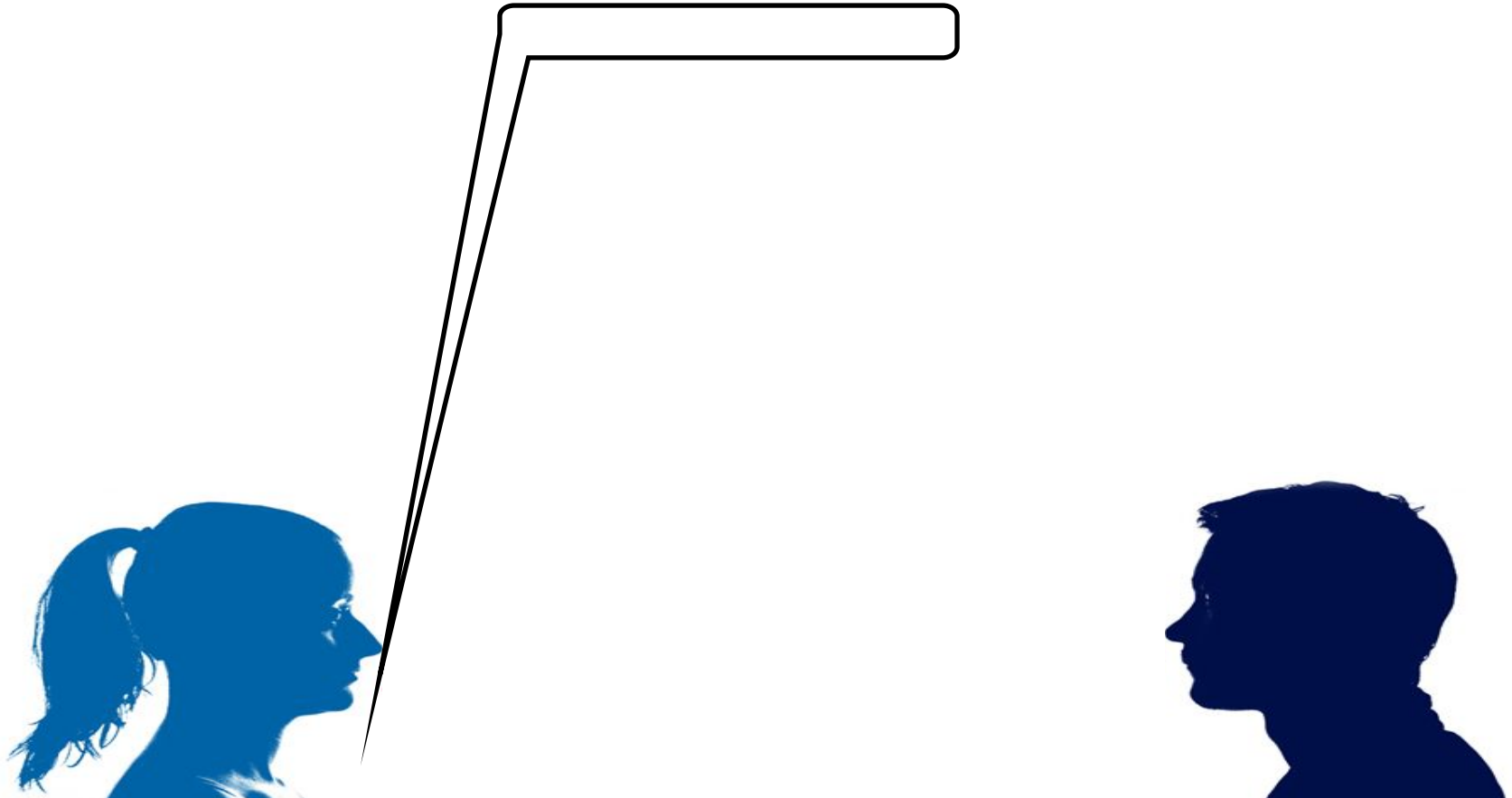
A User's Guide to Conversation Research

IACM 2021

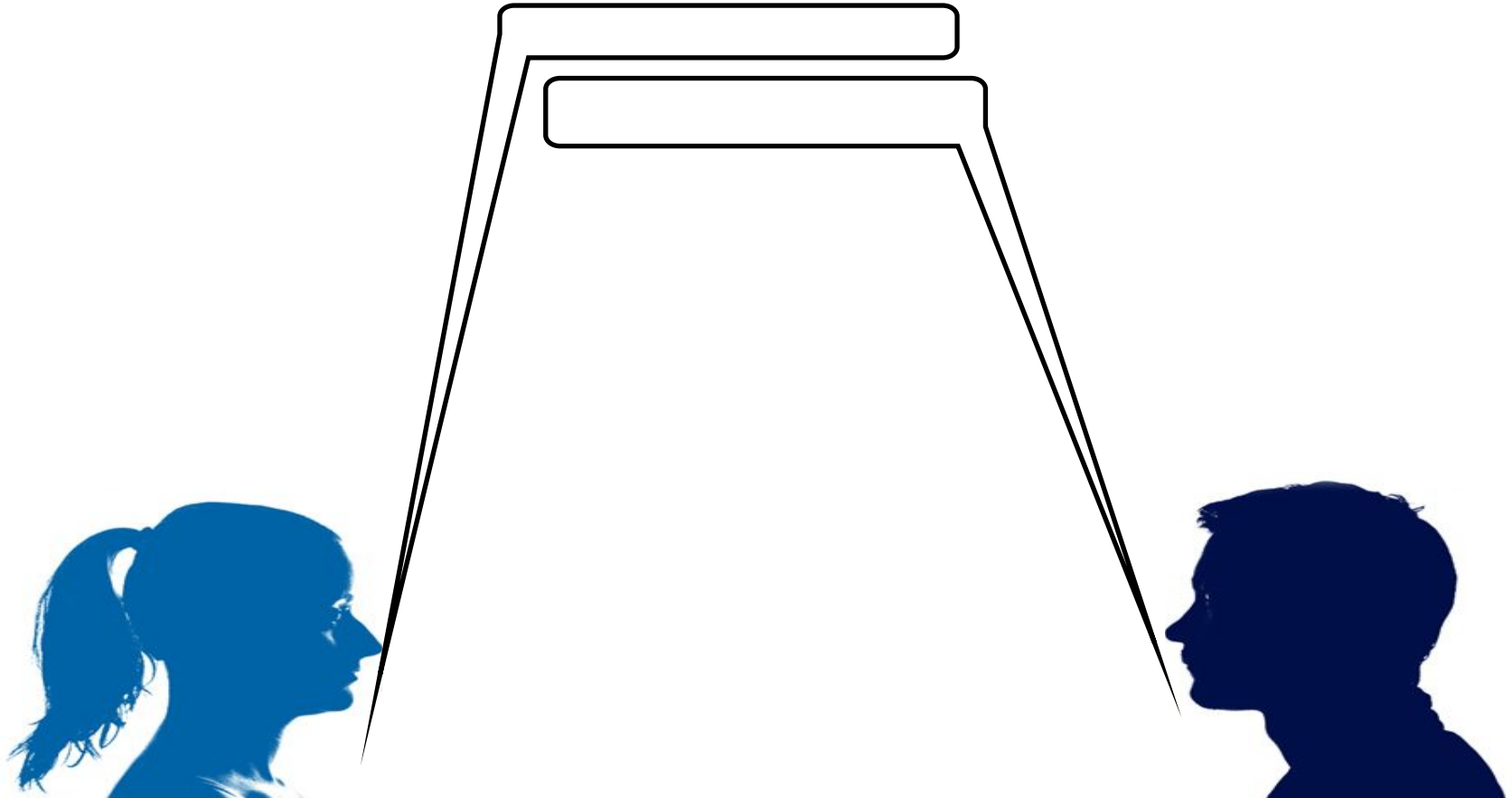
Conversations are Everywhere

Conversations

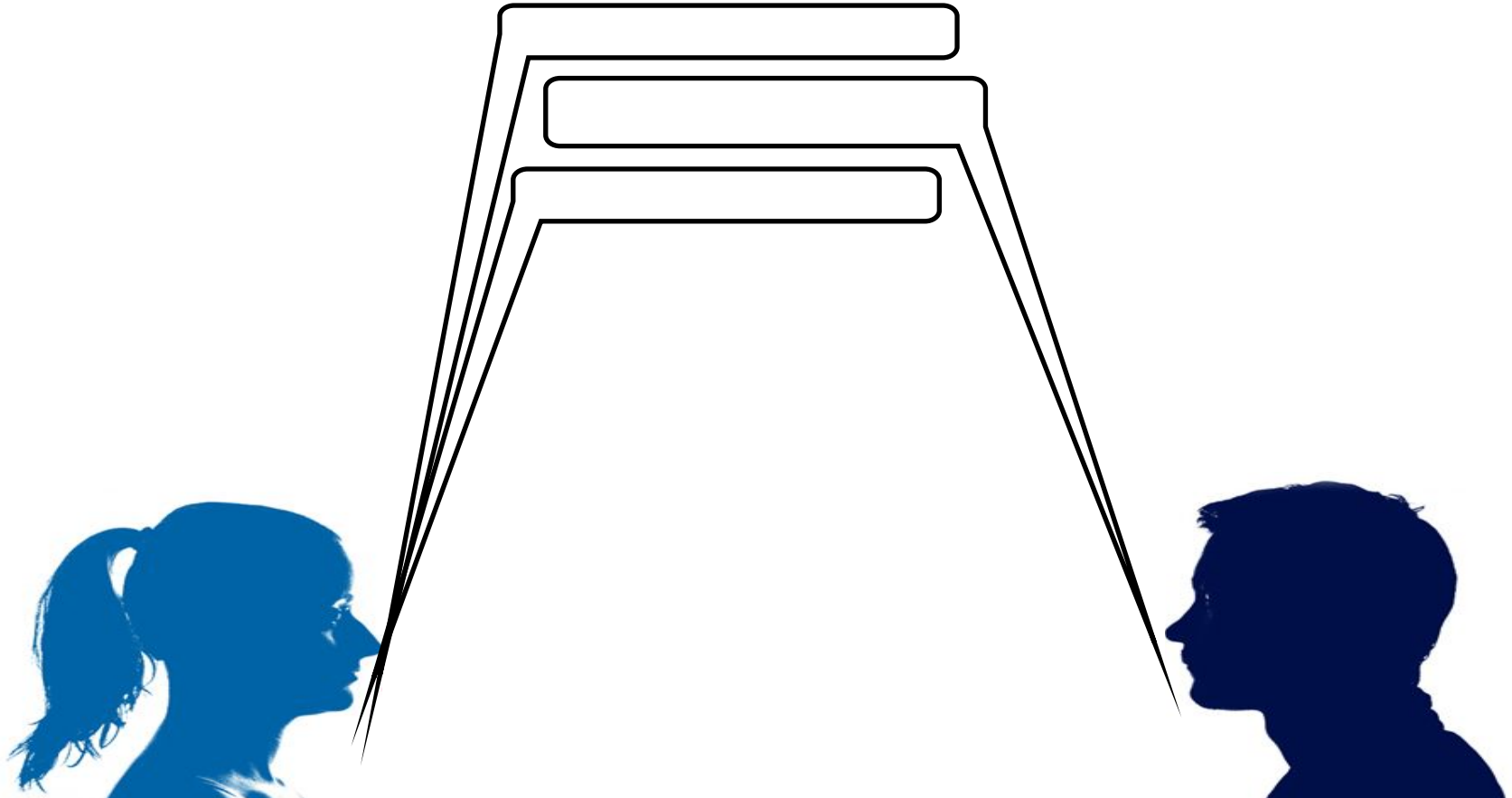
Conversations



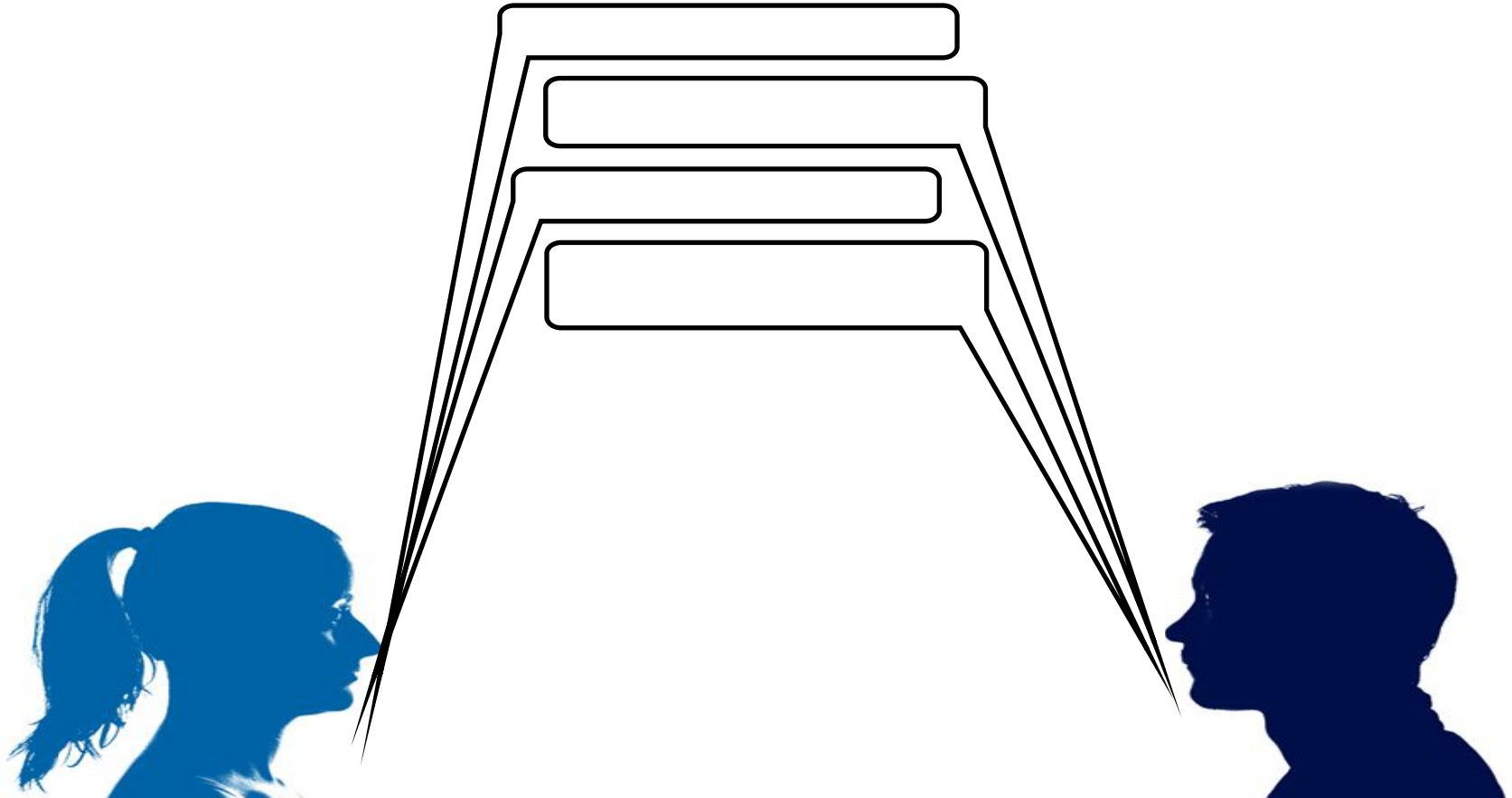
Conversations



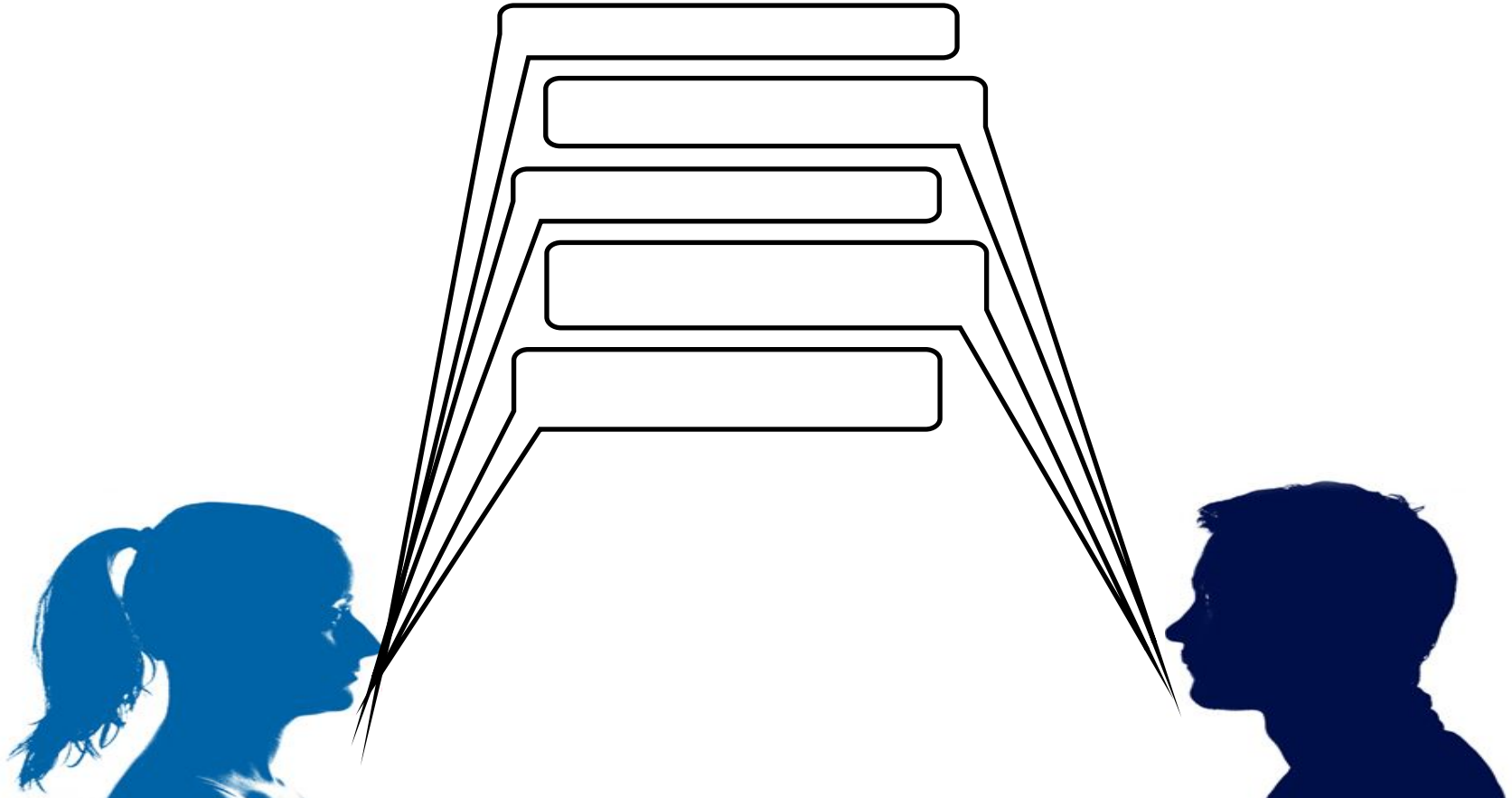
Conversations



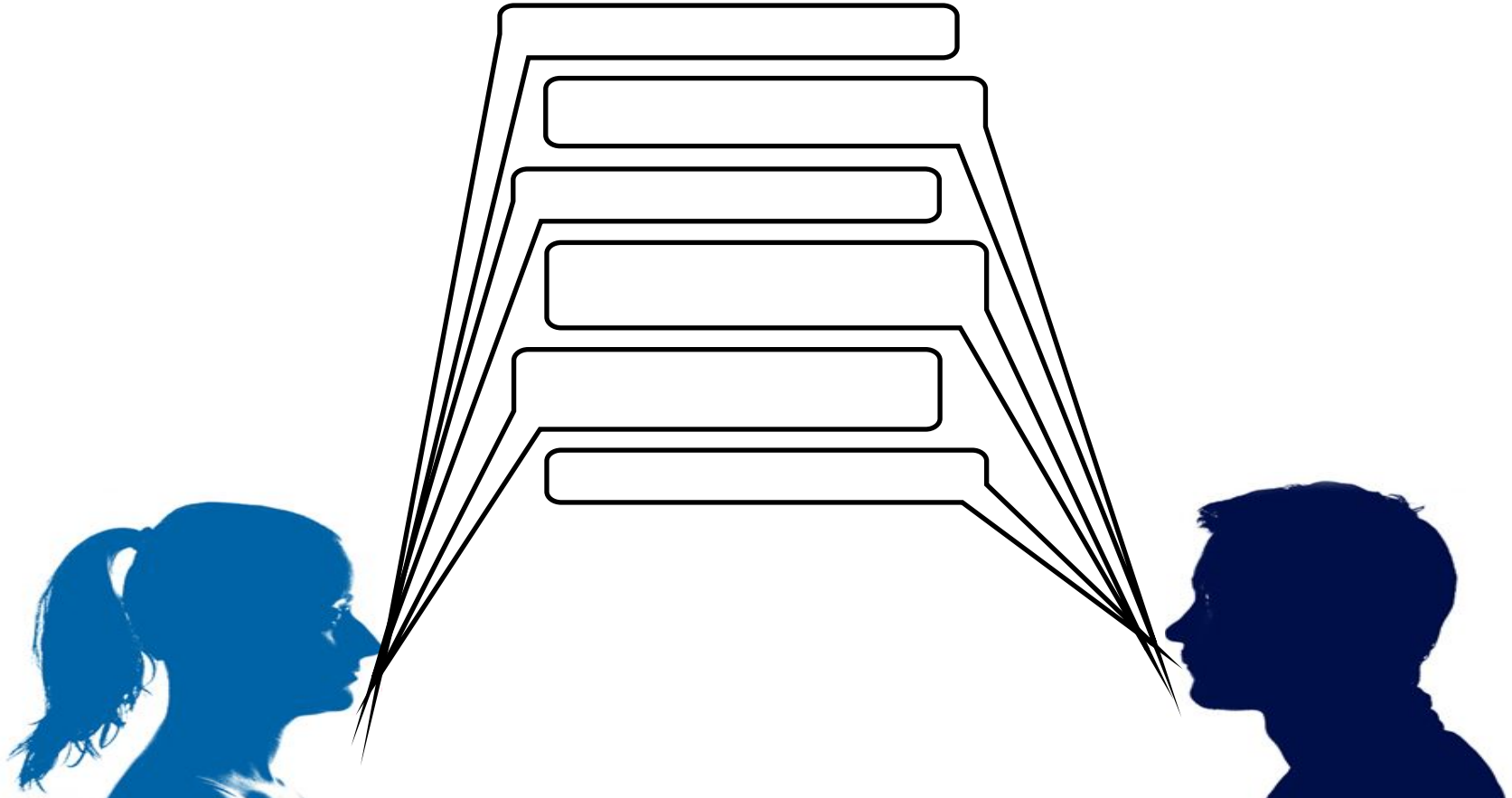
Conversations



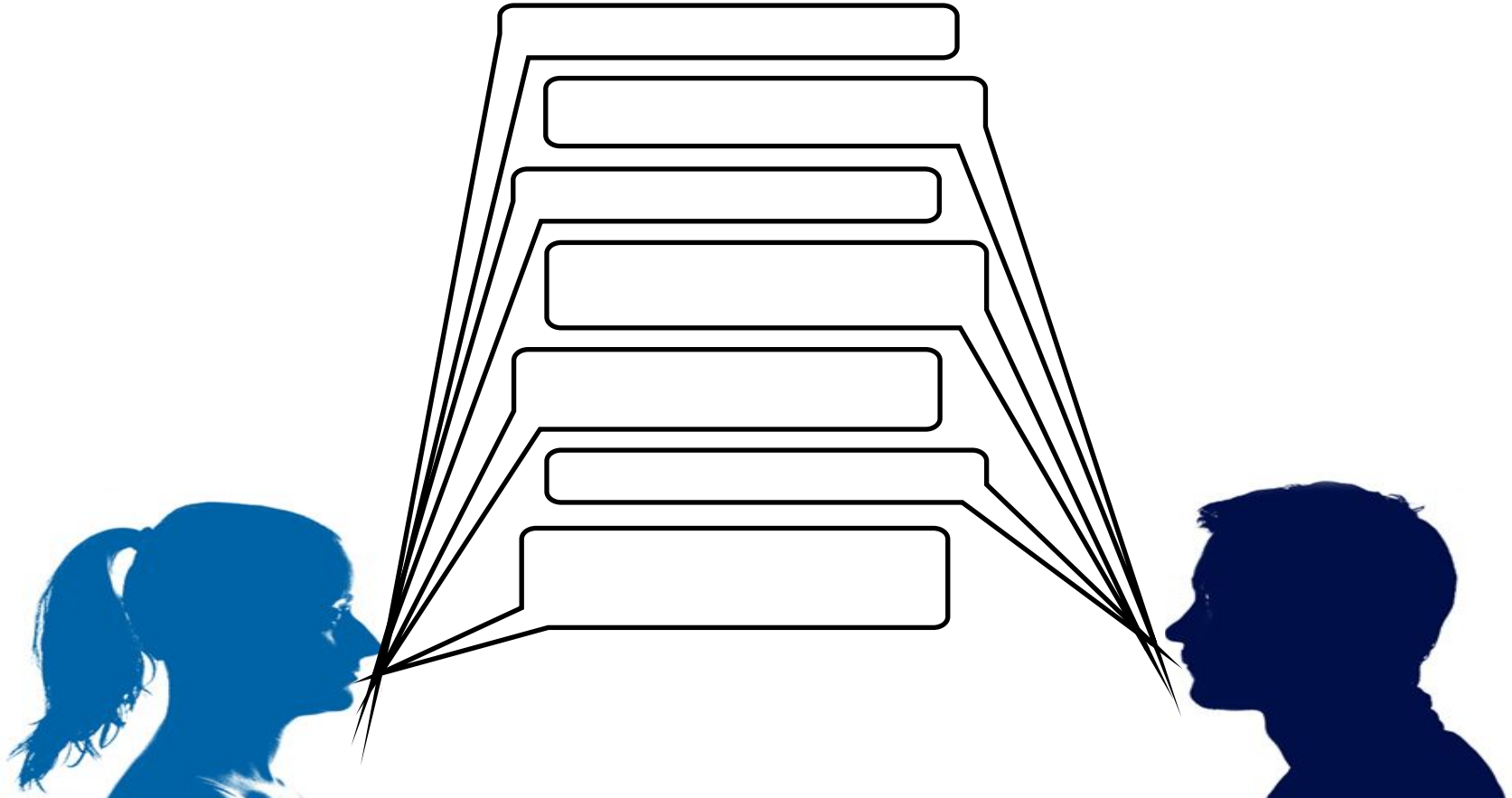
Conversations



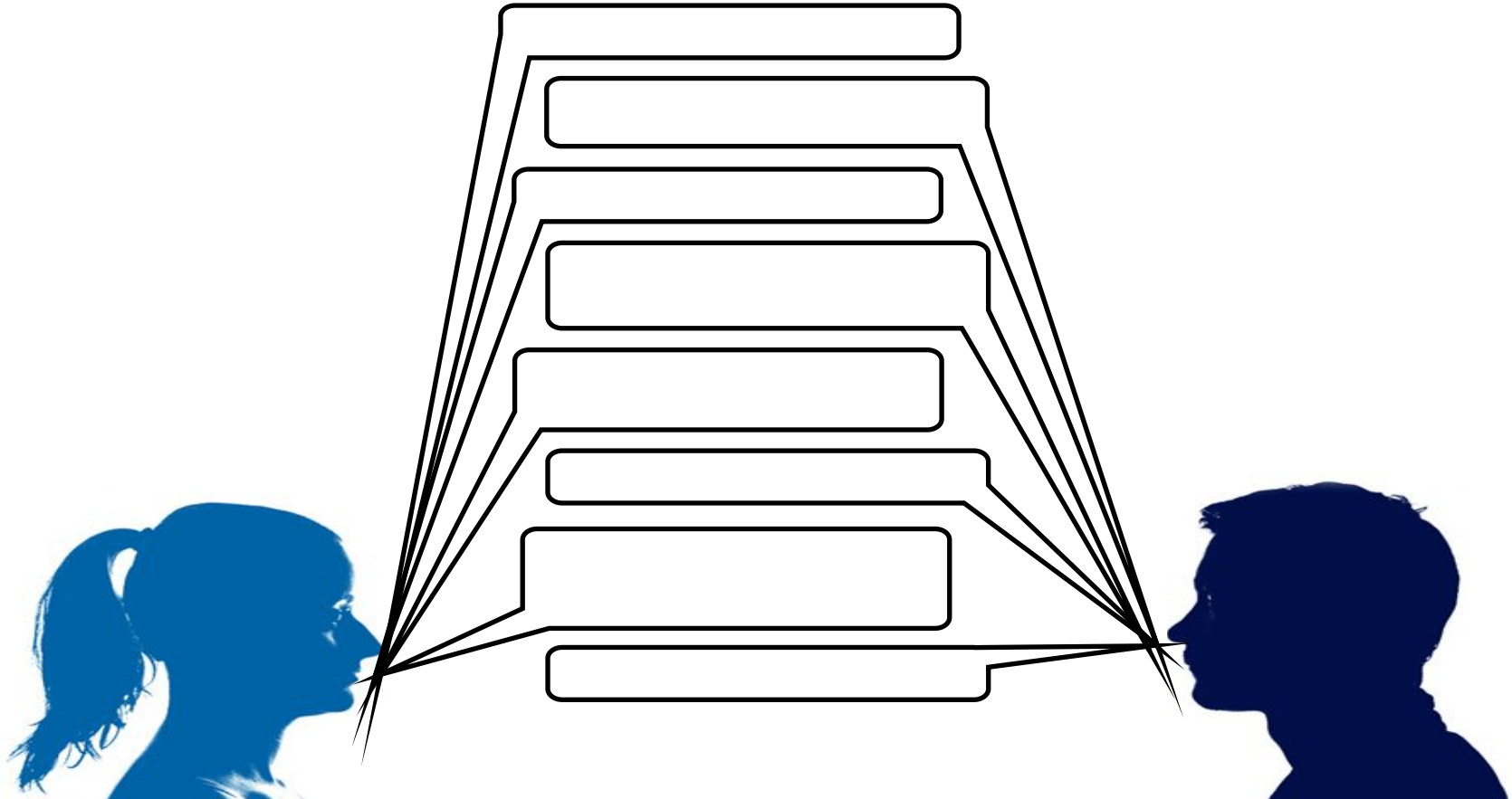
Conversations



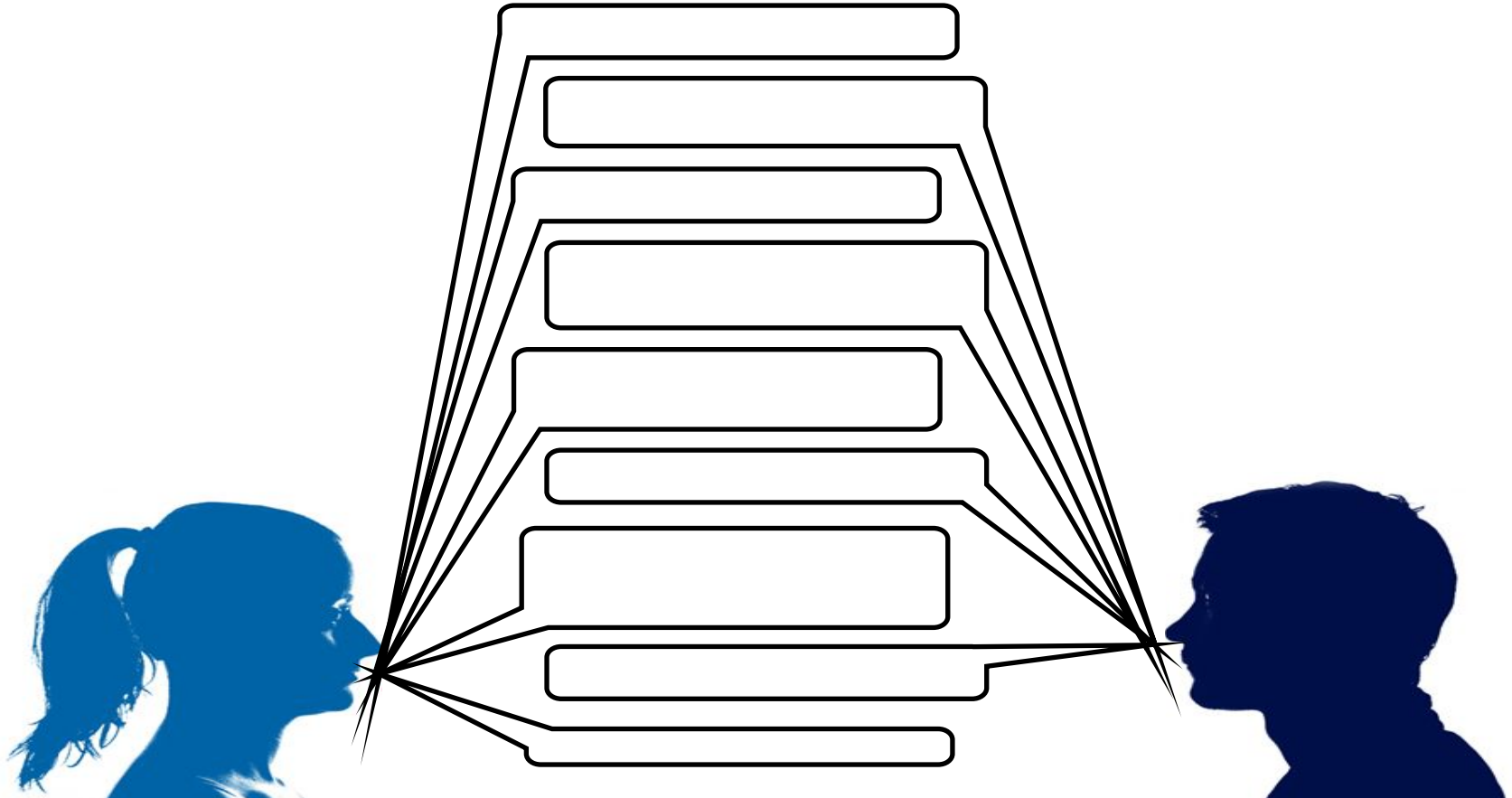
Conversations



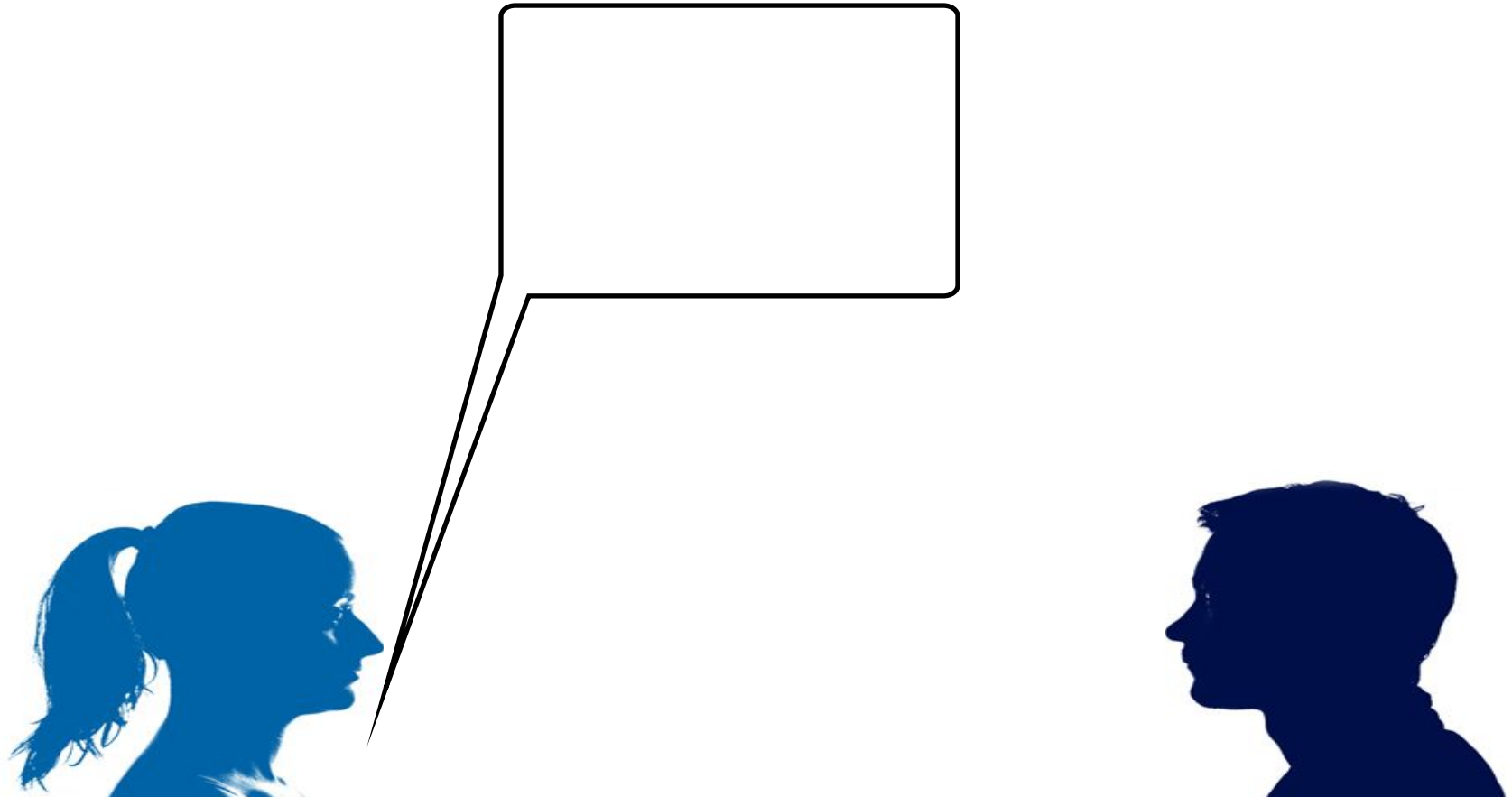
Conversations



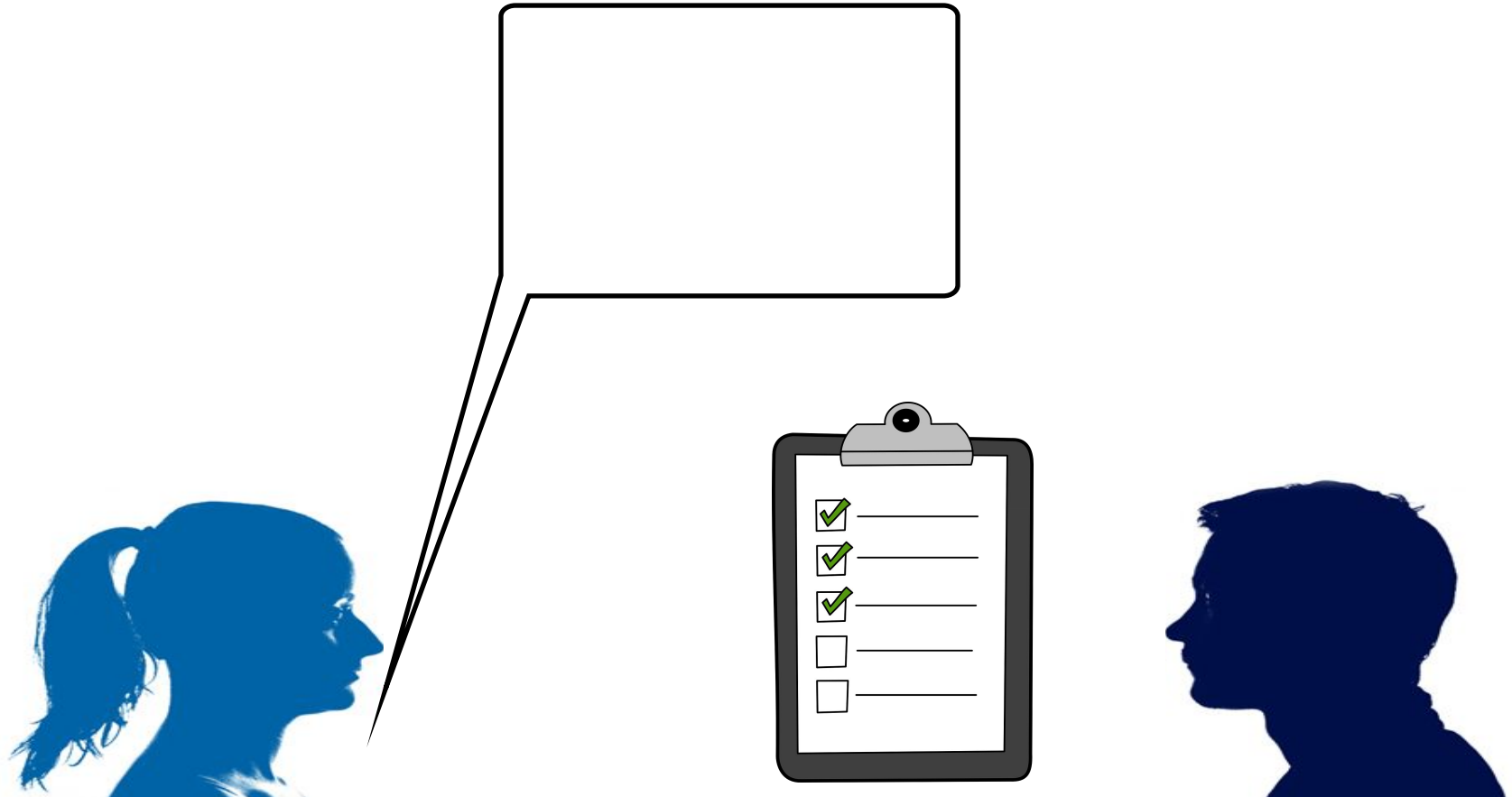
Conversations



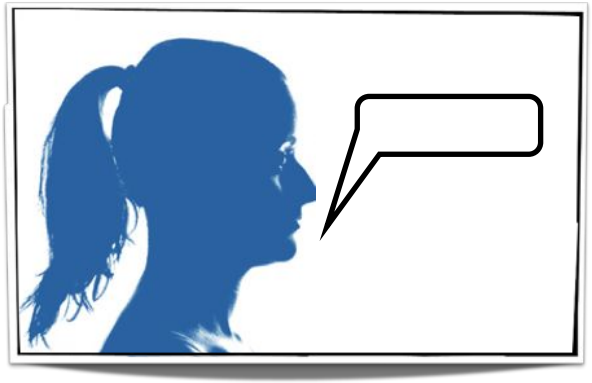
Conversations, Indirectly

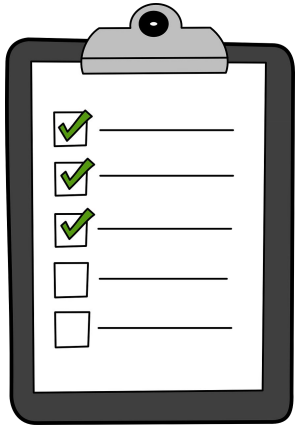


Conversations, Indirectly



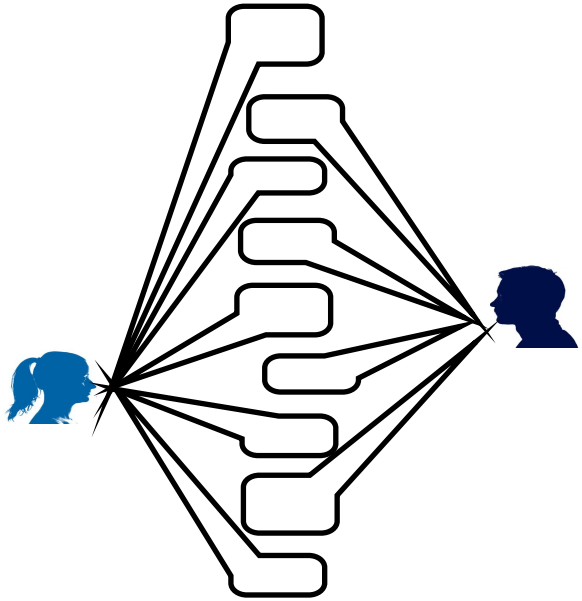
Conversations, Indirectly



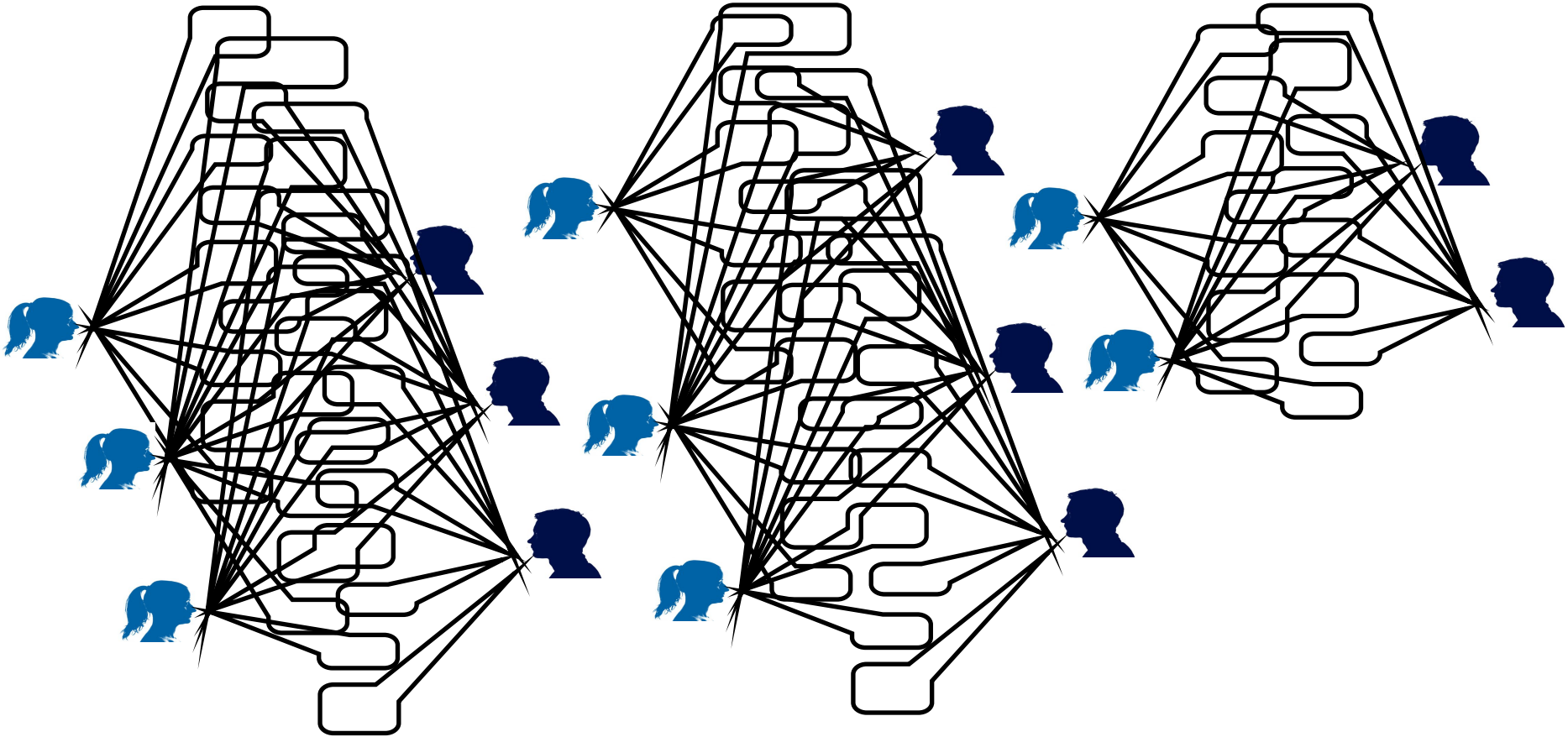


Conversations, Indirectly

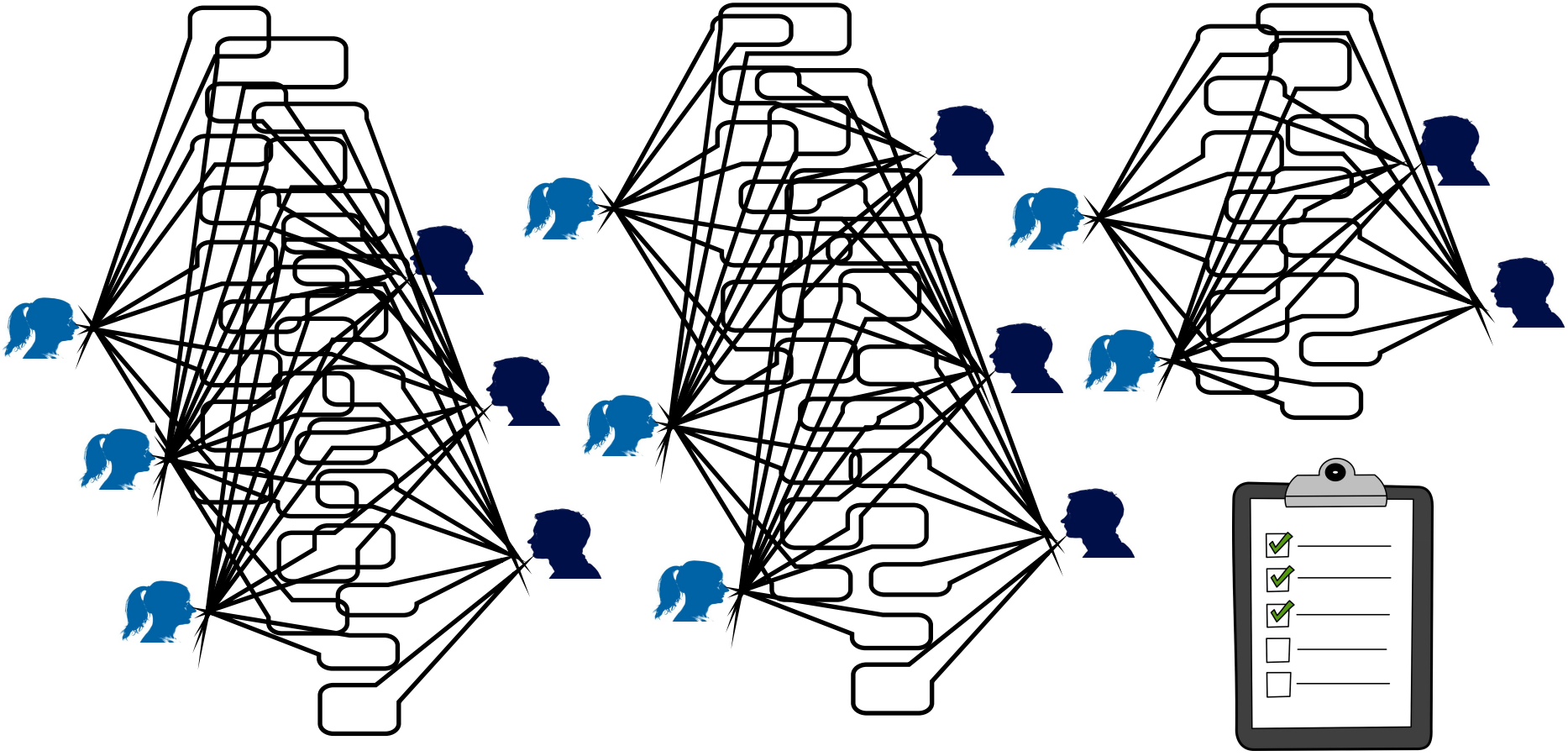
Conversations, Indirectly



Conversations, Indirectly



Conversations, Indirectly



Conversations

The fundamental behavior of all human social interaction

Conversations

The fundamental behavior of all human social interaction

So why do we so rarely study the behavior itself?

Conversations

The fundamental behavior of all human social interaction

So why do we so rarely study the behavior itself?

- Difficult to acquire/clean data

Conversations

The fundamental behavior of all human social interaction

So why do we so rarely study the behavior itself?

- Difficult to acquire/clean data
- Difficult to analyse (many measures, goals, contexts)

Conversations

The fundamental behavior of all human social interaction

So why do we so rarely study the behavior itself?

- Difficult to acquire/clean data
- Difficult to analyse (many measures, goals, contexts)

Goals of this workshop:

Improve science by making conversation accessible to researchers

Conversations

The fundamental behavior of all human social interaction

So why do we so rarely study the behavior itself?

- Difficult to acquire/clean data
- Difficult to analyse (many measures, goals, contexts)

Goals of this workshop:

Improve science by making conversation accessible to researchers

Build a community of inquiry around conversational behavior

A Quick Pointer to Example Data and Code (for later)

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

https://github.com/myeomans/convoResearch

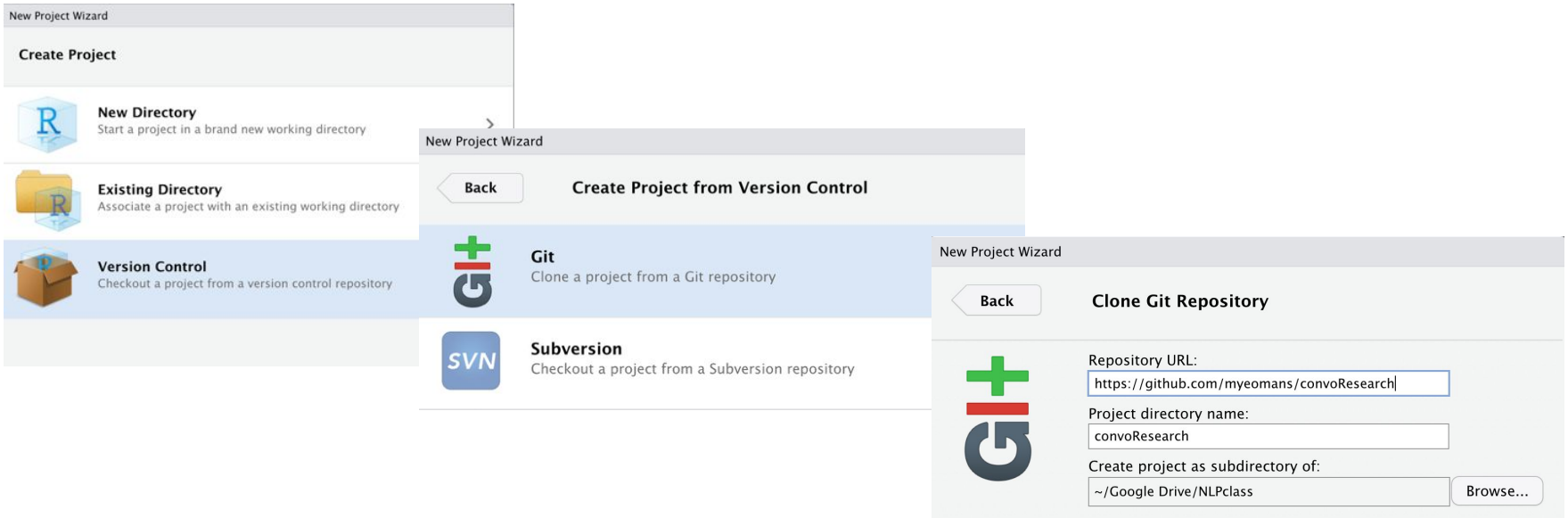
Download everything at once by **creating an RStudio Project**:

A Quick Pointer to Example Data and Code (for later)

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

https://github.com/myeomans/convoResearch

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



Agenda

Part 1 - Conversation Technology (5-15 min)

Presented by Nicole Abi-Esber (nabiesber@hbs.edu)

Part 2 - Humans in the Loop: Transcriptions and Annotations (15-25 min)

Presented by Katie Boland (kboland@hbs.edu)

Part 3 - The Structure of Conversation Data (25-35 min)

Presented by Hanne Collins (hcollins@hbs.edu)

Part 4 - Applied Example: Measuring politeness in social interactions (35-80 min)

Presented by Michael Yeomans (myeomans@imperial.ac.uk)

And beyond... **A User's Guide to Conversation Research (working paper)**

Conversation technology

Conversation technology

	Already exist	**Generated in lab**

Conversation technology

	Already exist	**Generated in lab**
Text		
Audio		

Conversation technology

	Already exist	Generated in lab
Text		
Audio		

Conversation technology

	Already exist	Generated in lab
Text		
Audio		

Conversation technology

	Already exist	Generated in lab
Text		qualtrics. chatplat
Audio		

Conversation technology

	Already exist	Generated in lab
Text		qualtrics. chatplat
Audio		zoom Otter.ai

Conversation technology

Text conversations that already exist:

Conversation technology

Text conversations that already exist:

- Slack, emails
  - (company-owned)
- Online forums
  - (publicly-available Reddit, Twitter, Wikipedia)
- Chatbots
  - Facebook, customer service

Conversation technology

4 7 63

MICHAEL JACKSON @mjackson · 2d
What's your technique for the icing? Do you make a mold?
1

Lynn Fisher @lynnandtonic · 2d
I used a 3D printed stamp cookie cutter before baking and then I traced the imprint with icing from a piping bag. Worked out pretty well!
1 5

MICHAEL JACKSON @mjackson · 2d
I'd say so. The sharp edges on that monomyth logo are so crisp for icing fron a bag!
1 2

Lynn Fisher @lynnandtonic · 2d
I use two consistencies of icing to get that effect (outline and fill). More work but worth it!
1 2

Brie
@Sktch_ComedyFan

@patel_232 hilarious set last night. You killed it!

9:20 · 4/2/19 · Twitter Web Client

1 Retweet **3** Likes

patel @patel_232 · 3m
Mentioned
Replying to @Sktch_ComedyFan
Aww shucks

Brie @Sktch_ComedyFan · 1m
Author
Can't wait to see your next show - I ar

Cleve @all_the_sportz · 2h
Ha, no way, I never listen to anything made after the 70's. Classic rock all the way.

Holli @domingo_124 · 2h
I'm with you, my record collection is all pre 80's

Jasi @k9lover85 · 1h
Back when music had a soul.

Katie O. @kay_tee_oh · 1h
But what about the 90s??

Suzie @sweetsuzzzie · 4h
This year's been good but indie rock from

GIF

Conversation technology

	Already exist	Generated in lab
Text		qualtrics. chatplat
Audio		zoom Otter.ai

Conversation technology

Audio conversations that are already transcribed:

Conversation technology

Audio conversations that are already transcribed:

- Supreme Court Hearings (Danescu-Niculescu-Mizil, Lee, Pang, & Kleinberg, 2012)
- FOMC meetings (Hansen, McMahon, & Prat, 2017)
- CEO Quarterly earnings calls (Crawford Camiciottoli, 2010)
- Sports post-game interviews (Fu, Danescu-Niculescu-Mizil, & Lee, 2016)
- Podcast interviews

Corporate Participants

Sherry Bahrambeygui - Chief Executive Officer

Michael McCleary - Chief Financial Officer

Conference Call Participants

Jon Braatz - Kansas City Capital

Operator

Good morning/afternoon everyone, and welcome to PriceSmart Incorporated Earnings Release Conference Call for the Third Quarter of Fiscal Year 2021, which ended on May 31, 2021. After remarks from our company's representatives, Sherry Bahrambeygui, Chief Executive Officer; and Michael McCleary, Chief Financial Officer, you will be given an opportunity to ask questions as time permits.

As a reminder, this conference call is limited to one hour and is being recorded today, Friday, July 9, 2021. A digital replay will be available following the conclusion of today's call through July 16, 2021, by dialing 1 (877) 344-7529 for domestic callers or 1 (412) 317-0088 for international callers, and by entering replay access code of 10156316.

For opening remarks, I would like to turn the call over to PriceSmart's Chief Financial Officer, Michael McCleary. Please proceed sir.

Michael McCleary

Thank you and welcome to the PriceSmart earnings call for the third quarter of fiscal year 2021. We will be discussing the information that we provided in our earnings press release in our 10-Q, which were both released yesterday afternoon, July 08, 2021. You can find these documents on our Investor Relations website at investors.pricesmart.com, where you can also sign up for email alerts.

As a reminder, all statements made on this conference call, other than

Coach, could you take us through the fourth-and-1 play with RB Mark Ingram II and the surprise there maybe for you and how that developed?

"Right. That's something that we've been practicing for many weeks now, something that the offensive coaches led by Greg Roman came up with. We've gotten a lot of reps on it, and of course Mark executed it really well as far as setting up the scheme itself in terms of his angles. And then the offensive line and tight ends just blocked it great. It's a critical play. It's something we had been preparing for that kind of situation. I give Greg and the offensive coaches and the players, the guys who executed it, all the credit. They're the ones that made it happen, and that was big. It was a big play in the game. It was the turning point in the game. Gave us a little breathing space and I'm proud of them for that."

Did you feel like this victory was sort of a testament to the fact that your team can win in a lot of different ways?

"You know, I haven't thought about it so much in terms of – we're just trying to win. We're trying to figure out every way we can to win and whatever tools are at your disposal, some things were working that we thought might and there was a lot of things that weren't working in this game because they took them away. I thought Anthony Weaver did a really nice job defensively with their game plan. He had a really creative plan and some different things for us and their guys were playing very disciplined, tackling well. Those were challenges that we were faced with. Ultimately overcame them but, it was not easy by any stretch."

Conversation technology

	Already exist	Generated in lab
Text		qualtrics. chatplat
Audio		zoom Otter.ai

Conversation technology:

Our research also generates text data

Conversation technology:

Our research also generates text data

- Chatplat
- iDecisionGames
- Open-ended qualtrics responses

Conversation technology

Lastly, the lab experiments we run are rich sources of live conversation data

Conversation technology

Lastly, the lab experiments we run are rich sources of live conversation data

- **Otter.ai:** auto-transcriptions (powered by Otter)
- **Macro.io:** automatic airtime analysis
- **Zoom:** auto-transcription, timestamps

Otter.ai

those

Speaker 1 2:07

three steps serve candidate should I can go first of us so I think a good things for my guy is that he has two years of work experience and information intelligence, and that he's trained in navigation skills. And then he can fly a Russian fighter plane. The bad stuff is like his head up. But out in the past, he often postpones things and he stopped those under high pressure.

Speaker 3 2:33

But any differences? I have just saying bad things too. There's a high pressure isn't burnout in the past. But another good thing that he has is He can speak five languages

2:48

that are solid.

Speaker 2 2:50

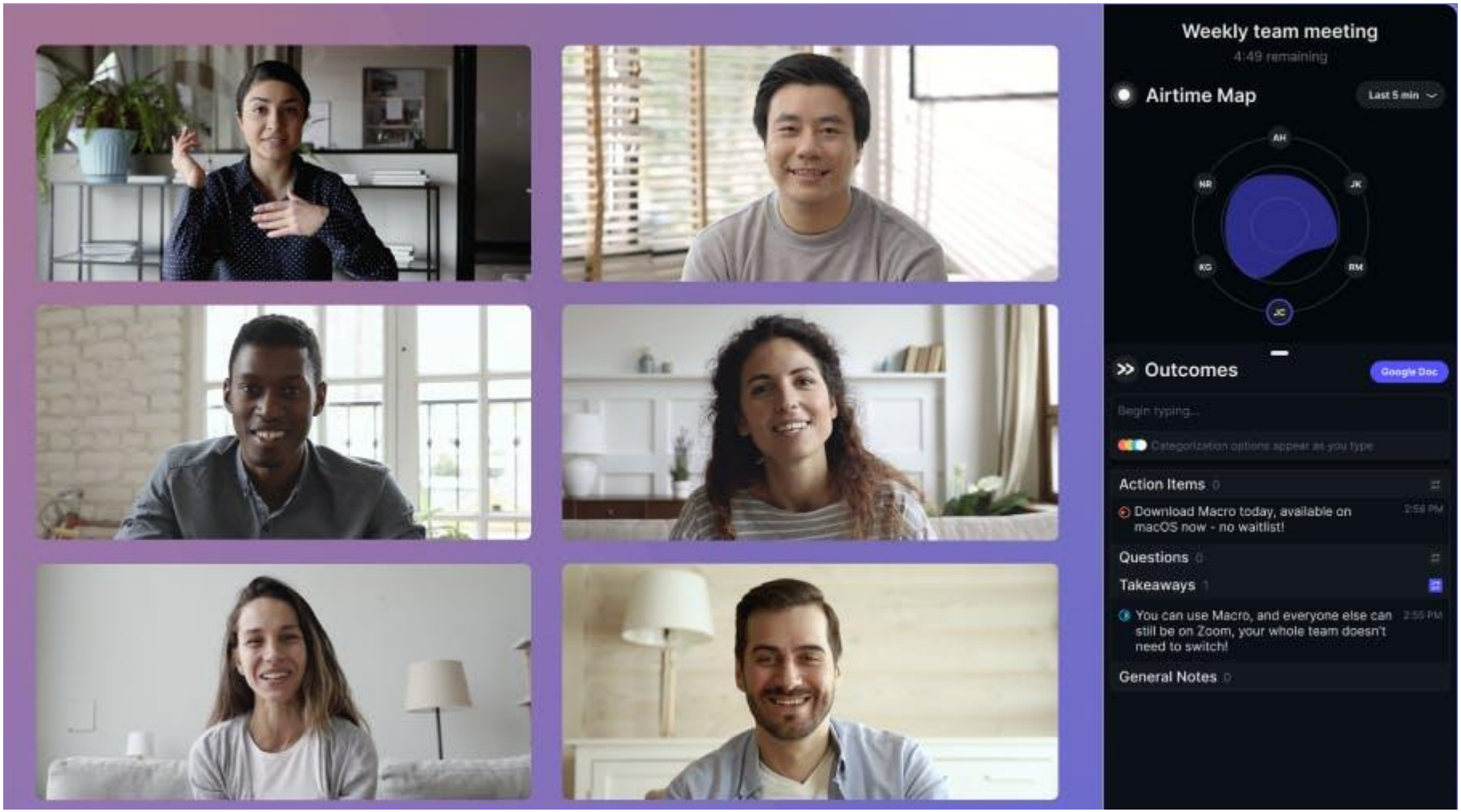
I have the same negatives as well. My positives, the only major one or the best one is probably good interrogation skills.

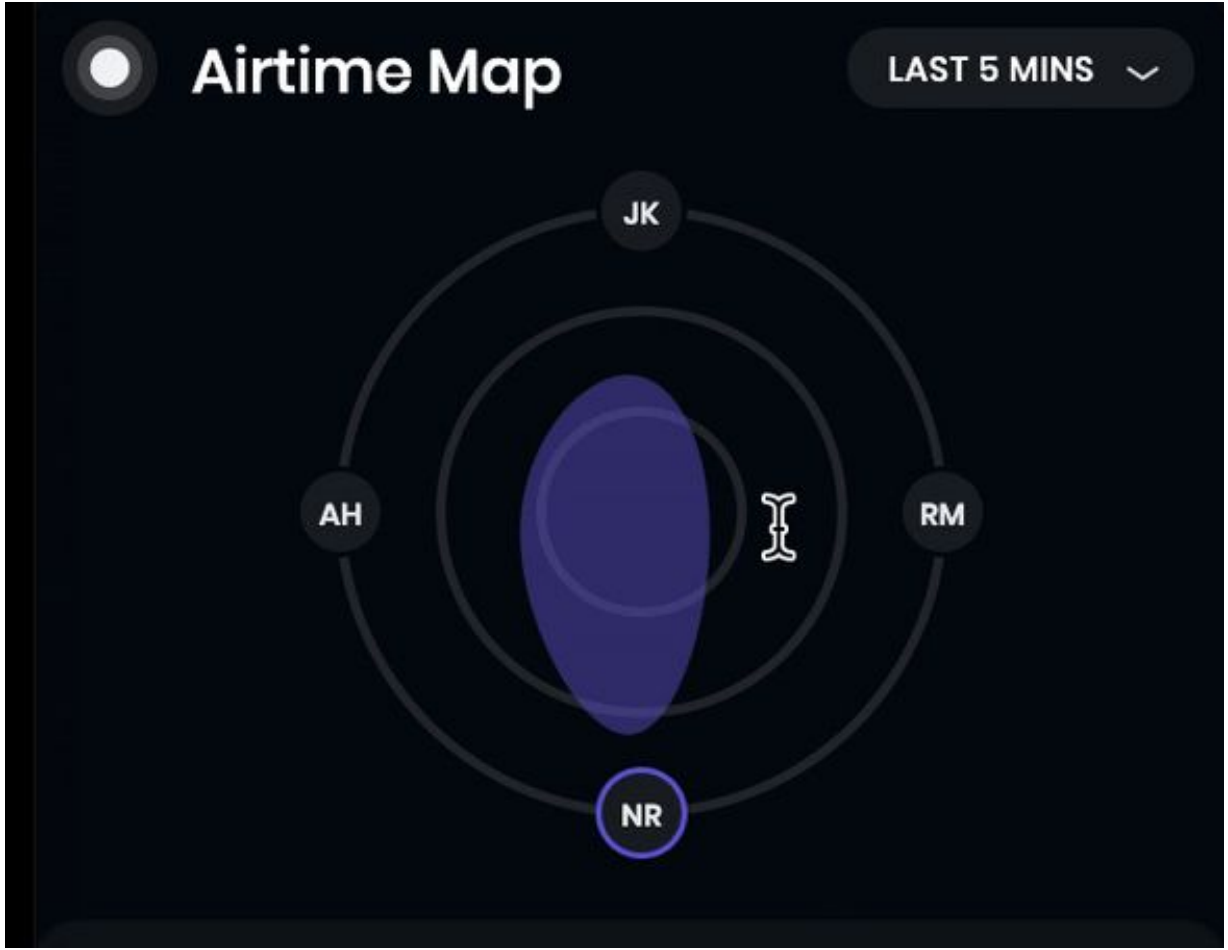
Speaker 2 3:06

I think in my opinion, I feel like this sweater is under high pressure to deal with such a cable.

All changes saved Undo Redo

Macro.io


Weekly team meeting
4:49 remaining
Airtime Map
Last 5 min
AH
JK
NR
RM
KG
JC
Outcomes
Google Doc
Begin typing...
Categorization options appear as you type
Action Items 0
Download Macro today, available on macOS now - no waitlist!
2:58 PM
Questions 0
Takeaways 1
You can use Macro, and everyone else can still be on Zoom, your whole team doesn't need to switch!
2:59 PM
General Notes 0


Airtime Map
LAST 5 MINS
JK
AH
RM
NR

From: Macro Meetings <meetings@macro.io>
Date: Wednesday, April 28, 2021 at 5:47 PM
To: "Boland, Katelynn" <kboland@hbs.edu>
Subject: Meeting Summary - Katie Boland's Personal Meeting Room 04/28/2021

Katie Boland's Personal Meeting Room

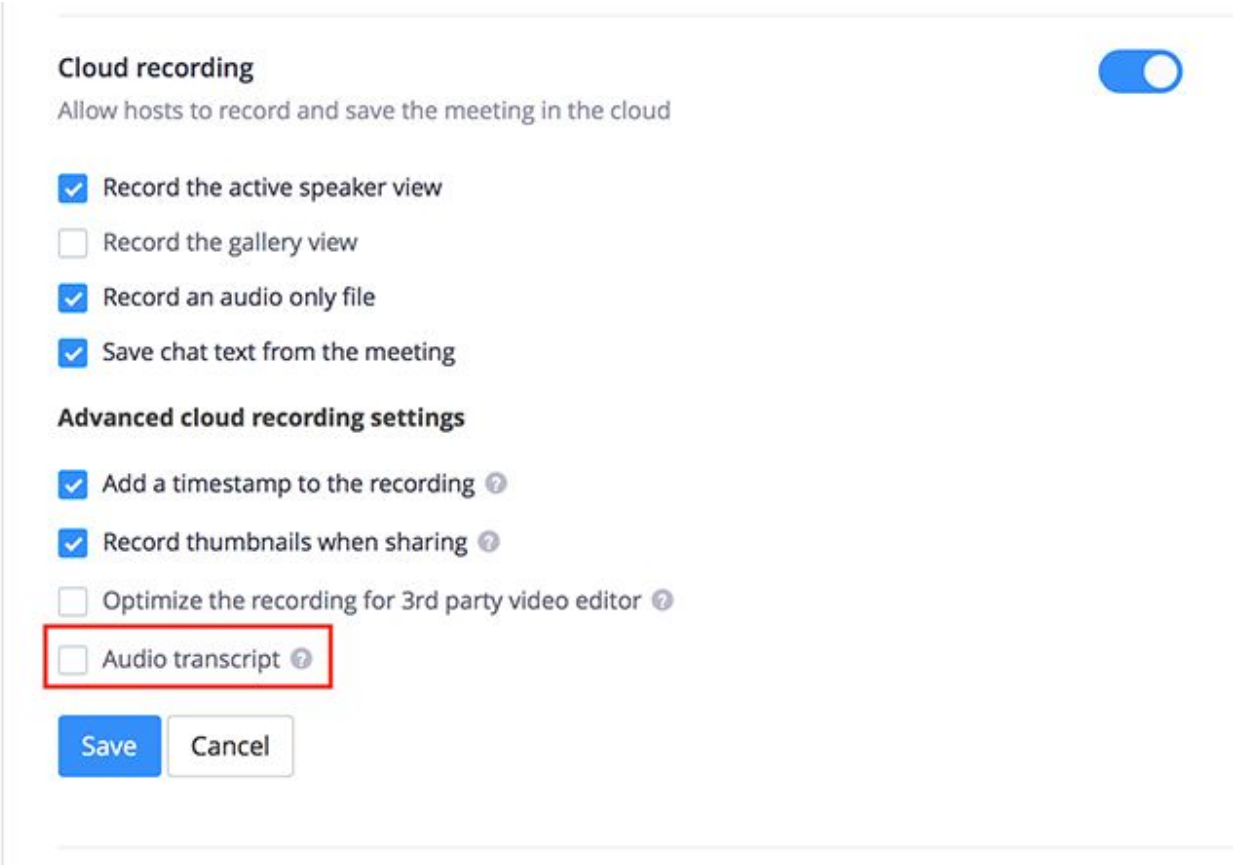
04/28/2021

All-meeting Airtime Breakdown:

- Katie: 57%
- Mittal: 20%
- jessicacarlton: 9%
- Mary Boehmer: 9%
- Aditya Nahar: 5%

Zoom auto-transcriptions

Zoom auto-transcriptions



97
00:24:08.520 --> 00:24:16.770
Delivery Manager: And if it takes off even more, we could switch to an order per two weeks or two, a month, depending on how well it takes off, and I would just be something that we see happen all the time.

98
00:24:18.570 --> 00:24:25.320
Central Warehouse Manager: yeah even me Oh, I would support they just insane model seems.

99
00:24:26.670 --> 00:24:44.670
Central Warehouse Manager: Our big aim is to minimize the holding costs and by increasing orders each and every time we will increase the holding cost and for a small business to have a really big holding costs seem to be.

100
00:24:46.800 --> 00:24:56.640
Central Warehouse Manager: at a disadvantage to them, and also to use the information that we have, like the perfect information.

101
00:24:58.350 --> 00:25:03.390
Central Warehouse Manager: That we provided in ingesting data will be the perfect thing.

102
00:25:07.770 --> 00:25:13.380
Retail Manager: How do we avoid kind of lagging behind demand without having an inventory.

103
00:25:16.890 --> 00:25:30.630
Retail Manager: say it sells out say it sells out, you know we have you know some number of boxes for customers to have and on the first day that they're available and we make people aware that they're all gone by noon.

Converting .vtt to .csv

https://github.com/seankross/vtt/

README.md

vtt

Read `.vtt` files into a data frame.

Installation

```
remotes::install_github("seankross/vtt")
```

Example

```
library(tidyverse)
library(vtt)

setwd("OneDrive_2_6-16-2021")

transcripts <- data.frame(File = list.files(pattern = ".vtt")) %>%
  mutate(DF = map(File, read_vtt)) %>%
  unnest(cols = "DF") %>%
  extract_speaker()

write_csv(transcripts, "transcription_vttfiles.csv")
```

Turn	Start	End	Person	Text
96	24:00.3	00:24:08.13	Delivery Manager	i'd say like an example is instead of order per day, and if it takes off, then we switch to an order per five days or order per week.
97	24:08.5	00:24:16.77	Delivery Manager	And if it takes off even more, we could switch to an order per two weeks or two, a month, depending on how well it takes off, and I would just be something that we see happen all the time.
98	24:18.6	00:24:25.32	Warehouse Manager	yeah even me Oh, I would support they just insane model seems.
99	24:26.7	00:24:44.67	Central Warehouse Manager	Our big aim is to minimize the holding costs and by increasing orders each and every time we will increase the holding cost and for a small business to have a really big holding costs seem to be.
100	24:46.8	00:24:56.64	Warehouse Manager	at a disadvantage to them, and also to use the information that we have, like the perfect information.
101	24:58.4	00:25:03.39	Warehouse Manager	That we provided in ingesting data will be the perfect thing.
102	25:07.8	00:25:13.38	Retail Manager	How do we avoid kind of lagging behind demand without having an inventory.
103	25:16.9	00:25:30.63	Retail Manager	say it sells out say it sells out, you know we have you know some number of boxes for customers to have and on the first day that they're available and we make people aware that they're all gone by noon.

nabiesber@hbs.edu

Agenda

Part 1 - Conversation Technology (5-15 min)

Presented by Nicole Abi-Esber (nabiesber@hbs.edu)

Part 2 - Humans in the Loop: Transcriptions and Annotations (15-25 min)

Presented by Katie Boland (kboland@hbs.edu)

Part 3 - The Structure of Conversation Data (25-35 min)

Presented by Hanne Collins (hcollins@hbs.edu)

Part 4 - Applied Example: Measuring politeness in social interactions (35-80 min)

Presented by Michael Yeomans (myeomans@imperial.ac.uk)

And beyond... **A User's Guide to Conversation Research (working paper)**

Humans in the Loop

International Conference on Natural Language and Speech Processing, ICNLSP 2015

Automatic Speech Recognition Errors Detection and Correction: A Review

Rahhal Errattahi[a,*], Asmaa El Hannani[a], Hassan Ouahmane[a]

[a]*Laboratory of Information Technologies, National School of Applied Sciences, University of Chouaib Doukkali, El Jadida - Morocco*

Abstract

Even though Automatic Speech Recognition (ASR) has matured to the point of commercial applications, high error rate in some speech recognition domains remain as one of the main impediment factors to the wide adoption of speech technology, and especially for continuous large vocabulary speech recognition applications. The persistent presence of ASR errors have intensified the need to find alternative techniques to automatically detect and correct such errors. The correction of the transcription errors is very crucial not only to improve the speech recognition accuracy, but also to avoid the propagation of the errors to the subsequent language processing modules such as machine translation. In this paper, basic principles of ASR evaluation are first summarized, and then the state of the current ASR errors detection and correction research is reviewed. We focus on emerging techniques using word error rate metric.

Keywords: Automatic Speech Recognition; ASR Error Detection; ASR Error Correction; ASR evaluation;

Humans in the Loop

1. Prepare
2. Record
3. Transcribe
4. Check

Humans in the Loop: Prepare

Audio Quality: Test the audio quality in the room(s) you will use to collect data

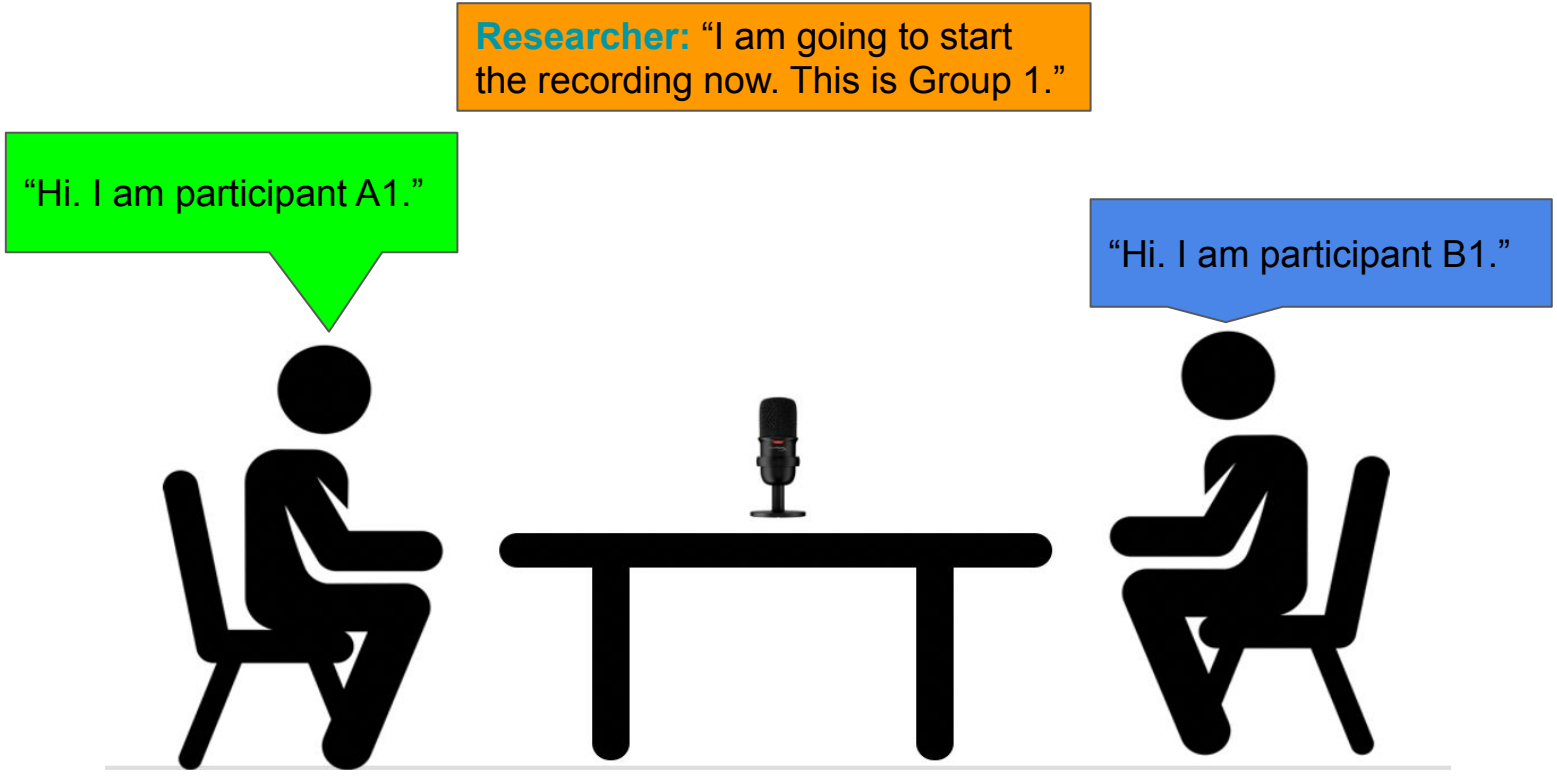
- Main factors:
  - Microphone quality
  - Speaker clarity
  - Background noise
  - Distance from the microphones
  - Echoes
- Try to rely on fixes that *do not* put a burden on the speakers

Humans in the Loop: Prepare

Number of recordings: How many recordings will you want to obtain?

Humans in the Loop: Prepare

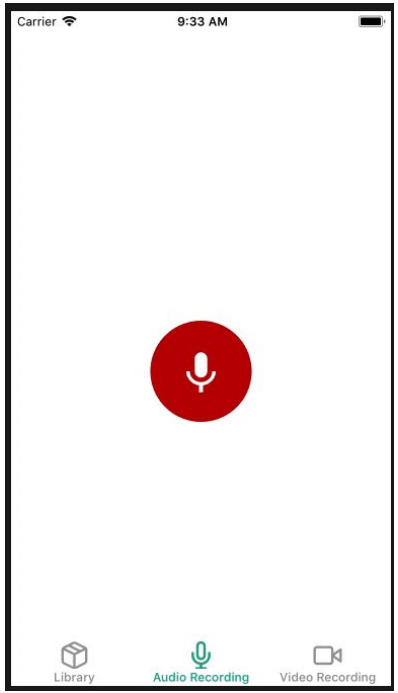
Organization: How will you associate the people speaking?



Humans in the Loop: Record

<u>Test Recordings:</u> Check to see that the file records, that you can play it back on your computer, and that it is saved in a format that is compatible with your transcription method

- ***Don’t forget:*** Hit record!



Humans in the Loop: Transcribe

Method of Transcription: We suggest a **hybrid model**

1. Auto-transcription service
2. Human correction

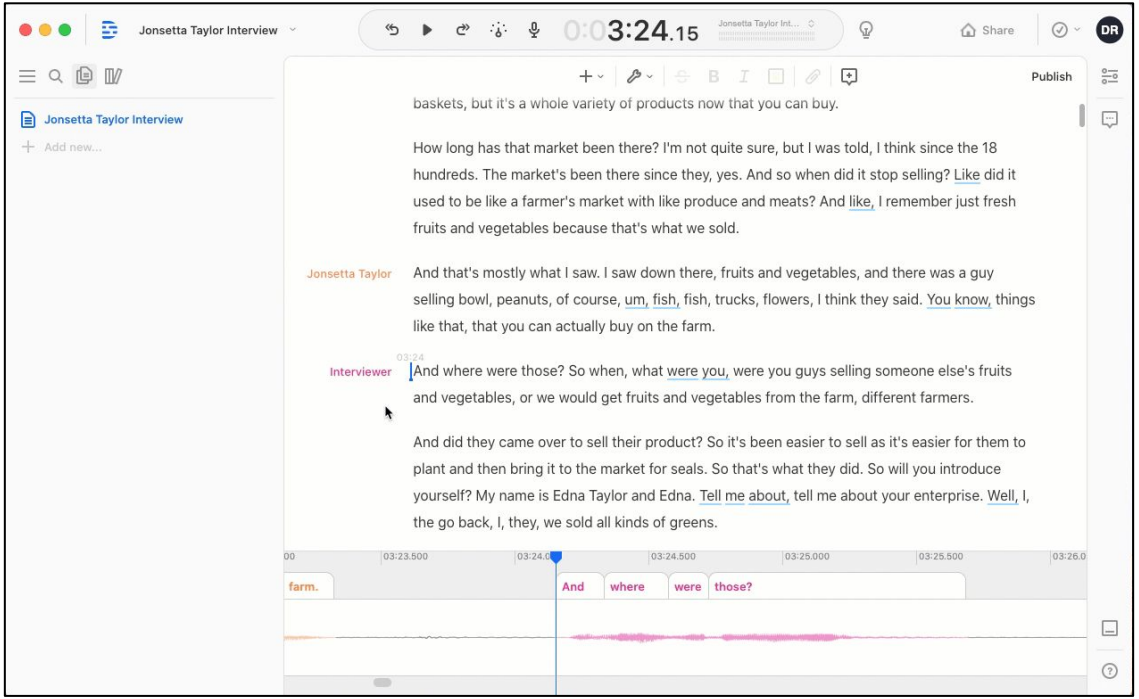
Humans in the Loop: Transcribe

Auto-Transcription Services: There are several services available!

1. Transcription accuracy
2. Speaker differentiation
3. Incorporation of time stamps
4. User friendliness
5. Pricing details

Humans in the Loop: Check

Human correction: Automated transcription services are not perfect and are the most accurate when humans clean the exported transcripts



Humans in the Loop: Check

Common Corrections: Listen to the audio recording while reading through the transcript; fix any mistakes and standardize the format

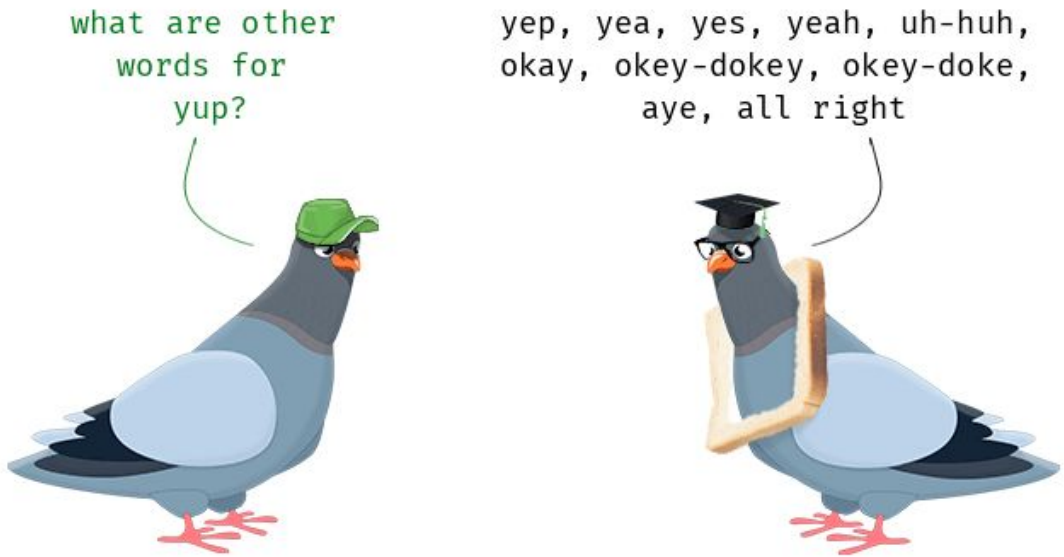
- Include and standardize the spelling of all non-verbals (e.g. "yeah", "uh-huh", "oh").
- Remove erroneously repeated words (e.g. “I thought... thought you were ready”)
- Include punctuation question marks, exclamation marks, periods, commas, ellipses...
- Change “gonna”, “sorta”, “dunno”, etc. to “going to”, “sort of”, “don’t know”, etc.
- Correct misspoken words where the intended meaning is clear to all parties

E
text
Hey, how are you? My name is Bruce, but my friends call me Barry.
Nice to meet you Bruce, I'm Sarah. Where are you from?
Thanks for asking! I'm from Chicago actually, you probably haven't heard of it. What about you?

E
text
Hey, how are you? My name is Bruce, but my friends call me Barry.
Nice to meet you Bruce, I'm Sarah. Where are you from?
Thanks for asking! I'm from [inaudible] Chicago actually, you probably haven't heard of it. What about you?

Humans in the Loop: Check

Fixing speaker mistakes: Correct any grammatical mistakes or clear misspeaking from speakers



Humans in the Loop: Final Result

	A	B	C	D	E
1	conversation_id	participant_id	start_time	end_time	text
2	1	A1	0:00:01	0:00:03	Hey, how are you? My name is Bruce, but my friends call me Barry.
3	1	B1	0:00:04	0:00:06	Nice to meet you Bruce, I'm Sarah. Where are you from?
4	1	A1	0:00:06	0:00:12	Thanks for asking! I'm from a small town outside of Chicago actually, you probably haven't heard of it. What about you?
5	1	B1	0:00:13	0:00:19	Probably not [laughter]. I've never been to Chicago. I'm from upstate Portland, Oregon. Have you ever been to Portland?
6	1	A1	0:00:20	0:00:22	No, I haven't! I've been to Seattle, but that's all.
7	1	B1	0:00:22	0:00:27	Seattle is ok. In Portland, we actually call it Vancouver's shoe [laughter].
8	1	A1	0:00:27	0:00:28	[laughter] That's funny.
9	1	B1	0:00:29	0:00:36	Umm. What's your favorite food?
10	1	A1	0:00:37	0:00:45	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.
11	1	B1	0:00:46	0:00:55	Cool. What was in your stew?

Agenda

Part 1 - Conversation Technology (5-15 min)

Presented by Nicole Abi-Esber (nabiesber@hbs.edu)

Part 2 - Humans in the Loop: Transcriptions and Annotations (15-25 min)

Presented by Katie Boland (kboland@hbs.edu)

Part 3 - The Structure of Conversation Data (25-35 min)

Presented by Hanne Collins (hcollins@hbs.edu)

Part 4 - Applied Example: Measuring politeness in social interactions (35-80 min)

Presented by Michael Yeomans (myeomans@imperial.ac.uk)

And beyond... **A User's Guide to Conversation Research (working paper)**

Structure of conversation data (Hanne)

- Section 2 from PTT
- Mike's Slide 143
- Conversation already has an inherent structure so that we can understand eachother--what does that look like
  - Turn taking
  - Back and forth
  - What are we talking about and what are our goals
  - Why do we need to be so disciplined about structure? What is said when matters--you're co-constructing a document together (order/timing)
    - Language choices in each turn
    - Timing (new to people)
    - Goals (new to people)
- Turn-level dataset
  - Time stamps, pauses, turn length
- Person-level dataset--talk about the different types (each person in only one conversation? multiple?); not usually from conversation directly (survey before, survey after, randomly assigned condition)-->UNIQUE ID TO JOIN DATSETS
- What do we add to the data to make it really rich?
  - Human coded things in conversation level dataset→ Topic choices, topic transitions, offer amounts (negotiation), laughter
  - Demographics/ratings in person level dataset → group conversation level data into person level data
- Use example to walk through code? Hedging across conditions, merging hedge sums into person-level data
  - MERGE CONVO DATA INTO PERSON LEVEL SO YOU DON'T FAKE THAT N → can use Mike's convo code from github (https://github.com/myeomans/NLPatLBS)
- Think back to nicole….all those different type of conversations? Here's one way to group them (speed dating and work meetings//sales calls, mental health hotlines) -- (a)symmetrical goals

The Structure of Conversation

Oh, hey, listen, Jim. Here's a little tip for your performance review. Tell Michael that we should be stocking more of the double-tabbed manila file folders.

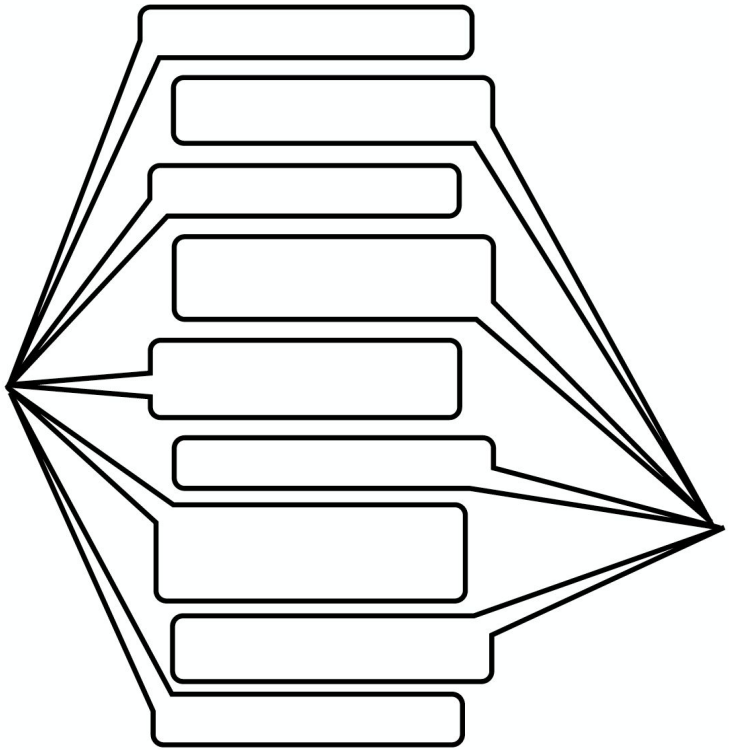
We don't have double-tabbed manila file folders.

Oh, yes, we do. It's a new product. So, you should just suggest that to him and he'll be sure to give you a raise.

Alright... well, I'm not asking for a raise. I'm gonna actually be asking for a pay decrease.

Oh, hey, listen, Jim. Here's a little tip for your performance review. Tell Michael that we should be stocking more of the double-tabbed manila file folders. We don't have double-tabbed manila file folders. Oh, yes, we do. It's a new product. So, you should just suggest that to him and he'll be sure to give you a raise. Alright... well, I'm not asking for a raise. I'm gonna actually be asking for a pay decrease.

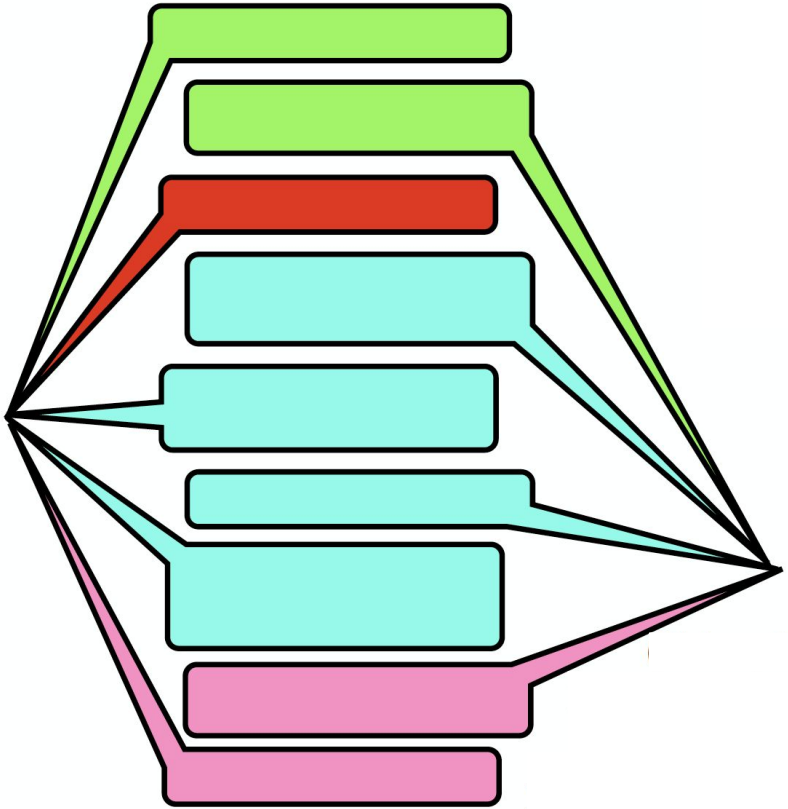
The Structure of Conversation



1. **<u>Turn Taking</u>**: *who is talking*

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

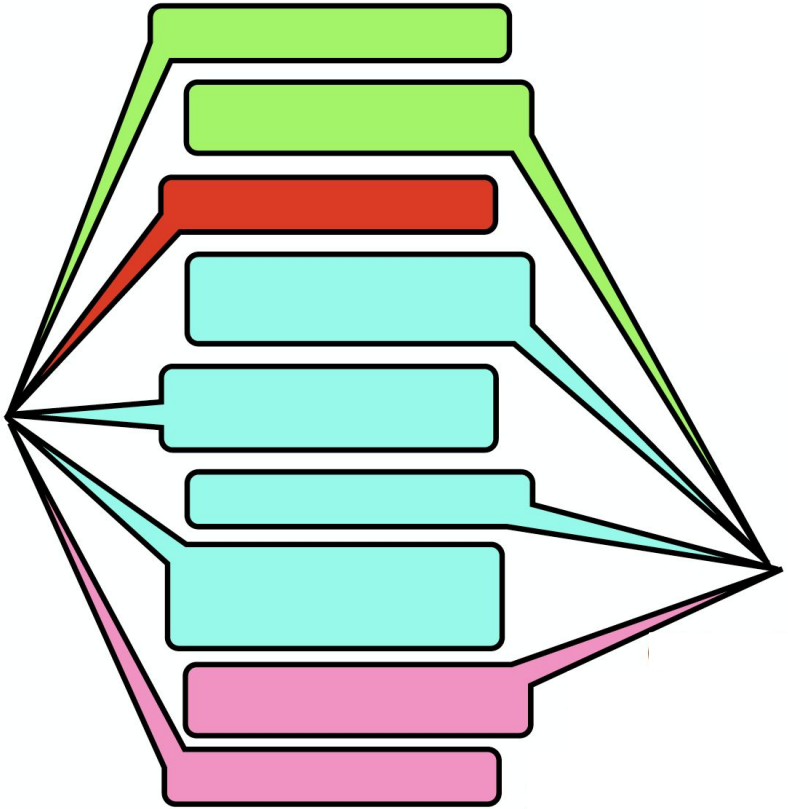
The Structure of Conversation



2. **<u>Topic Selection:</u>** *what we are talking about*

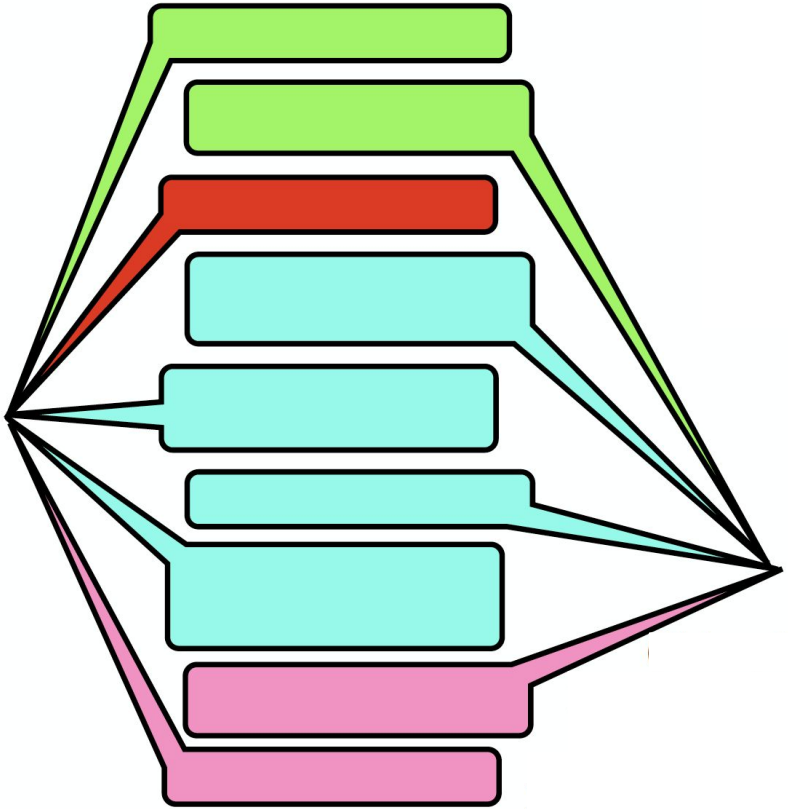
(Grice, 1975; Hardin & Higgins, 1996; Hearst, 1997; Passonneau & Litman, 1997; Drew & Holt, 1998; Galley et al., 2003; Blei, Ng & Jordan, 2003; Nguyen et al., 2014)

The Structure of Conversation



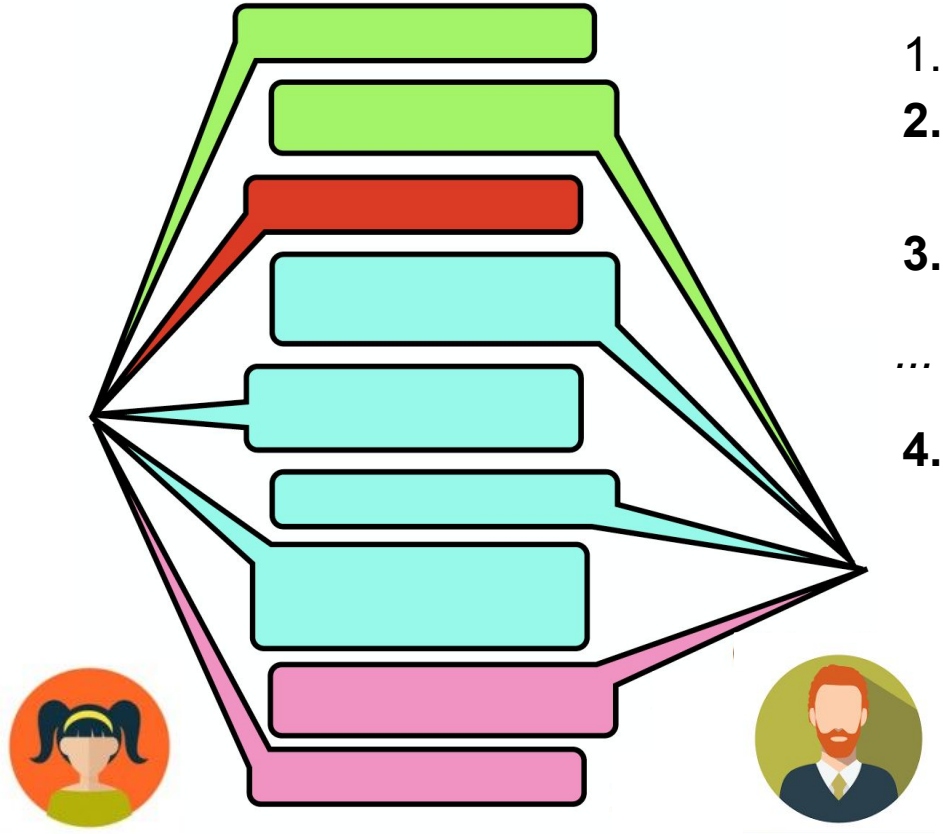
3. **<u>Timing:</u>** *when we are saying what*

The Structure of Conversation



1. **Turn Taking**: *who is talking*
2. **Topic Selection:** *what we are talking about*
3. **Timing:** *when we are saying what*

The Structure of Conversation



1. **<u>Turn Taking</u>**: *who is talking*
2. **<u>Topic Selection:</u>** *what we are talking about*
3. **<u>Timing:</u>** *when we are saying what*

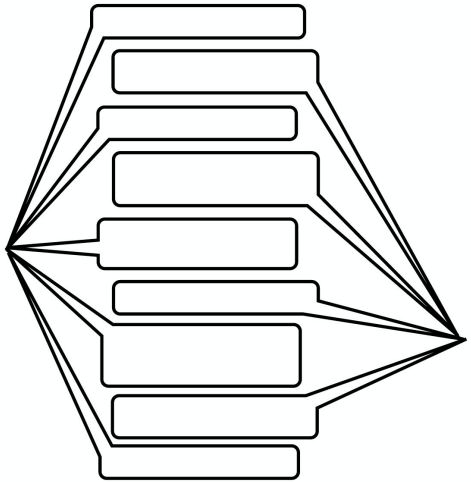
...

4. **<u>People:</u>**
   - Goals (e.g., information exchange vs. coordination)
   - Interests (e.g., topic preference vs. avoidance)
   - etc.

Conversation Data - Two Tables

Turn-Level Dataset: One row for every “turn” in the conversation

- *speaker/convo ID*
- *timestamps*
- *words*
- *features of words*



Conversation Data - Two Tables

Turn-Level Dataset: One row for every “turn” in the 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

Conversation Data - Two Tables

Turn-Level Dataset: One row for every “turn” in the conversation

- NLP-generated features
  - ‘Politeness’ Package in R (Yeomans et al., 2018)
  - ‘Doc2concrete’ Package in R (Yeomans et al., 2021)

Conversation Data - Two Tables

Turn-Level Dataset: One row for every “turn” in the conversation

- NLP-generated features
  - ‘Politeness’ Package in R (Yeomans et al., 2018)
  - ‘Doc2concrete’ Package in R (Yeomans et al., 2021)

- Human-coded features
  - Laughter
  - Topics
  - Pauses
  - Offer amounts (negotiations)

Conversation Data - Two Tables

Person-Level Dataset: One row for each speaker in each conversation (there will be multiple rows per conversation, and depending on study design, perhaps speaker as well)

- *speaker/convo/partner ID*
- *feature summaries from self & partner*
- *metadata on goals/demographics/outcomes/condition/role*

Conversation Data - Two Tables

Person-Level Dataset: One row for each speaker in each conversation (there will be multiple rows per conversation, and depending on study design, perhaps speaker as well)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	conversation_id	participant_id	partner_id	age	gender	partn_gende	role	condition	liking	partn_liking	enjoyment	perceived_lis	questions	laughter	turns	word_count
2	1	A1	B1	24	1	2	A	1	5	6	5	4	2	1	5	87
3	1	B1	A1	34	2	1	B	1	6	5	6	5	4	2	5	53
4	2	A2	B2	57	1	1	A	2	2	5	3	3	0	0	2	45
5	2	B2	A2	23	1	1	B	2	5	2	5	3	1	0	4	24

Conversation Data - Two Tables

Person-Level Dataset: One row for each speaker in each conversation (there will be multiple rows per conversation, and depending on study design, perhaps speaker as well)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	conversation_id	participant_id	partner_id	age	gender	partn_gende	role	condition	liking	partn_liking	enjoyment	perceived_lis	questions	laughter	turns	word_count
2	1	A1	B1	24	1	2	A	1	5	6	5	4	2	1	5	87
3	2	A1	B2	24	1	1	A	1	2	7	3	5	3	1	4	60
4	3	A1	B3	24	1	2	A	1	7	6	4	4	1	0	3	54
5	1	B1	A1	34	2	1	B	1	6	5	6	5	4	2	5	53
6	2	B1	A2	34	2	1	B	1	6	2	2	2	0	3	6	102
7	3	B1	A3	34	2	2	B	1	5	4	7	3	3	1	7	131
8	1	A2	B2	57	1	1	A	2	2	5	3	3	0	0	2	45
9	2	A2	B3	57	1	2	A	2	1	7	4	6	1	1	4	75
10	3	A2	B1	57	1	2	A	2	2	6	5	7	1	0	5	64
11	1	B2	A2	23	1	1	B	2	5	2	5	3	1	0	4	24
12	2	B2	A3	23	1	2	B	2	7	5	6	7	3	3	5	33
13	3	B2	A1	23	1	1	B	2	7	2	6	7	4	2	6	98
14	1	A3	B3	55	2	2	A	1	3	4	4	6	2	1	3	112
15	2	A3	B1	55	2	2	A	1	4	5	3	4	5	1	4	33
16	3	A3	B2	55	2	1	A	1	5	7	6	5	1	2	2	16
17	1	B3	A3	19	2	2	B	2	4	3	3	5	1	0	3	47
18	2	B3	A1	19	2	1	B	2	6	7	6	3	0	0	4	87
19	3	B3	A2	19	2	1	B	2	7	1	5	4	0	1	6	101

Conversation Data - Sharing

Open science is important. Privacy is really important.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	conversation_id	participant_id	partner_id	age	gender	partn_gende	role	condition	liking	partn_liking	enjoyment	perceived_lis	questions	laughter	turns	word_count
2	1	A1	B1	24	1	2	A	1	5	6	5	4	2	1	5	87
3	1	B1	A1	34	2	1	B	1	6	5	6	5	4	2	5	53
4	2	A2	B2	57	1	1	A	2	2	5	3	3	0	0	2	45
5	2	B2	A2	23	1	1	B	2	5	2	5	3	1	0	4	24

	A	B	C	D	E	G	H	I
1	conversation_id	turn	start_time	end_time	participant_id	question	laughter	wordcount
2	1	1	0:00:01	0:00:03	A1	1	0	14
3	1	2	0:00:04	0:00:06	B1	1	0	11
4	1	3	0:00:06	0:00:12	A1	1	0	21
5	1	4	0:00:13	0:00:19	B1	1	1	19
6	1	5	0:00:20	0:00:22	A1	0	0	10
7	1	6	0:00:22	0:00:27	B1	0	1	12
8	1	7	0:00:27	0:00:28	A1	0	1	3
9	1	8	0:00:29	0:00:36	B1	1	0	5

Conversation Data - Two Tables

Turn-Level Dataset: One row for every “turn” in the conversation

Person-Level Dataset: One row for each speaker in each conversation (there will be multiple rows per conversation, and depending on study design, perhaps speaker as well)

Agenda

Part 1 - Conversation Technology (5-15 min)

Presented by Nicole Abi-Esber (nabiesber@hbs.edu)

Part 2 - Humans in the Loop: Transcriptions and Annotations (15-25 min)

Presented by Katie Boland (kboland@hbs.edu)

Part 3 - The Structure of Conversation Data (25-35 min)

Presented by Hanne Collins (hcollins@hbs.edu)

Part 4 - Applied Example: Measuring politeness in social interactions (35-80 min)

Presented by Michael Yeomans (myeomans@imperial.ac.uk)

And beyond... **A User's Guide to Conversation Research (working paper)**