Stancetaking in Spontaneous Speech

Maya Asher

Data Science for Linguists – Spring ‘24

Table of contents

01

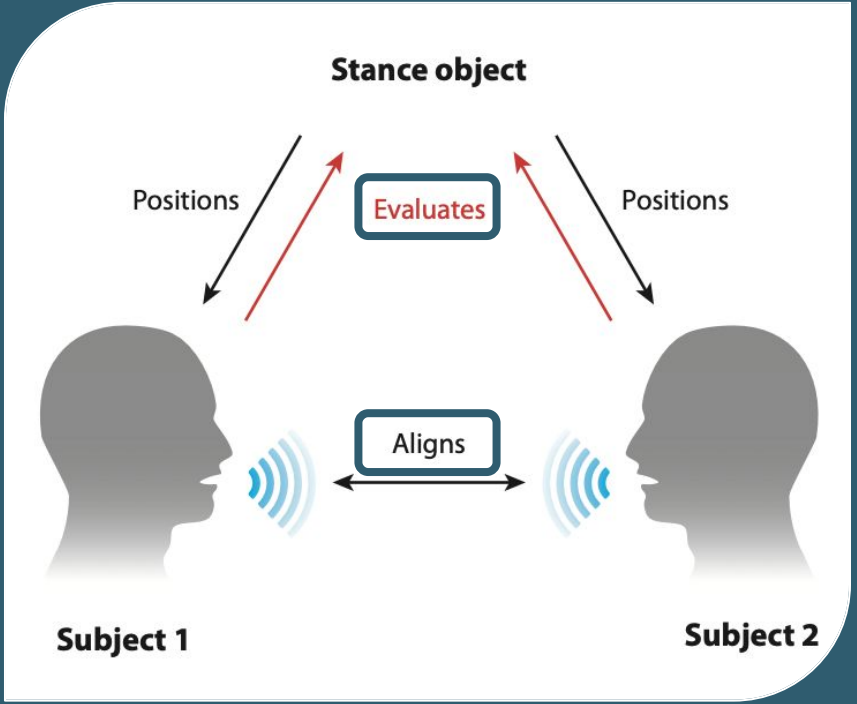
Background and Motivation

Stance and stancetaking

- ➔ Stance and stancetaking refer to the ways people position themselves in conversations
  - In terms of politeness, certainty, or emotion
  - In regard to interactants and content/objects of conversation
- ➔ Helps explain the patterning of language and the motivations for the use of lexical items, constructions, and discourse markers
- ➔ Method of analysis used in discourse analysis (and sociolinguistics)
- ➔ Many different theories and models…

(Kiesling, 2022)

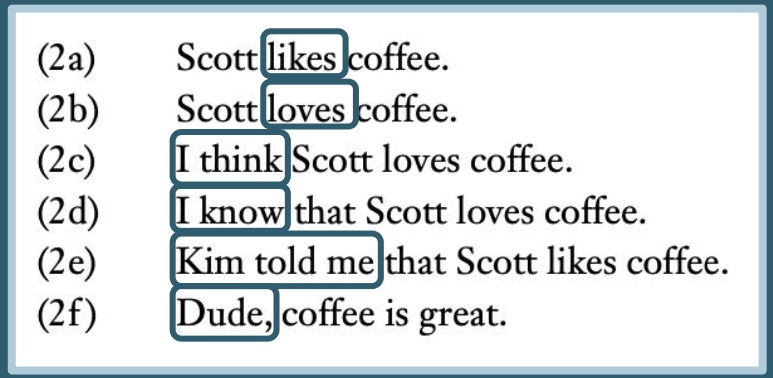
The Stance Triangle



(Du Bois, 2007, as cited in Kiesling, 2022, p. 418)

Updated Stancetaking Model

- ➔ 3 “speaker” participant roles
  - ◆ Author
  - ◆ Animator
  - ◆ Principal
- ➔ 3 dimensions
  - ◆ Evaluation
  - ◆ Alignment
  - ◆ **Investment**


(2a) Scott likes coffee.
(2b) Scott loves coffee.
(2c) I think Scott loves coffee.
(2d) I know that Scott loves coffee.
(2e) Kim told me that Scott likes coffee.
(2f) Dude, coffee is great.

(Kiesling, 2022, p. 420)

Investment

Investment can be modulated through epistemicity and evidentiality…BUT it can also be embedded in the meanings of…

- ➔ Lexical items and constructions (*like* vs. *love*)
- ➔ Discourse markers (*I mean*, *just sayin'*)
- ➔ Hedges and boosters

Classification	Example
Hedges	
modal verbs	*could*, *might*, *would*
epistemic adjectives, adverbs and nouns	*perhaps*, *likely*, *interpretation*
lexical verbs	*seem*, *assume*, *suggest*
Boosters	
modal verbs	*must*, *will*
epistemic adjectives, adverbs and nouns	*obvious*, *always*, *argument*
lexical verbs	*demonstrate*, *show*, *find*

(Wang & Jiang, 2018)

(Kiesling, 2022, p. 420)

The Present Study

Investigating investment through lexical items and discourse markers in a corpus of spontaneous speech

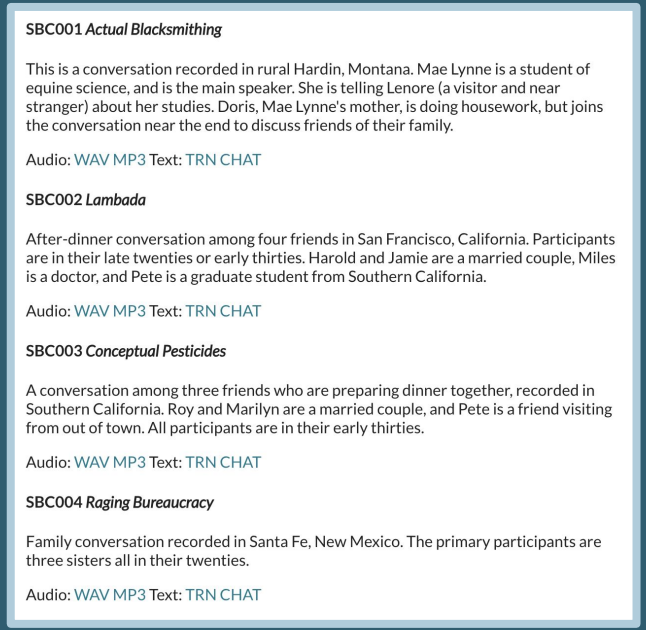
- ➔ Lexical items: *like* vs. *love*
- ➔ Discourse markers: *well, fine, good, great, right, I mean*

Research Questions:

- ➔ Which lexical item/discourse marker is used most frequently in this corpus?
- ➔ Which is most informative of the animator's investment?
- ➔ What generalizations can be made for the investment level of each item/marker?

Santa Barbara Corpus of Spoken American English (SBCSAE)

- **60** recordings/transcriptions and **~249,000** words
  - Only using **43 of the 60** (face-to-face spontaneous speech)
- All over the U.S. – wide variety of people of different regional origins, ages, occupations, genders, and ethnic and social backgrounds



SBC001 ***Actual Blacksmithing***

This is a conversation recorded in rural Hardin, Montana. Mae Lynne is a student of equine science, and is the main speaker. She is telling Lenore (a visitor and near stranger) about her studies. Doris, Mae Lynne's mother, is doing housework, but joins the conversation near the end to discuss friends of their family.

Audio: WAV MP3 Text: TRN CHAT

SBC002 ***Lambada***

After-dinner conversation among four friends in San Francisco, California. Participants are in their late twenties or early thirties. Harold and Jamie are a married couple, Miles is a doctor, and Pete is a graduate student from Southern California.

Audio: WAV MP3 Text: TRN CHAT

SBC003 ***Conceptual Pesticides***

A conversation among three friends who are preparing dinner together, recorded in Southern California. Roy and Marilyn are a married couple, and Pete is a friend visiting from out of town. All participants are in their early thirties.

Audio: WAV MP3 Text: TRN CHAT

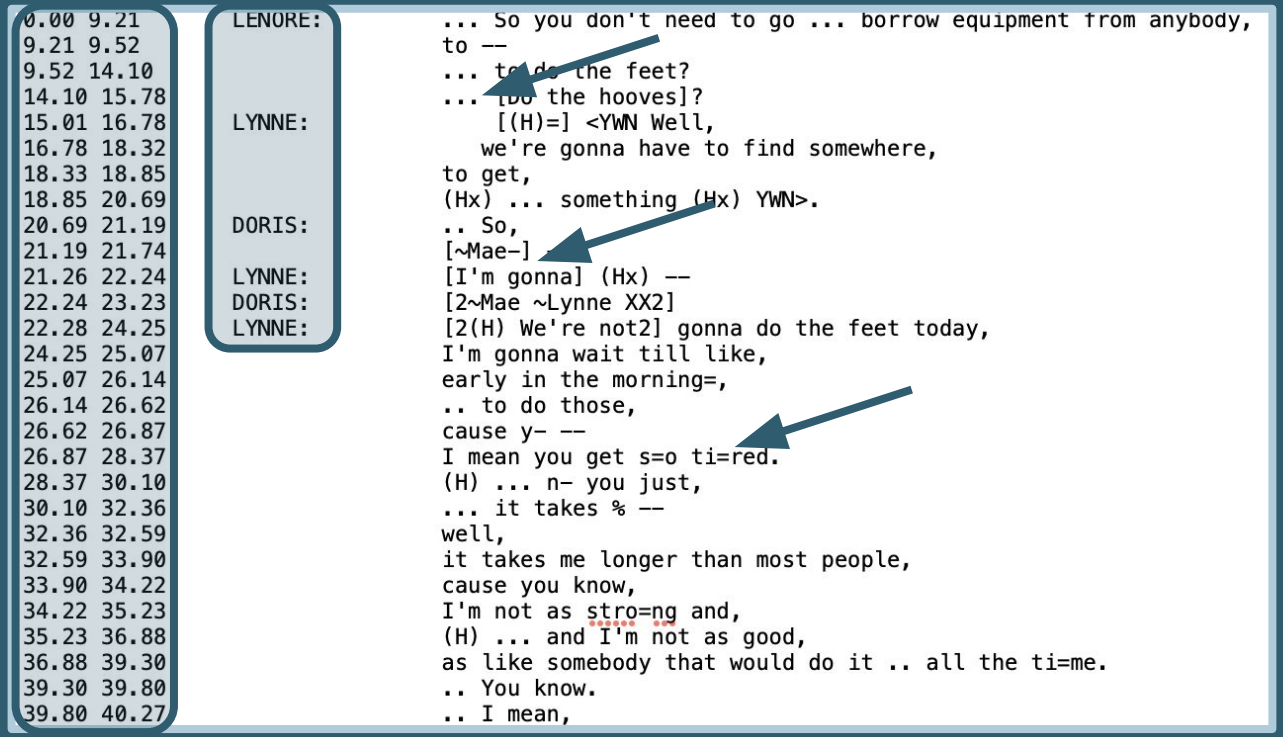
SBC004 ***Raging Bureaucracy***

Family conversation recorded in Santa Fe, New Mexico. The primary participants are three sisters all in their twenties.

Audio: WAV MP3 Text: TRN CHAT

(Du Bois, et al., 2000-2005)

SBCSAE – SBC001.trn



```
0.00 9.21     LENORE:    ... So you don't need to go ... borrow equipment from anybody,
9.21 9.52                to --
9.52 14.10               ... to do the feet?
14.10 15.78              ... [DO the hooves]?
15.01 16.78   LYNNE:          [(H)=] <YWN Well,
16.78 18.32                 we're gonna have to find somewhere,
18.33 18.85              to get,
18.85 20.69              (Hx) ... something (Hx) YWN>.
20.69 21.19   DORIS:     .. So,
21.19 21.74              [~Mae-]
21.26 22.24   LYNNE:     [I'm gonna] (Hx) --
22.24 23.23   DORIS:     [2~Mae ~Lynne XX2]
22.28 24.25   LYNNE:     [2(H) We're not2] gonna do the feet today,
24.25 25.07              I'm gonna wait till like,
25.07 26.14              early in the morning=,
26.14 26.62              .. to do those,
26.62 26.87              cause y- --
26.87 28.37              I mean you get s=o ti=red.
28.37 30.10              (H) ... n- you just,
30.10 32.36              ... it takes % --
32.36 32.59              well,
32.59 33.90              it takes me longer than most people,
33.90 34.22              cause you know,
34.22 35.23              I'm not as stro=ng and,
35.23 36.88              (H) ... and I'm not as good,
36.88 39.30              as like somebody that would do it .. all the ti=me.
39.30 39.80              .. You know.
39.80 40.27              .. I mean,
```

(Du Bois, et al., 2000-2005)

SBCSAE – Annotation key

Santa Barbara Corpus of Spoken American English Part-II

Units	
Intonation Unit	RETURN
Truncated intonation unit	--
word	SPACE
truncated word	_
Speakers	
Speaker identity/turn start	:
Speech overlap	[ ]
Transitional Continuity	
Final	.
Continuing	,
Appeal	?
Terminal Pitch Direction	
Fall	\
Rise	/
Level	-

Accent and Lengthening	
Primary accent	^
Secondary accent	'
Booster	!
Lengthening	=
Tone	
Fall	\
Rise	/
Fall-rise	\/
Rise-fall	/\
Level	-
Pause	
Long	...(N)
Medium	...
Short	..
Latching	(0)
Vocal Noises	
Vocal noises	( )
Inhalation	(H)
Exhalation	(Hx)
Glottal stop	%
Laughter	@
Quality	
Quality	<Y Y>
Laugh quality	<@ @>
Quotation quality	<Q Q>
Multiple quality features	<Y <Z Z> Y>

Phonetics	
Phonetic transcription	(/ /)
Transcriber's Perspective	
Researcher's comments	(( ))
Uncertain hearing	<X X>
Indecipherable syllable	X
Specialized notation	
Duration	(N)
Intonation unit continued	&
Intonation subunit boundary	|
Embedded intonation unit	<| |>
Reset	
False start	< >
Codeswitching	<L2 L2>
Non-transcription Lines	
Comment	$
Interlinear gloss	$G
Reserved Symbols	
Phonemic/orthographic	,
Morphosyntactic coding	= * # { }
User-definable	" ~ ;

(Du Bois, et al., 2000-2005)

Why spontaneous speech?

Spontaneous speech vs. read speech

- ➔ Boundaries between tone units at different points
- ➔ Different positions of stress
- ➔ Fewer pauses in read speech
- ➔ Locations of pauses differ

Discourse analysis

- ➔ Studies natural interactions between interlocutors
- ➔ Read speech could help gather acoustic measurements, but spontaneous speech is needed to conduct a discourse analysis

(Howell & Kadi-Hanifi, 163)

Motivation

- Discourse Analysis with Dr. Scott Kiesling
- Working with speech data with Dr. Dan Villarreal

21 S: I mean this- ↑these are expensive and look how gro:ss it is ((pan clinks)) and then this thing=

22 A: =but ↑why would they make them so expensive if you can't use it?

23 S: you can use it but you have to-they're s-they're so ↑go:od that they need so: much care=

24 A: =they're so good that they're ↓bad

A Damaged Pan

The conversation below takes place on a Wednesday morning between Anjali and her mother, Shilpa. Both of them are in the kitchen getting breakfast between college/work meetings when Shilpa begins passionately informing Anjali about a damaged pan.

Transcription Key

↑↓	rising and falling intonation
:	extension of a syllable
|	overlapping utterances
=	no interval between utterances
-	stutter
_	emphasis
((sound))	non-word sounds
(0.0)	interval of silence over ½ second

Transcript

01	Shilpa:	um this th-thing is like
02		it's so expensive and a:ll this stuff is so messed up and the-they're like
03	Anjali:	↑what?
04	Shilpa	they're like
05		I mean ↑look at it like it's
06		I-I did I did baking soda so now it looks |better|
07	Anjali:	|↑↓o:h|
08	Shilpa:	but I was they were like you're only supposed to u- they said this is so (1.1) sensitive you're only supposed to use ↓low to medium ↑heat you can only use ↑butter like there's all |these rules.|
09	Anjali:	|↑why:?| it's a ↑pa:n
10	Shilpa:	no but it's like (0.6) t- and ↑then I'm like I ↑said ↓um (0.9)
11		I've bought these I've given 'em as gifts (0.7) so I was like
12		so: are you gonna honor your warranty
13	Anjali:	↓↑↓o:h=
14	Shilpa:	=and they said well you're ↑using it ↓wrong so no:
15	Anjali:	((stuttered laugh))
16	Shilpa:	but then they said okay we wi:ll
17		so: they're gonna send me a new one and take sen- take this back (1.1)
18	Anjali:	|but what's the|
19	Shilpa:	|but now I have to| use it at low to medium-
20		look I mean lo:ok at it it's like horri- like
21		I mean this- ↑these are expensive and look how gro:ss it is ((pan clinks)) and then this thing=

02

Methods and Data

Extracting tokens

```python
wells = {}
fines = {}
goods = {}
greats = {}
rights = {}
imeans = {}

# iterate through each df in the dict
for df_name, df in dfs.items():

    # extract rows where the target word is found in the text column
    well_rows = df[df['text'].str.contains(r'\bwell\b', case=False, regex=True)]
    fine_rows = df[df['text'].str.contains(r'\bfine\b', case=False, regex=True)]
    good_rows = df[df['text'].str.contains(r'\bgood\b', case=False, regex=True)]
    great_rows = df[df['text'].str.contains(r'\bgreat\b', case=False, regex=True)]
    right_rows = df[df['text'].str.contains(r'\bright\b', case=False, regex=True)]
    imean_rows = df[df['text'].str.contains(r'\bI\smean\b', case=False, regex=True)]

    # list of tuples containing (row #, text) for each occurrence
    well_occurrences = [(index, row['text']) for index, row in well_rows.iterrows()]
    fine_occurrences = [(index, row['text']) for index, row in fine_rows.iterrows()]
    good_occurrences = [(index, row['text']) for index, row in good_rows.iterrows()]
    great_occurrences = [(index, row['text']) for index, row in great_rows.iterrows()]
    right_occurrences = [(index, row['text']) for index, row in right_rows.iterrows()]
    imean_occurrences = [(index, row['text']) for index, row in imean_rows.iterrows()]

    # save the occurrences in the dict
    wells[df_name] = well_occurrences
    fines[df_name] = fine_occurrences
    goods[df_name] = good_occurrences
    greats[df_name] = great_occurrences
    rights[df_name] = right_occurrences
    imeans[df_name] = imean_occurrences
```

```
1 goods
```

```
{'SBC001.trn': [(24, "(H) ... and I'm not as good,"),
use you don't want to .. cripple up a .. (H) really go
good stereo.'), (490, "[(H) That's pretty] good,"), (8
good .. lambada dancer?'), (1407, '... Well that was g
Good good.'), (111, '.. that sounds good.'), (410, ".
it's all fine [and good
[2That's2] good."),
at looks good.'), (1
are so good leftover
eally good."), (1476
t's good is] .. hibiscus
morni=ng,'), (331, 'good bye=,'), (610, ".. he's real
s,'), (852, '(H) And then give those p=asses out for g
et really good,')], 'SBC005.trn': [(65, 'k=- so good g
good time Saturday,'), (542, "that's a good [spot for
ood friend herself."), (1167, 'I2] had a good time wit
ew friends,'), (1226, '(H) was his good buddy.'), (141
[That's good P>]."),], 'SBC007.trn': [(93, 'a real gooc
good idea for you to go up there in the winter."),], 'S
```

Keys: file name
Values: list of tuples
(line #, text)

Isolating indices and exploring context

```
indices = {key: [tpl[0] for tpl in value] for key, value in goods.items()}
print(indices)
```

```
{'SBC001.trn': [24, 40, 496], 'SBC002.trn': [236, 490, 874, 940, 1407], 'SBC005
768, 932, 935, 947, 1003, 1098, 1263, 1430, 1433, 1453, 1476], 'SBC004.trn': [1
864], 'SBC005.trn': [65], 'SBC006.trn': [179, 542, 721, 744, 1167, 1206, 1226,
8], 'SBC008.trn': [1029, 1488], 'SBC009.trn': [688], 'SBC010.trn': [766, 993],
02, 652, 755, 849], 'SBC013.trn': [247, 339, 382, 786, 787, 854, 865, 968, 1232
7, 1789, 1813, 1821, 1994, 1996, 2207, 2234], 'SBC014.trn': [493, 617, 641, 657
764], 'SBC016.trn': [3, 13, 81, 386, 388, 523, 617, 694, 1070, 1135, 1148, 1417
C017.trn': [], 'SBC018.trn': [4, 106, 163, 176], 'SBC019.trn': [96, 247, 251, 2
190], 'SBC022.trn': [5, 23, 316, 336, 366, 367, 404, 405, 418, 453], 'SBC023.tr
1353, 1391, 1394], 'SBC024.trn': [80, 742, 869, 871], 'SBC029.trn': [58, 67, 25
8, 799, 981, 982, 1011, 1082, 1097], 'SBC031.trn': [45, 47, 162, 191, 291, 346,
8, 1489, 1510, 1511], 'SBC032.trn': [84, 345, 346, 582, 583, 849, 1306, 1368, 1
34, 338, 393, 508, 522], 'SBC034.trn': [26, 360, 504, 553], 'SBC035.trn': [157,
619, 814, 1054, 1118, 1143, 1201, 1226, 1293], 'SBC036.trn': [10, 53, 108, 378,
797], 'SBC037.trn': [86, 428, 528, 579, 629, 644, 663], 'SBC042.trn': [457, 679
47, 310, 311, 413, 442, 444, 488, 550, 847, 855, 1001, 1012, 1015, 1016, 1038,
1485], 'SBC044.trn': [113, 128, 129, 194, 266, 440, 448, 1348, 1366, 1394], 'SB
[156, 157, 328, 347, 896, 919, 922, 1010, 1058, 1071, 1100, 1103, 1109, 1157],
96, 414, 415, 417, 838, 842, 908, 926, 1000, 1002, 1011, 1025], 'SBC049.trn': [
1182], 'SBC050.trn': [19, 29, 90, 117, 175, 186, 273, 483, 486, 722, 798, 799,
1229, 1509, 1669], 'SBC056.trn': [89, 273, 282, 344, 345, 346, 411, 522, 533, 6
229, 1238, 1470], 'SBC057.trn': [117, 193, 300, 330, 526, 716, 967, 974, 990],
9, 564, 751, 863], 'SBC059.trn': [567, 569, 625, 628, 637, 696, 700, 713, 715,
384, 1607, 1637], 'SBC060.trn': [652, 654, 672, 800, 808, 995, 996]}
```

```
targets = [likes, loves, wells, fines, goods, greats, rights, imeans]
targets_str = ['likes', 'loves', 'wells', 'fines', 'goods', 'greats', 'rights',
count = 0

for target in targets:
    # establish indices
    indices = {key: [tpl[0] for tpl in value] for key, value in target.items()}

    # write out file
    with open(targets_str[count]+"_output.txt", 'w') as file:

        import sys
        sys.stdout = file

        for index in indices:
            df = dfs[index]
            print(index)
            key = index
            values = indices[key]
            # give 20 lines of context
            for v in values:
                first = v - 10
                last = v + 10
                print(df.iloc[first:last])
    count+=1

    sys.stdout = sys.__stdout__
```

Output files

	time		speaker	text
983	842.52	843.37		... [Yeah <X I know X>],
984	842.94	845.04	PHIL:	[(H)] !Jack and I and !Jim .. met with him.
985	845.04	845.34	BRAD:	.. Oh,
986	845.34	845.94		you mean his,
987	845.94	847.86		.. !Jack's [financial] friend that --
988	846.63	847.25	PHIL:	[!Jack's friend].
989	847.86	848.81	BRAD:	.. is retired.
990	848.81	849.26	PHIL:	Yeah.
991	849.26	850.06		.. [He's a] banker,
992	849.23	849.58	BRAD:	[Yeah].
993	850.06	851.41	PHIL:	.. and he'd be good on one hand.
994	851.41	851.78		.. [But],
995	851.56	851.81	BRAD:	[Yeah].
996	851.81	852.50	PHIL:	I would like,
997	852.50	854.37		(H) ideally XX I'd want em both.
998	854.37	854.87	BRAD:	.. Yeah.
999	854.87	855.54		... [Yeah].
1000	855.38	855.80	BRAD:	[You know].
1001	855.80	856.25		.. And,
1002	856.25	858.29		... [I'm hoping],

	time		speaker \
486	532.42	533.20	
487	533.20	534.27	
488	534.27	535.29	
489	535.29	536.80	
490	536.80	538.15	
491	538.15	539.39	
492	539.39	540.01	
493	540.01	541.32	
494	541.32	543.29	
495	543.29	544.42	
496	544.42	547.07	
497	547.07	547.37	
498	547.37	548.37	
499	548.37	550.02	
500	550.02	550.92	
501	550.92	552.07	
502	552.07	552.87	
503	552.87	553.82	
504	553.82	554.04	
505	554.04	554.36	

Some got disjointed, but the indices are correct

	text
486	they have to go,
487	.. a long ways to go get em,
488	like back East somewhere,
489	to get these .. horse hooves.
490	(H) For the college.
491	(H) They go back % .. East,
492	and they get em.
493	and they freeze em you know?
494	.. (H) So we have this frozen horse hoof,
495	that we have to start out on,
496	cause you don't want to .. cripple up a .. (H)...
497	and like,
498	my first hoof,
499	(H) .. that horse would have been,
500	.. lame,
501	@ @ <@ like cra=zy.
502	@ @ (H) But @>,
503	(H) it was just,
504	and,
505	oh,

(Du Bois, et al., 2000-2005)

03

Findings

Transcript – “fine”

	time_start	time_end	speaker	text
41	58.688	59.596	MELISSA:	[One side] of a page?
42	59.596	60.598		It takes me a long time,
43	60.598	62.461		because I've got to go over the sentences,
44	62.461	63.323		(H) figure out,
45	63.323	65.768		.. if I'm gonna rewrite them or leave them the...
46	65.768	66.909		(H) [and just] write them out.
47	65.768	66.280	JAN:	[Well y-] --
48	66.909	68.772	MELISSA:	I can't write them exactly the way they are,
49	68.772	69.681		because they stink.
50	69.681	71.730	JAN:	.. Then you need to go downstairs and finish it.
51	71.730	72.755	MELISSA:	(H) I'm fine.
52	73.989	74.781	FRANK:	... ~Melissa,
53	74.711	75.829		[it's nine o'clock].
54	74.804	76.807	MELISSA:	[I'm not gonna do any better] downstai=rs.
55	76.761	78.880	FRANK:	It's [2nine o'clock .. in the evening2].
56	77.063	79.509	JAN:	[2You also have algebra2] to do.
57	79.509	80.953	MELISSA:	.. I can skip algebra.
58	80.953	82.024	FRANK:	... No you [can't].
59	81.465	82.024	JAN:	[No].
60	82.024	82.420		.. [2you- (Hx)2] --

(Du Bois, et al., 2000-2005)

SBC019.trn: *Doesn't Work in this Household*

"A family conversation, recorded in Michigan. Frank and Jan (a married couple) are talking with Ron--Jan's brother who is visiting from California. Brett and Melissa are Frank and Jan's junior-high-age children, who are doing homework and also taking part in the conversation."

Transcript – “I mean”

SBC002.trn

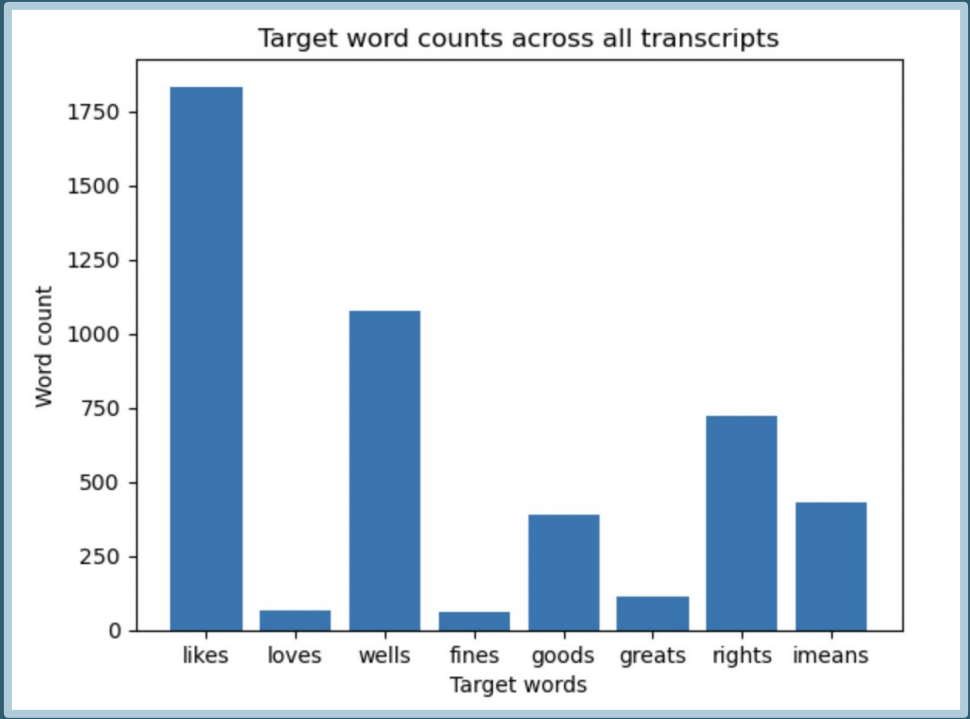
	time		speaker	text
45	45.91	47.53	MILES:	and I guess he really goes fa=st.
46	47.53	47.86		[<X And X>],
47	47.59	47.94	JAMIE:	[Yeah].
48	47.94	48.49	MILES:	(H) he k- --
49	48.49	50.09	JAMIE:	... He does[n't explain anything precisely].
50	48.69	52.20	MILES:	[He has to double it dow=n to] like...
51	52.20	53.31		before they can g=- pick it up-,
52	53.31	54.23		... @@
53	54.23	55.91	HAROLD:	Well I'm sure !Thomas is all over it.
54	55.91	61.26	JAMIE:	... Prob[ably XX] [2XXX2] --
55	59.81	60.56	HAROLD:	[I mean he] [2has a bro-2] --
56	60.26	61.36	MILES:	[2XXXX could have2] see=n him.
57	61.36	63.58	HAROLD:	I guess that means his broken leg is [3@doing ...
58	62.97	63.86	PETE:	[3I was w...
59	63.86	65.25		I was imagining [4he had broke an arm4] or som...
60	64.24	64.99	JAMIE:	[4<HI Oh yeah= HI>4].
61	65.25	66.05	PETE:	But it was his leg?
62	66.05	66.70	HAROLD:	.. Yeah[=].
63	66.35	66.76	PETE:	[That's like],
64	66.76	68.71		.. <X I guess X> that he was being hauled arou...

(Du Bois, et al., 2000-2005)

SBC002.trn: *Lambada*

"After-dinner conversation among four friends in San Francisco, California. Participants are in their late twenties or early thirties. Harold and Jamie are a married couple, Miles is a doctor, and Pete is a graduate student from Southern California."

Figure


Target word counts across all transcripts
Word count
1750
1500
1250
1000
750
500
250
0
likes
loves
wells
fines
goods
greats
rights
imeans
Target words

04

Reflection and Future Steps

Reflection

- What went well
  - Topic
  - Theoretical basis

- What could've been better
  - Data processing
  - Time management

- What I learned
  - Licensing
  - Python libraries

Future Steps

- ➔ Analyze more of the data
- ➔ Do value add
- ➔ Make more figures
- ➔ Draw concrete conclusions

Thanks!

Questions?

References

Du Bois, J. W. (2007). The stance triangle. *Stancetaking in Discourse: Subjectivity, Evaluation, Interaction*, ed. by Robert Englebretson, 139–182

Du Bois, J. W., Chafe, W. L., Meyer, C., Thompson, S. A., Englebretson, R., & Martey, N. (2000-2005). Santa Barbara corpus of spoken American English, Parts 1-4. *Philadelphia: Linguistic Data Consortium*

Howell, P., & Kadi-Hanifi, K. (1991). Comparison of prosodic properties between read and spontaneous speech material. *Speech Communication*, *10*(2), 163–169. https://doi.org/10.1016/0167-6393(91)90039-V

Kiesling, S. F. (2022). Stance and Stancetaking. *Annual Review of Linguistics*, 8, 409-426. https://doi.org/10.1146/annurev-linguistics-031120-121256

Wang, J., & Jiang, F. (Kevin). (2018). Chapter 9. Epistemic stance and authorial presence in scientific research writing: Hedges, boosters and self-mentions across disciplines and writer groups. In P. Mur-Dueñas & J. Šinkūnienė (Eds.), *AILA Applied Linguistics Series* (Vol. 18, pp. 195–216). John Benjamins Publishing Company. https://doi.org/10.1075/aals.18.09wan