

The Future  
of the Past

# The Testimony of the Multitude:

Towards a computational model of  
listening to Holocaust testimonies

Renana Keydar, Amit Pinchevski,  
Maxim Ifergan, Omri Abend



# Motivation

“Fiction gives eyes to the horrified narrator. Eyes to see and to weep. The present state of literature on the Holocaust provides ample proof of this . . .

one either counts the cadavers or one tells the stories of the victim”

PAUL RICŒUR, TIME AND NARRATIVE.

- Mass atrocity and **mass** testimony – disputing Ricoeur's dichotomy
- Dual challenge: abundance of archived testimonies and danger of fixation of the testimonial narrative

# Motivation

- **Distant listening**: engaging with multiple Holocaust testimonies simultaneously by means of a dedicated algorithm
- Using computational tools **against the grain**: not aggregation of data, but attending to the **integrity of testimonial narrative**
- Adopting **multiplicity** as essential to attending to the memory mass atrocity

# Approach: computational distant listening

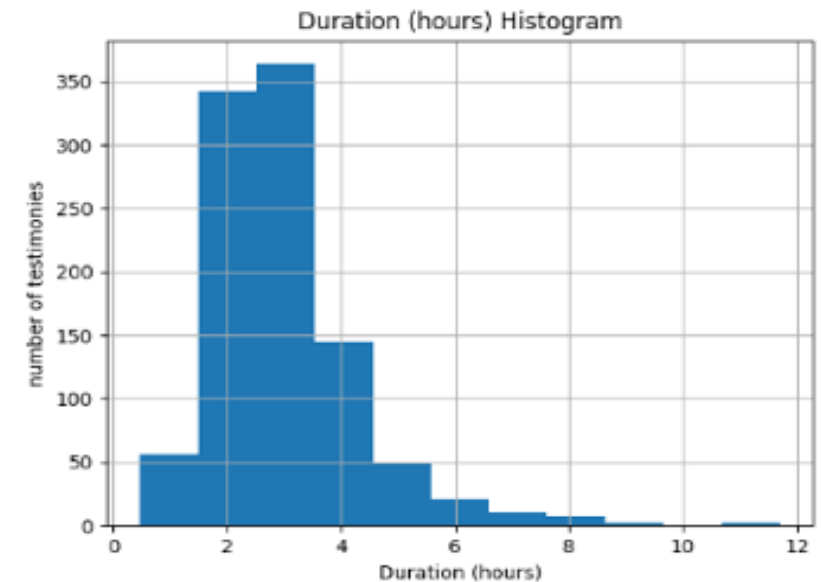
- Identifying **typical** narrative scheme within corpus of testimonies according to a given category
- Detecting testimonies that **diverge from the typical** scheme
- **Recursive iteration**: the process can be repeated indefinitely so that with each run different typical and exceptional are produced

# Approach: computational distant listening

- Contrasting while maintaining **the tension** between the particular story and its categorization
- Technology is **indispensable** in integrating multiplicity and singularity
- This is in order to foster **new ways of listening** that may **unsettle fixed and familiar** representations of witnessing

# Corpus and data

- Transcripts of **1000** Holocaust survivor testimonies in English from USC Shoah Foundation
- Range between **2,609 - 88,105** words per testimony; average testimony length: **22,536** words
- Each testimony was split into Q&A pairs of 250-450 words (each pair = **document**); average testimony length: **62** documents



# Methodology

- Training a **Contextualized Topic Model** (CTM) on the documents
- **Unsupervised** model to find latent “topics” in the corpus (number of topics is set in advance by researchers)
- Topics = **clusters of words** likely to appear together in documents on which the model was trained
- The goals of the model: (1) identify latent topics and (2) identify in each document distribution of topics

# Topics

7: ['work', 'job', 'new', 'company', 'year', 'college', 'york', 'busine', 'marry', 'married', 'university', 'month', 'husband', 'start', 'move']

10: ['jewish', 'jew', 'community', 'antisemitism', 'anti', 'population', 'organization', 'synagogue', 'religiou', 'zionist', 'belong', 'party', 'hitler', 'gentile', 'orthodox']

11: ['name', 'date', 'english', 'interview', 'bear', 'birth', 'spell', 'please', 'today', 'conduct', 'survivor', 'language', 'november', 'state', 'december']

12: ['unit', 'army', 'british', 'military', 'soviet', 'officer', 'italian', 'general', 'force', 'nazi', 'communist', 'war', 'french', 'france', 'join']

16: ['barrack', 'bunk', 'coffee', 'shoe', 'shift', 'factory', 'block', 'count', 'stone', 'cold', 'cement', 'number', 'soup', 'bed', 'assign']

17: ['letter', 'cousin', 'israel', 'write', 'find', 'uncle', 'palestine', 'send', 'meet', 'stay', 'united', 'brother', 'dp', 'aunt', 'relative']

23: ['room', 'house', 'bedroom', 'apartment', 'live', 'kitchen', 'dining', 'floor', 'flat', 'big', 'building', 'bathroom', 'garden', 'maid', 'living']

31: ['god', 'feel', 'message', 'question', 'holocaust', 'think', 'world', 'answer', 'future', 'human', 'never', 'experience', 'affect', 'important', 'generation']

34: ['food', 'buy', 'money', 'egg', 'meat', 'bread', 'sell', 'flour', 'eat', 'ration', 'potato', 'milk', 'sugar', 'market', 'cook']



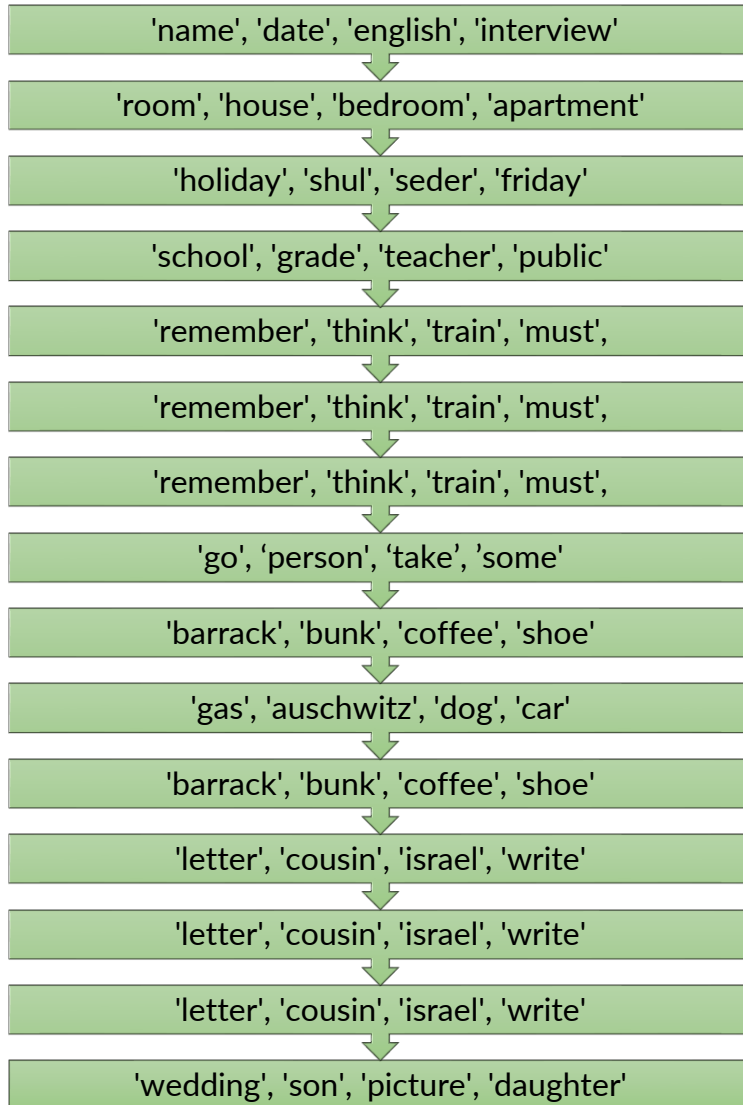
# From topics to narrative scheme

- Based on training on the entire corpus, the model infers for each document the topical distribution across the testimony
- Since testimonies vary in lengths and number of documents, we aligned the testimonies by splitting them into equal number of segments (15 in the examples to follow)
- For every segment, we calculated the most prominent topic

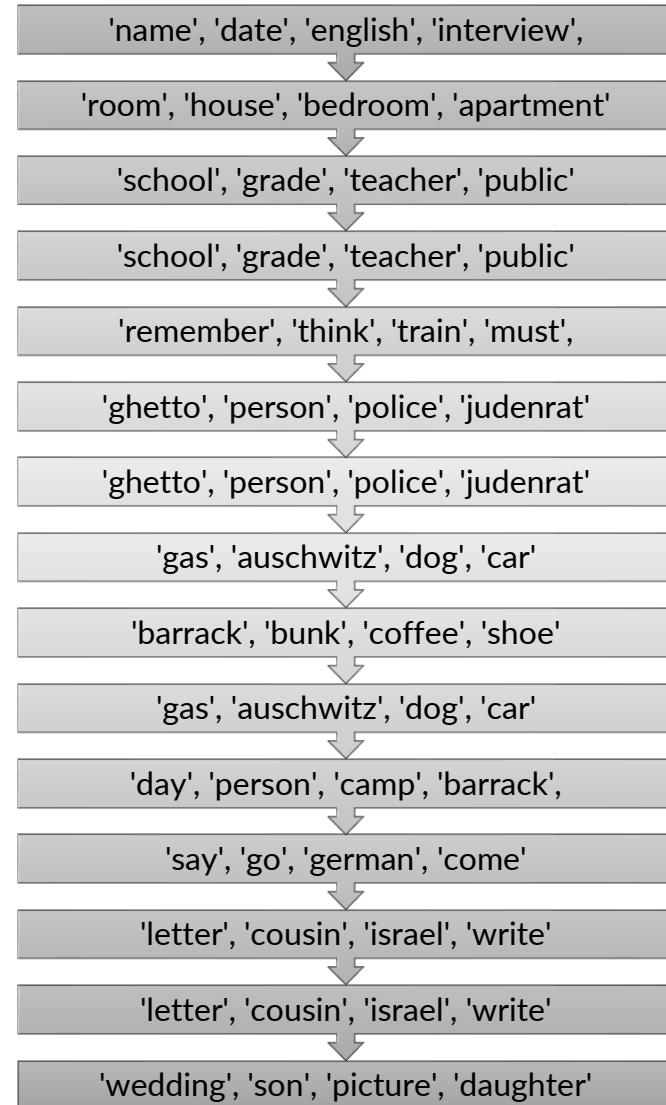
# Typical USC testimonial narrative schema



# Typical Female



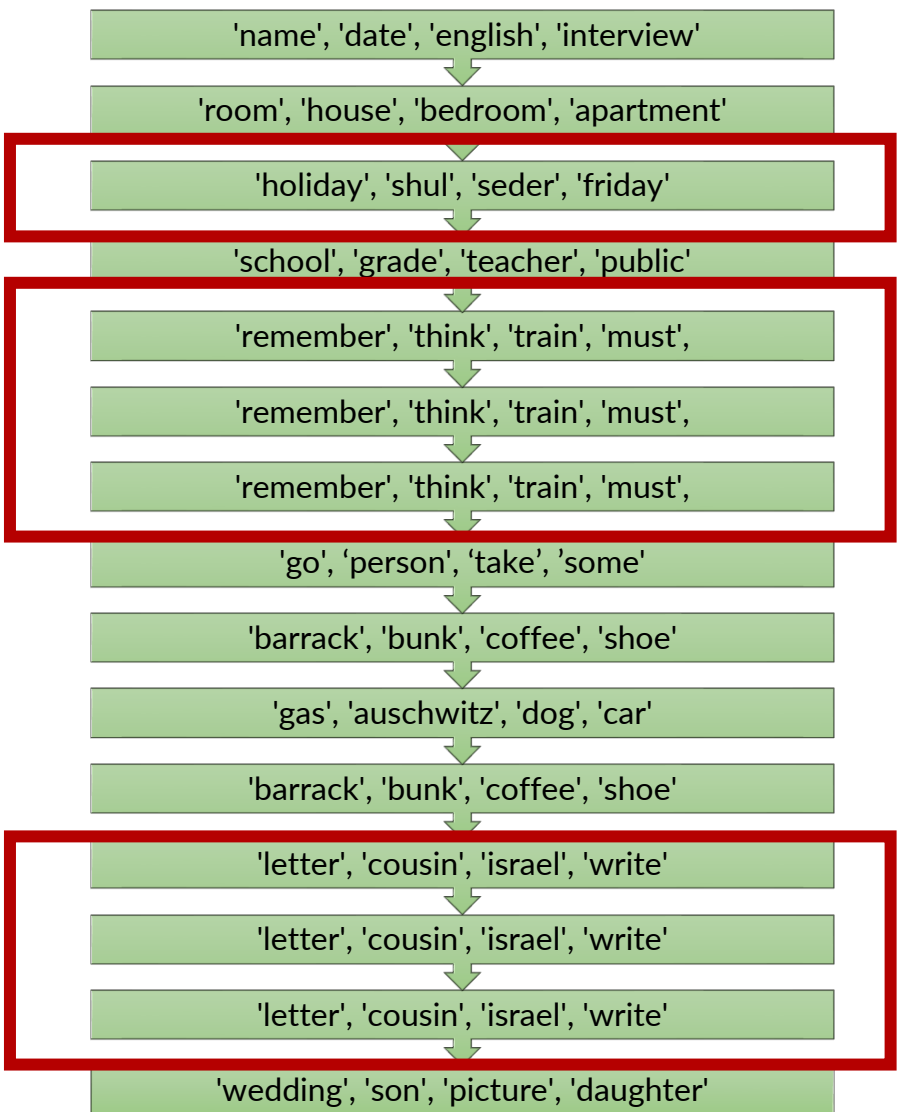
# Typical General



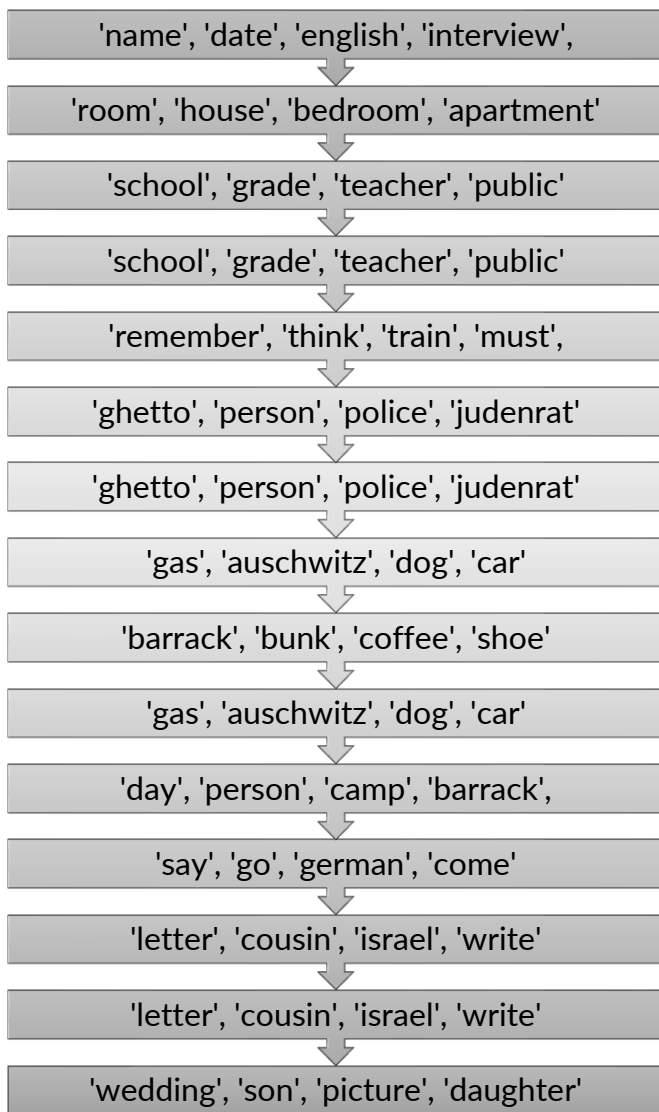
# Typical Male



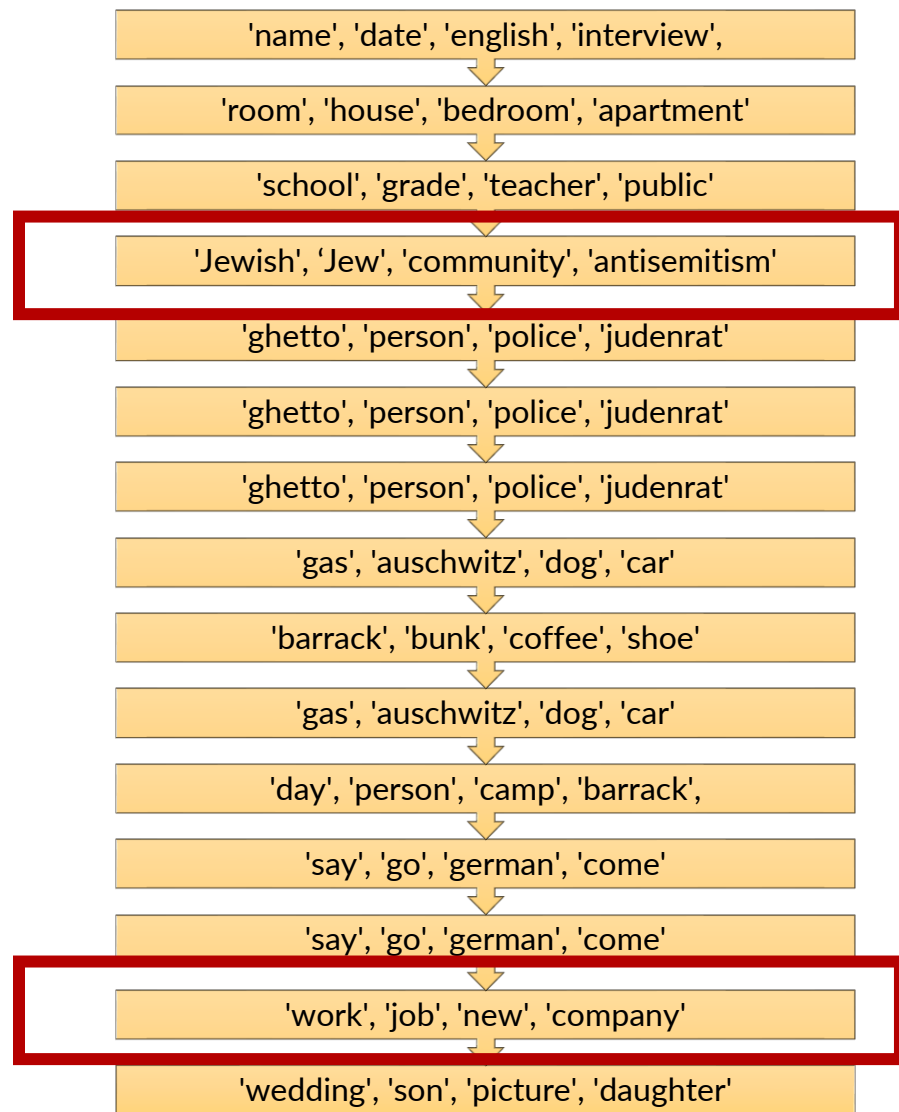
# Typical Female



# Typical General



# Typical Male



Topic/Part	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	-0.16	-0.24	-0.31	-0.34	-0.22	-0.31	-0.24	-0.46	-0.5	-0.44	-0.45	-0.62	-0.65	-0.57	-0.26
1	-0.16	-0.4	-0.36	-0.43	-0.44	-0.51	-0.76	-0.63	-0.53	-0.55	-0.63	-0.9	-0.64	-0.54	-0.42
2	0.07	0.1	0.14	0.12	0.25	0.15	0.2	0.84	0.35	0.55	0.72	0.83	0.53	0.55	-0.02
3	0.15	0.17	0.45	0.11	0.28	0.48	0.54	0.5	0.38	0.53	0.49	0.57	0.44	0.25	0.11
4	-0.04	-0.13	-0.43	-0.58	-0.69	-0.61	-0.43	-0.42	-0.23	-0.28	-0.48	-0.6	-0.6	-0.31	-0.16
5	0.03	0.01	0.09	0.18	0.21	0.07	0.24	0.25	0.24	0.14	0.1	0.07	0.11	-0.07	0.04
6	0.21	0.7	-0.21	-0.39	0.51	0.57	0.16	-0.03	-0.07	-0.01	-0.08	0.01	-0.01	0.13	0.03
7	0.04	-0.03	0.06	0.05	0.07	0.15	0.06	-0.03	-0.09	-0.22	0	0.12	0.34	-0.2	0
8	0.07	0.08	0.14	-0.04	-0.07	0.03	-0.03	-0.02	0	-0.07	0.01	0.08	-0.24	-0.38	-0.36
9	0.47	0.57	0.56	0.63	0.7	0.74	0.66	0.55	0.64	0.69	0.64	0.6	0.58	0.91	0.81
10	0.9	1.59	1.35	1.09	1	1.27	0.6	0.41	0.49	0.32	0.19	0.24	0.42	0.36	0.39
11	-1.88	0.36	0.21	0.1	0.06	0.15	0.15	0.13	0.12	0.11	0.1	0.16	0.1	0.14	0.21
12	0.39	0.84	1.12	1.33	1.43	1.5	1.36	1.47	1.48	1.84	2.18	2.11	1.94	1.73	0.99
13	0.05	0.06	0.14	0.08	0.02	0.22	0.23	0.31	0.37	0.56	0.42	0.28	0.34	0.2	0.05
14	0.08	0.23	0.42	0.31	0.35	0.51	0.31	0.42	0.37	0.45	0.39	0.35	0.71	0.58	0.64
15	-0.34	-0.76	-0.29	-0.33	-0.46	-0.39	-0.24	-0.3	-0.31	-0.33	-0.34	-0.39	-0.32	-0.21	-0.29
16	0.11	0.11	0.23	0.27	0.21	0.34	0.39	0.6	0.37	0.12	-0.09	0.34	0.21	0.04	0.03
17	-0.02	-0.07	0.12	-0.15	-0.08	-0.36	-0.28	-0.37	-0.2	-0.49	-0.53	-0.59	-0.68	-0.76	-0.29
18	0.03	0.13	0.16	0.29	0.5	0.39	0.71	0.88	0.76	0.91	1.07	1.17	0.91	0.64	0.31
19	-0.07	-0.36	-0.41	-0.53	-1.04	-0.86	-0.88	-0.96	-0.6	-0.4	-0.47	-0.48	-0.47	-0.2	-0.26
20	-0.1	-0.09	-0.01	0.11	0.07	0.08	0.18	0.19	0.09	0.05	-0.09	-0.19	0.08	-0.06	0.06
21	0.21	0.22	-0.19	-0.16	-0.25	-0.23	-0.08	-0.24	-0.26	-0.21	-0.21	-0.31	-0.21	-0.03	-0.03
22	-0.24	-0.4	-0.36	-0.47	-0.37	-0.54	-0.52	-0.55	-0.65	-0.6	-0.8	-1.13	-0.89	-0.66	-0.64
23	0.6	-0.23	-0.63	-0.51	-0.46	-0.42	-0.58	-0.4	-0.15	-0.38	-0.12	-0.06	-0.1	-0.15	-0.19
24	-0.46	-2.23	-1.44	-0.29	-0.23	-0.37	-0.35	-0.47	-0.33	-0.44	-0.29	-0.24	-0.19	-0.2	-0.14
25	-0.19	-0.41	-0.43	-0.51	-0.75	-1.04	-0.86	-1.22	-1.1	-1.16	-0.97	-0.83	-0.8	-0.57	-0.32
26	0.16	0.21	0.14	0.28	0.18	0.62	0.51	0.59	0.29	0.22	0.12	0.18	0.13	0.14	0.09
27	-0.21	-0.28	-0.23	-0.29	-0.03	-0.08	-0.28	-0.26	-0.28	-0.26	-0.31	-0.58	-0.84	-0.61	-0.56
28	0.01	0.12	0.21	0.2	0.23	0.25	0.44	0.51	0.61	0.4	0.75	0.86	0.82	0.75	0.55
29	0.18	0.05	0.03	-0.02	-0.14	-0.13	-0.14	-0.18	-0.1	-0.04	-0.1	-0.11	-0.32	-0.31	-0.22
30	-0.14	-0.18	-0.27	-0.58	-0.78	-1.12	-0.88	-0.91	-0.91	-0.72	-0.85	-0.71	-0.53	-0.53	-0.21
31	0.02	0.05	-0.01	0.17	0.02	0.1	-0.24	-0.02	0.12	0.08	0.17	-0.09	0.02	0.26	0.14
32	0.15	0.25	0.48	0.67	0.82	0.58	0.62	0.76	0.66	0.39	0.43	0.45	0.41	0.23	0.17
33	-0.08	-0.1	-0.38	-0.59	-0.71	-0.98	-0.68	-0.88	-0.94	-0.54	-0.62	-0.5	-0.55	-0.25	-0.17
34	0.13	0.05	-0.11	0.22	-0.2	-0.28	0.12	-0.08	-0.09	-0.21	-0.36	-0.12	-0.02	-0.28	-0.07

- Difference in topical distribution between male and female testimonies

- **Red** = topics more prominently mentioned by **women**

- **Green** = topics more prominently mentioned by **men**



# Conclusion

- Accounting for both **statistical categorization** of multiple testimonies and the **integrity of each** particular story
- Re-enacting **the tension** between multiplicity and singularity
- Repeatedly discovering divergences as a way to render **memory indeterminate**