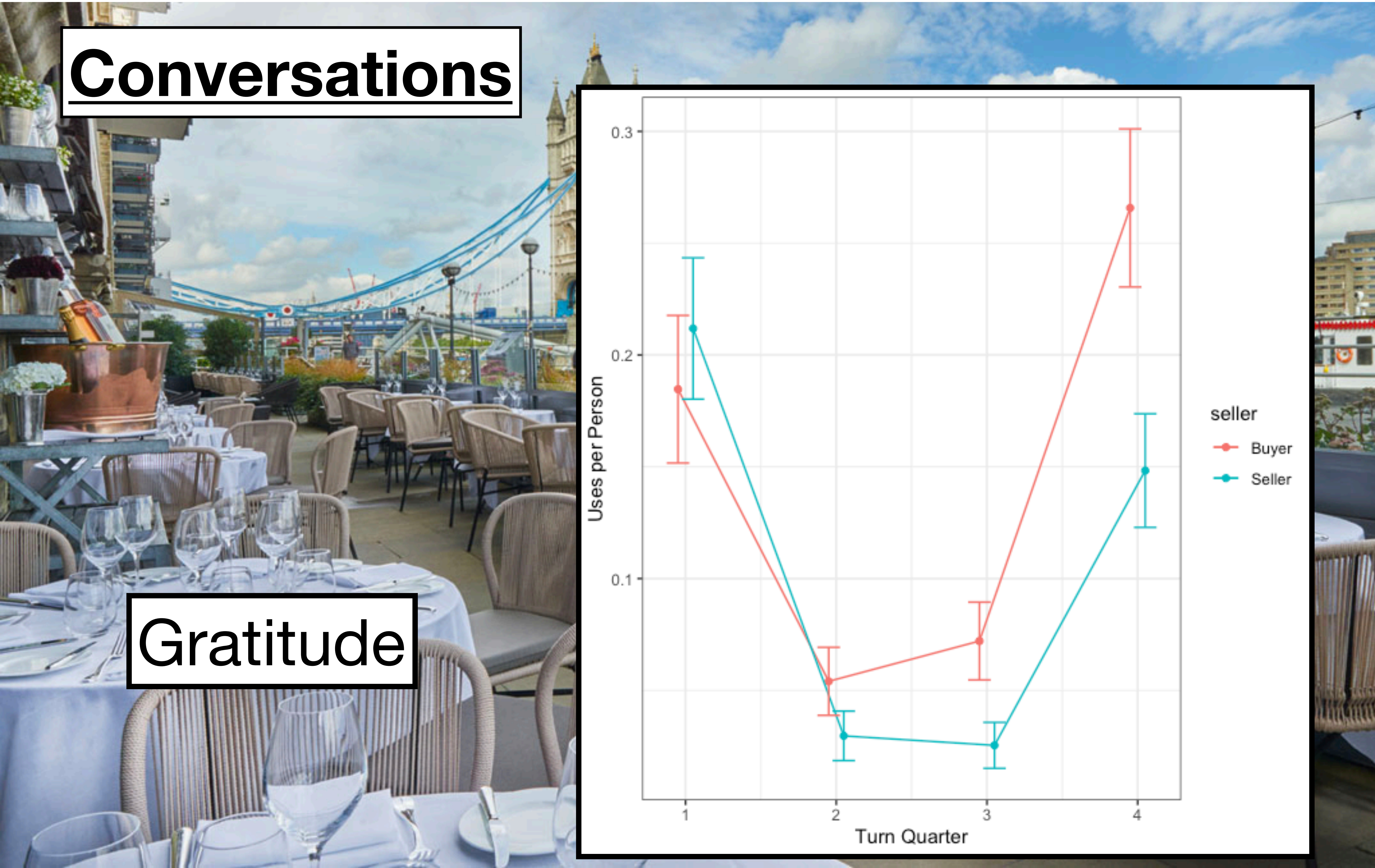


Application in Your Code

Application in Your Code



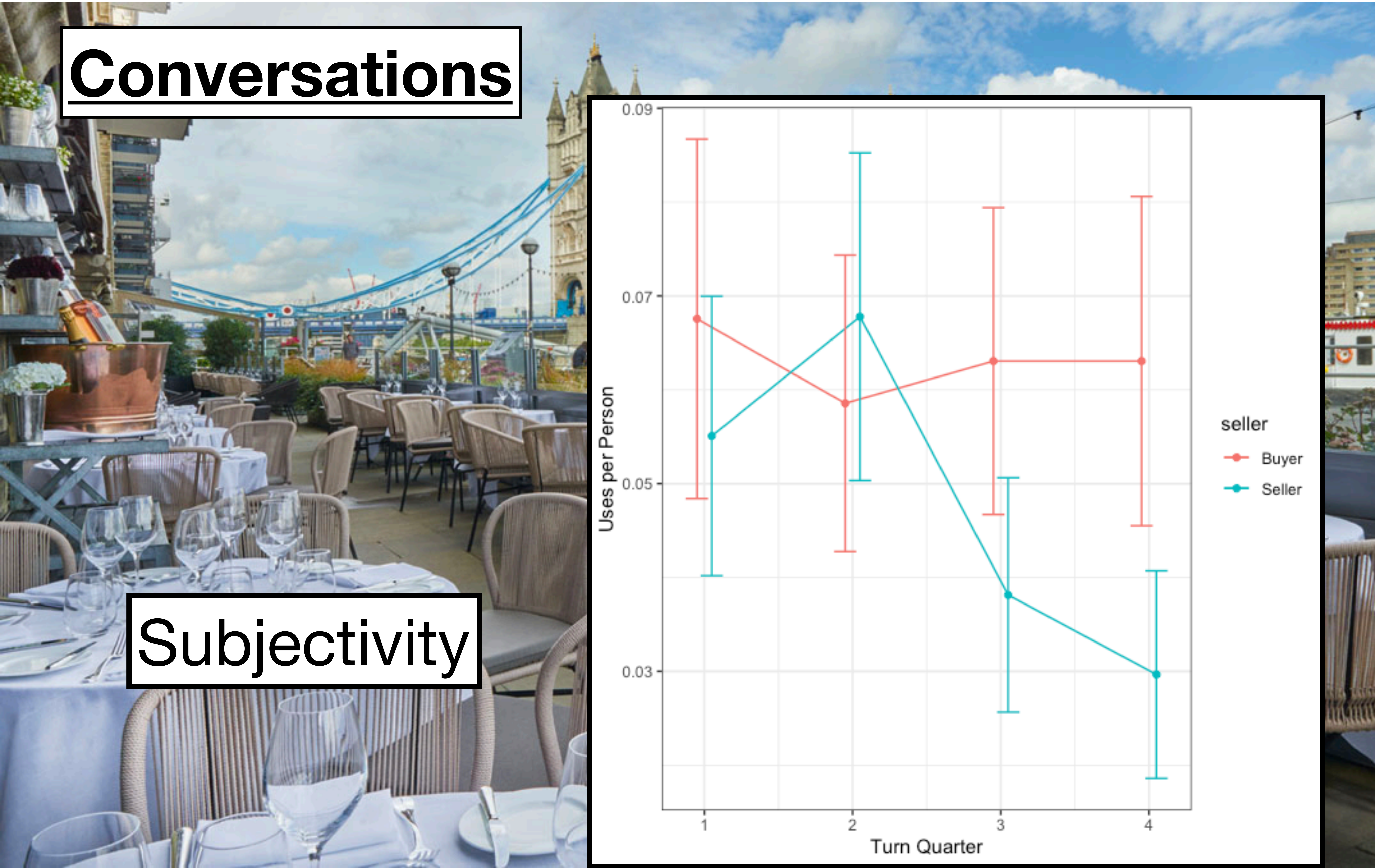
Application in Your Code



Application in Your Code

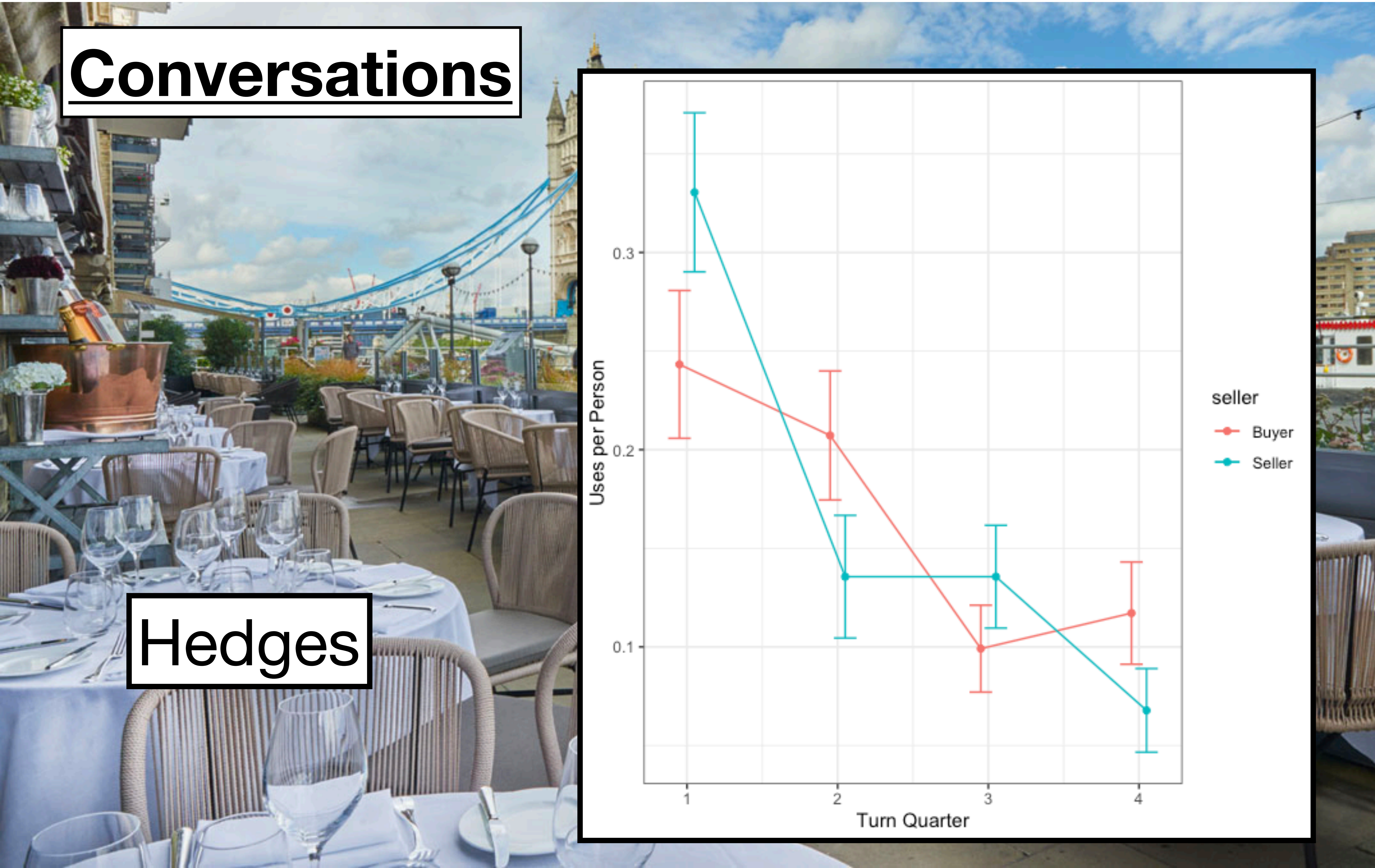
Conversations

Subjectivity



Application in Your Code

Application in Your Code



Other Conversational Cues

Nonverbal Acts

- backchannels (e.g. “yeah”, “uh-huh”)
- pauses
- interruptions
- laughter

(Passonneau & Litman, 1997; Jurafsky & Martin, 2019)

Other Conversational Cues

Nonverbal Acts

- backchannels (e.g. “yeah”, “uh-huh”)
- pauses
- interruptions
- laughter

(Passonneau & Litman, 1997; Jurafsky & Martin, 2019)

Question Types

- mirror *"how about you?"*
- follow-up *"cool, when was that?"*
- switch *"did you hear the news?"*

(Huang, Yeomans, Brooks, Minson & Gino, 2017)

Other Question Types

Type	Question motifs	Answer fragments	Example question-answer pairs
0: Issue update (16,693)	{what, are←taking}, {will←update}	continue→work, met→discuss	Q: **What** steps **are** the Department **taking** to create a system for asylum-seekers? A: We **continue** to **work** with the Department of Education to ensure an equitable [...]
1: Shared concerns (35,954)	{will←take}, {may←urge}	grateful←am, shall←consider	Q: **Will** he **take** steps to support other MPs to employ apprentices? A: I **am grateful** for that suggestion [...]
2: Narrow factual (16,467)	{what←made}, {what←happen, will←happen}	is←considering, have←discussed	Q: **What** representations has the Minister **made** on the future of rural policing [in] Dyfed-Powys? A:The Home Office **is considering** the matter [...]
3: Prompt for comment (16,588)	{what←say, say→to}, {will←tell}	must←say, said→was	Q: **What** has the Prime Minister to **say to** President Reagan for sending troops to Honduras? A: [...] I **must say** that we deplore the reported incursion by Nicaraguan forces [...]
4: Agreement (32,835)	{does←agree, agree→is}, {is→important}	agree→with, agree→completely	Q: **Does** [he] **agree** that one of the best ways to improve the trade balance **is** to continue the Government's strong economic policies? A: I **agree with** my hon. Friend [...]
5: Self promotion (26,351)	{is→aware}, {will←consider}	will←appreciate, am→certain	Q: **Is** my Friend **aware** that members of my parish church are pleased to have received a grant [...] ? A: [My Friend] **will appreciate** the significant performance of parishes up and down the country [...]
6: Concede, accept (31,653)	{will←accept}, {is→not, is→true}	not←accept, not←believe	Q: **Will** [he] **accept** that [the UK exiting the EU] would undermine our security [...]? A: No, I do **not accept** that [...]
7: Condemnatory (21,320)	{can←explain}, {how←justify, can←justify}	knows→well, is→wrong	Q: **Can** the Secretary **explain** why the Government are scrapping child poverty targets? A: The hon. Lady **is wrong** in what she says [...]

(Zhang, Spirling & Danescu-Niculescu-Mizil, 2018)

Other Conversational Cues

Nonverbal Acts

- backchannels (e.g. “yeah”, “uh-huh”)
- pauses
- interruptions
- laughter

(Passonneau & Litman, 1997; Jurafsky & Martin, 2019)

Question Types

- mirror *"how about you?"*
- follow-up *"cool, when was that?"*
- switch *"did you hear the news?"*

(Huang, Yeomans, Brooks, Minson & Gino, 2017)

Emergent Features

Emergent Features

Air Time

Is someone dominating the conversation?

Emergent Features

Air Time

Is someone dominating the conversation?

Variability

Is someone's speaking style consistent?

Emergent Features

Air Time

Is someone dominating the conversation?

Variability

Is someone's speaking style consistent?

Speech rate

How fast is someone talking?

Emergent Features

Air Time

Is someone dominating the conversation?

Variability

Is someone's speaking style consistent?

Speech rate

How fast is someone talking?

Accommodation

Are people repeating one another?

Semantic vs stylistic - unclear...

Estimating Conversation Effects

Causation vs. Prediction vs Detection

Estimating Conversation Effects

Causation vs. Prediction vs Detection

Are person-level data pre-, mid- or post-conversation variables?

Estimating Conversation Effects

Causation vs. Prediction vs Detection

Are person-level data pre-, mid- or post-conversation variables?

How to estimate causality mid-conversation?

Estimating Conversation Effects

Causation vs. Prediction vs Detection

Are person-level data pre-, mid- or post-conversation variables?

How to estimate causality mid-conversation?

Random assignment of instructions

(e.g. Jeong et al., 2019; Yeomans et al., 2020)

Estimating Conversation Effects

Causation vs. Prediction vs Detection

Are person-level data pre-, mid- or post-conversation variables?

How to estimate causality mid-conversation?

Random assignment of instructions
(e.g. Jeong et al., 2019; Yeomans et al., 2020)

Random assignment of partners
(e.g. Zhang, Mullainathan & Danescu-Niculescu-Mizi, 2019)

Estimating Conversation Effects

Causation vs. Prediction vs Detection

Are person-level data pre-, mid- or post-conversation variables?

How to estimate causality mid-conversation?

Random assignment of instructions
(e.g. Jeong et al., 2019; Yeomans et al., 2020)

Random assignment of partners
(e.g. Zhang, Mullainathan & Danescu-Niculescu-Mizi, 2019)

Beginning vs. End of conversation
(e.g. Voigt et al., 2017; Zhang et al., 2019)

Estimating Conversation Effects

Causation vs. Prediction vs Detection

Are person-level data pre-, mid- or post-conversation variables?

How to estimate causality mid-conversation?

Random assignment of instructions
(e.g. Jeong et al., 2019; Yeomans et al., 2020)

Random assignment of partners
(e.g. Zhang, Mullainathan & Danescu-Niculescu-Mizi, 2019)

Beginning vs. End of conversation
(e.g. Voigt et al., 2017; Zhang et al., 2019)

Otherwise: Control for observables, Focus on cross-person effects

Guidelines for Conversation Research

Guidelines for Conversation Research

NLP will improve (not replace) your reading

Guidelines for Conversation Research

NLP will improve (not replace) your reading

If humans can't do it, neither can an algorithm

Guidelines for Conversation Research

NLP will improve (not replace) your reading

If humans can't do it, neither can an algorithm

Interpretability is essential for credibility

Guidelines for Conversation Research

NLP will improve (not replace) your reading

If humans can't do it, neither can an algorithm

Interpretability is essential for credibility

Expect domain specificity as a rule

Guidelines for Conversation Research

NLP will improve (not replace) your reading

If humans can't do it, neither can an algorithm

Interpretability is essential for credibility

Expect domain specificity as a rule

Consider goals of speakers, structure, and time

Guidelines for Conversation Research

NLP will improve (not replace) your reading

If humans can't do it, neither can an algorithm

Interpretability is essential for credibility

Expect domain specificity as a rule

Consider goals of speakers, structure, and time

Conversation is different (harder) than descriptions

Let's keep the conversation going