Identifying Narrative Patterns and Outliers in Holocaust Testimonies Using Topic Modeling

Maxim Ifergan[†] **Renana Keydar**[‡] **, Omri Abend**[†] **Amit Pinchevski**[◇]

[†] Department of Computer Science [‡] Faculty of Law and Digital Humanities
[◇] Department of Communication and Journalism
Hebrew University of Jerusalem
{first_name}.{last_name}@mail.huji.ac.il

Abstract

The vast collection of Holocaust survivor testimonies presents invaluable historical insights but poses challenges for manual analysis. This paper leverages advanced Natural Language Processing (NLP) techniques to explore the USC Shoah Foundation Holocaust testimony corpus. By treating testimonies as structured question-and-answer sections, we apply topic modeling to identify key themes. We experiment with BERTopic, which leverages recent advances in language modeling technology. We align testimony sections into fixed parts, revealing the evolution of topics across the corpus of testimonies. This highlights both a common narrative schema and divergences between subgroups based on age and gender. We introduce a novel method to identify testimonies within groups that exhibit atypical topic distributions resembling those of other groups. This study offers unique insights into the complex narratives of Holocaust survivors, demonstrating the power of NLP to illuminate historical discourse and identify potential deviations in survivor experiences.

Keywords: Topic Modeling, Narrative, Testimonies, Holocaust

1. Introduction

In recent decades, significant efforts have been made to gather the accounts of the remaining Holocaust survivors. The passing of the last living witnesses and the beginning of the era of the post-testimony occurs simultaneously with technological developments in NLP.The wealth of testimonies in the archives presents a challenge: how to preserve the significance of individual stories within a vast collection of a thousand testimonies, while also giving voice to the collective body of testimonies in a manner that honors the individuality of each story. By employing techniques such as contextualized topic modeling and topic narrative analysis, we aim to uncover broad trends within the collection, while ensuring the preservation of the uniqueness and integrity of each personal narrative.

Despite advancements in NLP, representation of long texts still poses a challenge to state-of-the-art models (Piper et al., 2021; Castricato et al., 2021; Mikhalkova et al., 2020; Dong et al., 2023). Antoniak et al. (2019) pioneered the representation and visualization of narratives as sequences of interpretable topics. And while previous topic modeling analyses of Holocaust testimonies (Blanke et al., 2019) have provided valuable insights, they treated the corpus as a monolithic body of text, obscuring the unique narrative structure of individual testimonies. Furthermore, using non-contextualized topic modeling such as LDA ((Blei et al., 2001)) treated the text as a body of words without order. Recent advancements in topic modeling techniques such as BERTopic (Grootendorst, 2022), and other

Contextualized Topic Modeling (Bianchi et al., 2020; Angelov, 2020; Pham et al., 2023) leverage language model representation to better identify and predict the text topics. While such methods were applied to Holocaust testimonies (Wagner et al., 2022), the main focus was on the segmentation of the testimonies for topic modeling. Our contributions are as follows:

- We apply a novel contextualized topic modeling approach, BERTtopic, to holocaust testimonies, revealing the main themes and their distribution.

- We examine the evolution of topics across aligned sections of testimonies, revealing a typical narrative scheme.

- We investigate how age and gender are expressed in the narrative structure of testimonies, highlighting distinctions between survivor subgroups.

- We introduce a novel method for identifying divergent testimonies, i.e., testimonies within a given group that exhibit atypical topic distributions, resembling patterns more characteristic of other groups. We demonstrate it in a case-study of different age groups.

We note that related contributions appear in an unpublished paper of ours (under review; anonymized) that uses an earlier contextualized topic model (CTM; Bianchi et al., 2020) for a similar process. The current paper uses a better performing model (Grootendorst, 2022) regarding topic di-
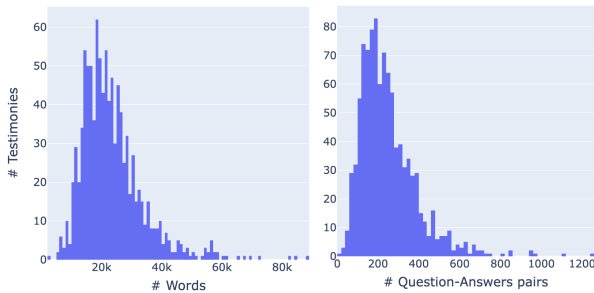
Figure 1: Testimonies number of words and number QA-pairs histogram.

versity and coherence and a more detailed and precise narrative analysis approach.

2. Corpus Level Statistics

This paper analyzes transcripts from the USC Shoah Foundation, a corpus containing 1000 oral testimonies in English. Survivors originated from over 30 countries, with a significant representation from Poland and Germany. The testimonies were recorded between 1996 and 2015, offering insights into the survivors' experiences decades after the events of the Holocaust. The length of the testimonies ranges from 3K to 88K words, with a mean length of 23K words. Each testimony contains an average of 250 questions, with the majority of question-answer pairs (95%) consisting of no more than 400 words. Fig. 1 illustrates the distribution of testimony lengths.

3. BERTopic: Topic Analysis

We use BERTopic to identify the topics within the corpus. Preprocessing involves the merging of consecutive very short sections (question-answer pairs <200 words) and the division of very long sections (>450 words) to mitigate potential outlier effects. BERTopic leverages all-MiniLM-L6-v2 (Wang et al., 2020) document embeddings and a TF-IDF based clustering approach, providing a context-aware analysis that surpasses traditional methods like LDA (Blei et al., 2001). For dimensionality reduction, UMAP (McInnes and Healy, 2018) is employed before clustering with HDBSCAN (McInnes et al., 2017). Unlike LDA, BERTopic dynamically determines the number of topics only by determining the minimum cluster size for HDBSCAN, resulting in greater flexibility. Our dataset yielded 58 topics, with approximately 4% outliers classified as "unknown topic". We set the minimum cluster size to be 50 sections.

To ensure interpretability, BERTopic extracts c-TF-IDF[1] word representations from each section's

cluster, revealing the importance of words within each topic. The most representative word is selected for initial topic representation. A domain expert then manually reviews these word sets and assigns a descriptive title to each topic, ensuring both accuracy and clarity. Notably, the topics detected by the model align with those outlined in the USC Shoah Foundation's interviewer guidelines [2] but also extend way beyond them. The guidelines encourage interviewers to ask about pre-war life, family, religion, politics, community, and experiences of antisemitism. The model's successful detection of these themes confirms the effectiveness in identifying core key topics.

4. Narrative analysis

This study analyzes individual survivor testimonies as narratives – sequences of interpretable topics (Antoniak et al., 2019). We aim to construct comprehensive narratives from the corpus testimonies that enable comparisons without sacrificing their temporal structures. Several challenges arise in this analysis. First, each testimony comprises a large number of sections (250 on average), conflicting with the direct interpretation and visualization purposes. Secondly, variations in testimony length complicate direct comparisons of narrative structures.

To address this, we divide testimonies into a fixed manageable number of parts, defining the part's theme representation as the distribution of its sections' topics. This division requires considering the trade-off between preserving temporal detail and achieving clear visualization and comparison. A large number of parts yields more nuanced narratives but risks excessive details and redundancy, whereas fewer parts allow better interpretation and visualization at the expense of obscuring finer temporal shifts in topics. After careful examination and consultation with domain experts in Holocaust studies and digital humanities, we divide testimonies into 15 equal parts. This strikes a balance between the need for detail and the goals of clear visualization and comparisons of topic distribution across parts.

4.1. Typical Testimony Narrative Schema

This section examines the most common topics covered in each part of a Holocaust survivor testimony, as well as the variation in topic representation between the different testimony parts. The analysis is based on Fig. 3, which shows the distribution of topics across the 15 parts into which each testimony was divided.

---

[1]https://maartengr.github.io/BERTopic/api/ctfidf.html

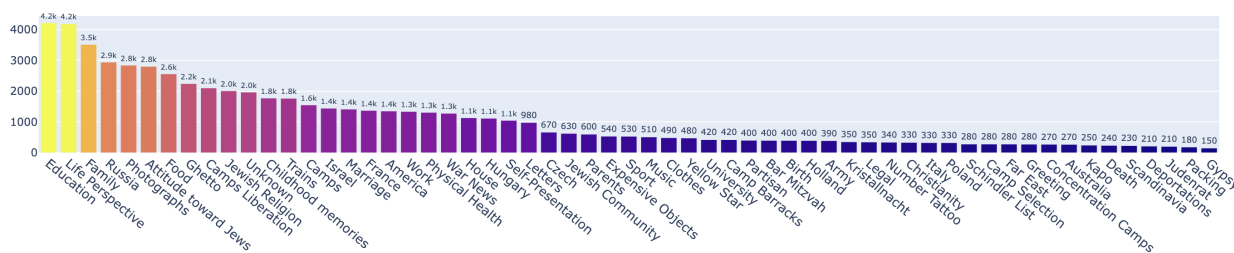[2]https://sfi.usc.edu/content/interviewer-guidelines
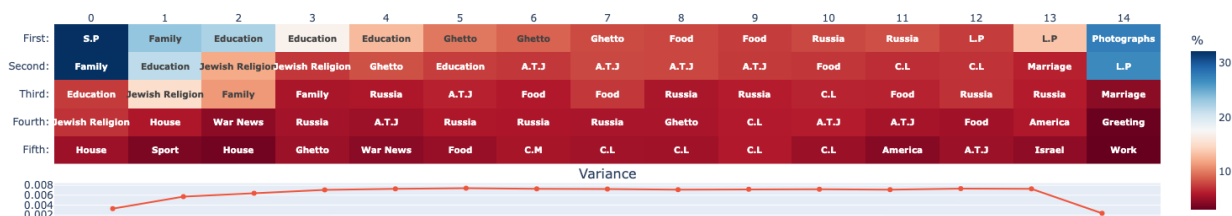
Figure 2: Corpus level QA-s topics histogram.

Figure 3: The 5 most prevalent topics and topics variance for each part. A.T.J = Attitude toward Jews, S.P = Self-Presentation, C.L = Camps Liberation, L.P = Life Perspective, and C.M = Childhood Memories

The first part of all testimonies is dominated by the topic of self-presentation followed by the family topic. It is perhaps unsurprising that many testimonies begin this way, as survivors introduce themselves and their families to the interviewer. The fact that the self-presentation topic rarely appears later may not constitute a significant finding, but it is nevertheless of importance as it validates the model's analytic capability. The next two parts of the testimonies also reveal a number of common topics, associated with the description of community life before the war. These include family, education, religion, house, and sport. The latter part also contains the topic of war news, hinting at the events to come.

In contrast to the dominance of common topics at the beginning of the testimonies, the middle parts show greater variance in the topic distributions. Each part typically features several common topics with similar percentages (around 5-15%). This might reflect the diversity of experiences among Holocaust survivors. The middle part topics vary starting with ghettos and war news to concentration/death camps and food, resolving in the rise in dominance of the camps liberation topic.

In the final parts of the testimonies narratives once again the model identifies a few dominant topics. These include interview-related topics such as presenting family pictures and discussing life after the Holocaust topics. Additionally, topics related to life after the war emerged, such as immigration and establishing work and marriage in new countries.

In conclusion, the BERTopic model successfully identifies a typical structure for Holocaust survivor testimonies, particularly at the beginning and end.

The middle sections show more variation, reflecting the different experiences of individual survivors.

4.2. Gender and Age as Expressed in Testimonies Narratives

This section introduces a method for comparing the narrative trajectories present within different groups of Holocaust survivor testimonies. We apply this method to investigate gender- and age-based differences in testimonial narratives.

To begin, we compute a typical testimony path for each group under consideration (e.g., male vs. female, young vs. older survivors). This 'typical' testimony schema represents the average topic distribution across the 15 fixed parts. Next, we perform t-tests for each part to quantify the differences in topic prevalence between groups. Topics with a substantial t-value (above 3.5%) and a low probability of such deviation arising by chance (p-value under 0.01) are flagged as characteristic of the group in which they are more prevalent.

Let's consider the age-based comparison between younger survivors, born 1925-1940, experiencing the Holocaust as children (522 testimonies), and older survivors, born 1902-1925, adults during the Holocaust (467 testimonies). Fig. 4 reveals interesting distinctions. Topics like "Childhood Memories" and "Food" dominate the middle parts of younger survivors' testimonies, while "Life Perspective" features in the final parts. Conversely, "Marriage", "Work", and "War News" are more prominent in the middle of older survivors' accounts. Interestingly, while education-related topics seem more prevalent at the beginning of older survivors' tes-
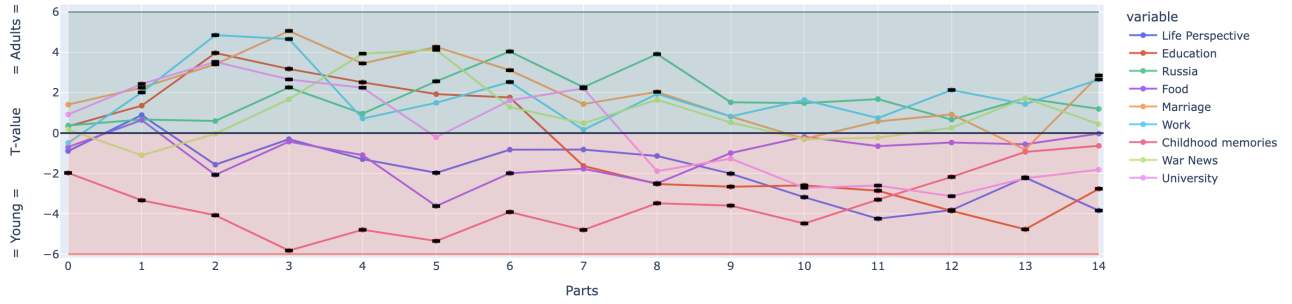
46

Figure 4: Adults vs. young survivors typical testimony t-test. The Black Point represents values with p-values under 0.01.
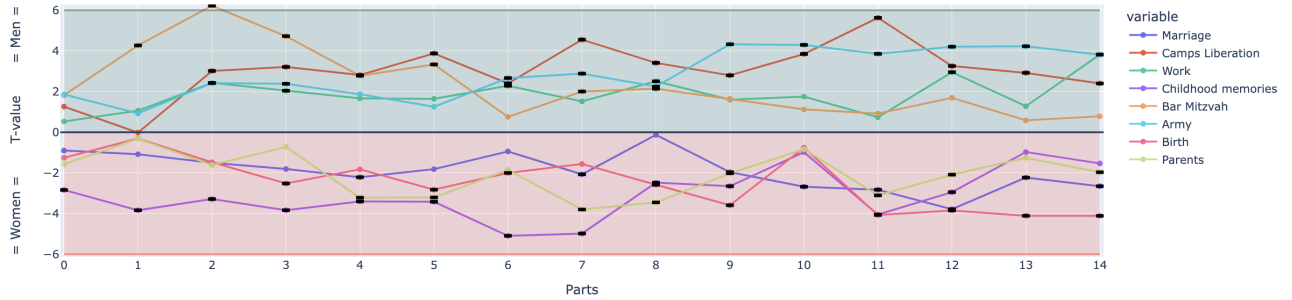

Figure 5: Men vs. women survivors typical testimony t-test. The Black Point represents values with p-values under 0.01.

timonies, they tend to re-emerge near the end for the younger group. Turning to the gender-based analysis, with a balanced corpus of 531 male and 469 female testimonies, Fig. 5 highlights potential differences. Topics like "Bar Mitzvah", "Army", "Camp Liberation", and "Work" are more characteristic of men's testimonies. In contrast, "Birth", "Childhood Memories", "Parents", and "Marriage" are more prevalent in women's testimonies. This analysis reveals how men and women may structure their narratives differently, particularly in the middle sections of their testimonies.

The USC Shoah Foundation's interviewer guidelines do not provide specific instructions for ordering topics or tailoring questions based on the subject's age or gender. This suggests that the observed differences in narrative structure between these groups are not a direct result of the guidelines. Rather, they may stem from the interviewers' individual approaches or the survivors' unique experiences and perspectives.

5. Exploratory Study Identifying Diverging Narratives

This study introduces a novel method for identifying testimonies within a specific group that exhibit topic distribution patterns more characteristic of another group. Our goal is to pinpoint narratives that stand out as atypical within their designated

category. We achieve this by defining a scoring function that quantifies the similarity between a testimony's topic distribution and the typical narrative patterns of a different group. Let us formalize the scoring function which yields a high score for testimonies from A that resemble the narrative patterns typical of B. Let $t = (t_1, t_2, ..., t_{15})$ represent the testimony's topic distributions from group A, where each t_i is a vector of topic probabilities for part i. And, $C_A = \{(i_1, j_1), (i_2, j_2), ..., (i_n, j_n)\}$ denotes the characteristic topic-part pairs for group A, where i_x is a part index and j_x is a topic index. These pairs have high t-values (>3.5) and low p-values (<0.01) in the group comparison. C_B similarly represents the characteristic topic-part pairs for group B.

$$R_B = \sum_{(i,j) \in C_B} t_i[j] \cdot |Tvalue_B(i,j)|$$

$$R_A = \sum_{(i,j) \in C_A} t_i[j] \cdot |Tvalue_A(i,j)|$$

$$S(t, C_A, C_B) = R_B - R_A$$

Finally, we apply an argmax operation to spot those testimonies exhibiting the highest resemblance to group B's typical narrative:

$$argmax_{t \in A} S(t, C_A, C_B)$$

When comparing older and younger survivor groups, Fig. 7 presents the distribution of resem-
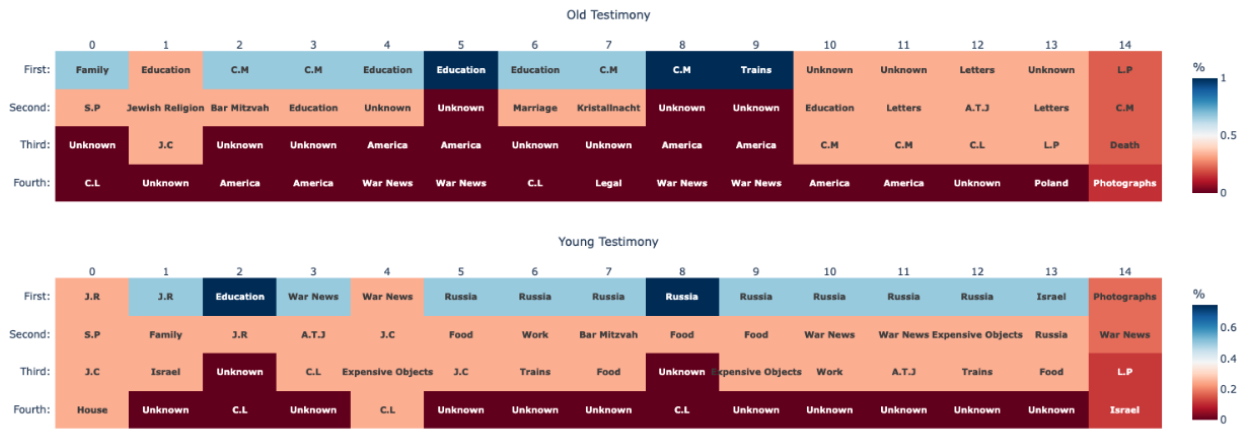
47

Figure 6: Outliers testimonies. A.T.J = Attitude toward Jews, S.P = Self-Presentation, C.L = Camps Liberation, L.P = Life Perspective, J.C = Jewish Community, J.R = Jewish Religion, and C.M = Childhood Memories
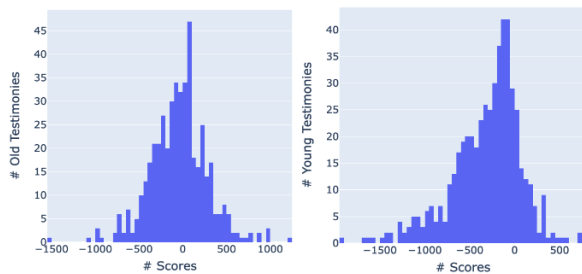

Figure 7: Testimonies scores histogram.

blance scores for testimonies within the groups. The uneven distribution favoring negative scores reveals that higher scores tend to be related to non-conforming narratives. Using this method, Fig. 6 highlights two specific examples: a younger survivor whose narrative strongly resembles the older group, and vice-versa emphasizing topics characteristic of the opposite group.

6. Conclusion and Future Work

This study applies NLP techniques to explore the complex narratives within the USC Shoah Foundation's Holocaust testimonies. Contextualized topic modeling with BERTopic reveals key themes and their distributions within the corpus. And, by aligning testimonies into fixed parts, we unveiled a common narrative trajectory along with age- and gender-based variations. Our method detects divergent testimonial narratives, identifying those within one group that exhibit topic patterns characteristic of another group. Future Work will extend the analysis by comparing survivor narratives across corpora[3] and other testimonial archives to identify both shared and distinct narratives patterns.

---
[3]Yale Fortunoff Archive

7. Acknowledgements

8. Bibliographical References

Dimitar Angelov. 2020. Top2Vec: Distributed Representations of Topics. *ArXiv*, abs/2008.09470.

Maria Antoniak, David M. Mimno, and Karen E. C. Levy. 2019. Narrative Paths and Negotiation of Power in Birth Stories. *Proceedings of the ACM on Human-Computer Interaction*, 3:1 – 27.

Federico Bianchi, Silvia Terragni, and Dirk Hovy. 2020. Pre-training is a Hot Topic: Contextualized Document Embeddings Improve Topic Coherence. In *Annual Meeting of the Association for Computational Linguistics*.

Tobias Blanke, Michael Bryant, and Mark Hedges. 2019. Understanding memories of the Holocaust - A new approach to neural networks in the digital humanities. *Digit. Scholarsh. Humanit.*, 35:17–33.

David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2001. Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3:993–1022.

Louis Castricato, Stella Biderman, David Thue, and Rogelio Cardona-Rivera. 2021. Towards

a model-theoretic view of narratives. In *Proceedings of the Third Workshop on Narrative Understanding*, pages 95–104, Virtual. Association for Computational Linguistics.

Zican Dong, Tianyi Tang, Lunyi Li, and Wayne Xin Zhao. 2023. A survey on long text modeling with transformers. *ArXiv*, abs/2302.14502.

Maarten R. Grootendorst. 2022. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *ArXiv*, abs/2203.05794.

Leland McInnes and John Healy. 2018. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *ArXiv*, abs/1802.03426.

Leland McInnes, John Healy, and S. Astels. 2017. HDBSCAN: Hierarchical Density Based Clustering. *J. Open Source Softw.*, 2:205.

Elena Mikhalkova, Timofei Protasov, Polina Sokolova, Anastasiia Bashmakova, and Anastasiia Drozdova. 2020. Modelling narrative elements in a short story: A study on annotation schemes and guidelines. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 126–132, Marseille, France. European Language Resources Association.

Chau Minh Pham, Alexander Miserlis Hoyle, Simeng Sun, and Mohit Iyyer. 2023. TopicGPT: A Prompt-based Topic Modeling Framework. *ArXiv*, abs/2311.01449.

Andrew Piper, Richard Jean So, and David Bamman. 2021. Narrative theory for computational narrative understanding. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 298–311, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Eitan Wagner, Renana Keydar, Amit Pinchevski, and Omri Abend. 2022. Topical Segmentation of Spoken Narratives: A Test Case on Holocaust Survivor Testimonies. In *Conference on Empirical Methods in Natural Language Processing*.

Wenhui Wang, Furu Wei, Li Dong, Hangbo Bao, Nan Yang, and Ming Zhou. 2020. MiniLM: Deep Self-Attention Distillation for Task-Agnostic Compression of Pre-Trained Transformers. *ArXiv*, abs/2002.10957.

A. Topic list

Topic Title	Top-15 words in topic
Life Perspective	children', 'think', 'holocaust', 'thank', 'say', 'life', 'grandchildren', 'years', 'want', 'experiences', 'people', 'much', 'god', 'thats', 'never'
Photographs	picture', 'taken', 'thank', 'photograph', 'left', 'name', 'right', 'photo', 'next', 'daughter', 'son', 'crew', 'sister', 'year', 'wife'
Family	name', 'mother', 'father', 'born', 'family', 'mothers', 'remember', 'sister', 'brother', 'fathers', 'lived', 'years', 'sisters', 'brothers', 'grandfather'
Education	school', 'jewish', 'antisemitism', 'hitler', 'jews', 'remember', 'teacher', 'hebrew', 'yiddish', 'german', 'friends', 'language', 'teachers', 'polish', 'years'
Russia	russian', 'russians', 'russia', 'people', 'away', 'germans', 'army', 'take', 'train', 'told', 'come', 'want', 'food', 'says', 'already'
Jewish Religion	synagogue', 'holidays', 'used', 'shabbos', 'remember', 'religious', 'shabbat', 'shul', 'friday', 'father', 'passover', 'home', 'holiday', 'jewish', 'family'
Self-Presentation	name', 'born', 'birth', 'spell', 'interview', 'please', 'date', 'english', 'today', '1997', 'conducting', 'interviewer', 'language', '1998', 'maiden'
Ghetto	ghetto', 'people', 'germans', 'jews', 'work', 'food', 'place', 'little', 'lived', 'remember', 'jewish', 'house', 'already', 'street', 'away'
Food	food', 'bread', 'soup', 'eat', 'potatoes', 'little', 'day', 'used', 'people', 'water', 'piece', 'work', 'something', 'hungry', 'put'
Marriage	married', 'wedding', 'met', 'husband', 'wife', 'marry', 'israel', 'years', 'meet', 'name', 'marriage', 'laughs', 'come', 'want', 'family'
Israel	israel', 'palestine', 'zionist', 'british', 'kibbutz', 'organization', 'jewish', 'people', 'army', 'hebrew', 'tel', 'war', 'come', 'also', 'years'
France	french', 'france', 'paris', 'belgium', 'people', 'brussels', 'germans', 'train', 'german', 'antwerp', 'war', 'border', 'think', 'vichy', 'little'
Physical Health	hospital', 'sick', 'doctor', 'typhus', 'camp', 'fever', 'people', 'doctors', 'typhoid', 'day', 'couldnt', 'food', 'take', 'put', 'work'
Hungary	hungarian', 'hungary', 'hungarians', 'budapest', 'jews', 'romania', 'romanian', 'jewish', 'people', 'war','many', 'labor', 'started', 'germans', 'father'
America	ship', 'states', 'united', 'york', 'boat', 'new', 'visa', 'america', 'come', 'arrived', 'american', 'quota', 'affidavit', 'days', 'papers'
Camps	auschwitz', 'birkenau', 'camp', 'people', 'gas', 'work', 'saw', 'day', 'barracks', 'train', 'see', 'barrack', 'right', 'knew', 'women'
House	room', 'house', 'apartment', 'kitchen', 'rooms', 'lived', 'bedroom', 'big', 'living', 'floor', 'remember', 'home', 'dining', 'describe', 'used'
Camps Liberation	camp', 'german', 'american', 'germans', 'saw', 'americans', 'day', 'see', 'people', 'soldiers','planes', 'war', 'prisoners', 'started', 'liberated'
Trains	train', 'people', 'cattle', 'trains', 'wagon', 'camp', 'march', 'long', 'water', 'many', 'station', 'car', 'food', 'days', 'see'
Work	job', 'business', 'worked', 'work', 'money', 'working', 'company', 'store', 'bought', 'make', 'years', 'week', 'started', 'factory', 'want'
Attitude Toward Jews	says', 'jewish', 'told', 'come', 'jews', 'mother', 'want', 'away', 'german', 'see', 'knew', 'people', 'look', 'gave', 'man'
Music	music', 'sing', 'singing', 'song', 'songs', 'opera', 'played', 'piano', 'sang', 'play', 'remember', 'used', 'yiddish', 'voice', 'laughs'
Childhood memories	mother', 'father', 'remember', 'think', 'parents', 'see', 'sister', 'never', 'really', 'knew', 'always', 'happened', 'mean', 'couldnt', 'tell'
Czech	prague', 'czech', 'czechoslovakia', 'theresienstadt', 'train', 'people', 'see', 'come', 'army', 'left', 'knew', 'stayed', 'transport', 'home', 'want'
Letters	letters', 'letter', 'wrote', 'mother', 'war', 'sister', 'parents', 'sent', 'write', 'cross', 'found', 'red', 'brother', 'knew', 'family'
Bar Mitzvah	bar', 'mitzvah', 'remember', 'torah', 'synagogue', 'mitzvahed', 'school', 'jewish', 'rabbi', 'hebrew', 'shul', 'family', 'mitzvahs', 'father', '13'

Topic Title	Top-15 words in topic
War News	'radio', 'poland', 'jews', 'polish', 'war', 'news', 'jewish', 'people', 'german', 'knew', 'germans', 'germany', 'heard', 'poles', 'warsaw'
Greeting	'thank', 'yourn', 'much', 'muchrn', 'sharing', 'welcome', 'welcomern', 'add', 'youre', 'mrs', 'testimony', 'concludes', 'thanks', 'say', 'want'
Legal	'trial', 'trials', 'court', 'nuremberg', 'crimes', 'witnesses', 'case', 'courtroom', 'witness', 'judge', 'cases', 'justice', 'defense', 'evidence', 'interpreter'
Sport	'play', 'soccer', 'used', 'played', 'sports', 'school', 'games', 'remember', 'sport', 'friends', 'swimming', 'playing', 'ball', 'club', 'liked'
Partisan	'partisans', 'partisan', 'group', 'russian', 'germans', 'forest', 'killed', 'people', 'army', 'fighting', 'woods', 'fight', 'food', 'knew', 'thats'
Far East	'shanghai', 'japanese', 'chinese', 'china', 'japan', 'hongkew', 'refugees', 'people', 'war', 'see', 'money', 'ship', 'american', 'boat', 'harbor'
Yellow Star	'star', 'wear', 'yellow', 'wearing', 'stars', 'jewish', 'david', 'jews', 'remember', 'armband', 'wore', 'germans', 'jew', 'people', 'think'
Number Tattoo	'number', 'tattooed', 'tattoo', 'numbers', 'auschwitz', 'camp', 'arm', 'barracks', 'given', 'birkenau', 'tattooing', 'people', 'triangle', 'put', 'prisoners'
Camp Selection	'mengele', 'selection', 'auschwitz', 'gas', 'camp', 'right', 'people', 'saw', 'side', 'twins', 'selected', 'see', 'left', 'told', 'experiments'
Schindler List	'schindler', 'factory', 'list', 'schindlers', 'plaszow', 'oskar', 'goeth', 'brxfcnnlitz', 'people', 'camp', 'brunnlitz', 'inaudible', 'working', 'knew', 'auschwitz'
Jewish Community	'jewish', 'town', 'population', 'jews', 'lived', 'community', 'city', 'people', 'synagogue', 'families', 'big', 'school', 'area', 'lot', 'business'
Expensive Objects	'jewelry', 'money', 'gold', 'ring', 'things', 'silver', 'coins', 'hide', 'give', 'take', 'put', 'buy', 'remember', 'sold', 'whatever'
Army	'army', 'training', 'basic', 'infantry', 'drafted', 'fort', 'draft', 'corps', 'sergeant', 'officer', 'unit', 'military', 'citizen', 'british', 'service'
Concentration Camps	'bergenbelsen', 'buchenwald', 'camp', 'people', 'belsen', 'barrack', 'dead', 'arrived', 'saw', 'barracks', 'remember', 'block', 'liberated', 'prisoners', 'many'
Kristallnacht	'kristallnacht', 'synagogue', 'november', 'happened', '1938', 'remember', 'father', 'arrested', 'night', 'school', 'mother', 'home', 'glass', 'jewish', 'day'
Italy	alian', 'italy', 'italians', 'athens', 'switzerland', 'rome', 'mussolini', 'people', 'camp', 'modena', 'train', 'germans', 'bari', 'chichibo', 'nonantola'
Birth	'baby', 'hospital', 'doctor', 'child', 'pregnant', 'husband', 'mother', 'father', 'sick', 'born', 'cancer', 'died', 'told', 'home', 'never'
Parents	'father', 'mother', 'come', 'says', 'take', 'little', 'told', 'saw', 'knew', 'see', 'place', 'away', 'want', 'ill', 'thought'
Scandinavia	'sweden', 'denmark', 'danish', 'swedish', 'danes', 'copenhagen', 'stockholm', 'king', 'people', 'malmo', 'jews', 'goteborg', 'jewish', 'swedes', 'government'
University	'university', 'college', 'school', 'degree', 'years', 'new', 'high', 'york', 'job', 'engineering', 'masters', 'worked', 'work', 'social', 'graduate'
Australia	'australia', 'melbourne', 'australian', 'sydney', 'boat', 'fremantle', 'ship', 'arrived', 'come', 'australians', 'people', 'job', 'english', 'friends', 'years'
Holland	'dutch', 'holland', 'amsterdam', 'westerbork', 'jews', 'people', 'rotterdam', 'german', 'camp', 'germans', 'war', 'jewish', 'happened', 'nazis', 'germany'
Christianity	'catholic', 'church', 'priest', 'baptized', 'communion', 'religion', 'jewish', 'catholicism', 'mother', 'school', 'convert', 'never', 'convent', 'prayers', 'think'
Deportations	'deported', 'deportation', 'people', 'deportations', 'jews', 'day', 'work', 'ghetto', 'happened', 'heard', 'think', 'told', 'believe', 'sent', 'knew'
Clothes	'shoes', 'clothes', 'shower', 'hair', 'shaved', 'camp', 'clothing', 'naked', 'put', 'auschwitz', 'pair', 'showers', 'barracks', 'barrack', 'women'
Packing	'take', 'clothing', 'little', 'suitcase', 'remember', 'things', 'knitting', 'left', 'made', 'used', 'yarn', 'everything', 'mother', 'maybe', 'something'
Judenrat	'judenrat', 'judenrate', 'ghetto', 'police', 'jewish', 'people', 'jews', 'germans', 'council', 'orders', 'killed', 'away', 'gestapo', 'work', 'town'
Gypsy	'gypsies', 'gypsy', 'camp', 'block', 'people', 'auschwitz', 'mengele', 'lager', 'saw', 'jews', 'gassed', 'barracks', 'eichman', 'see', 'told'

Topic Title	Top-15 words in topic
Kapo	'kapos', 'kapo', 'prisoners', 'camp', 'block', 'barrack', 'people', 'german', 'political', 'killed', 'work', 'put', 'prisoner', 'somebody', 'remember'
Poland	'warsaw', 'polish', 'lublin', 'place', 'train', 'find', 'army', 'walked', 'also', 'anyway', 'station', 'says', 'poles', 'people', 'praga'
Camp Barracks	'barrack', 'barracks', 'beds', 'bunk', 'people', 'bunks', 'slept', 'camp', 'wooden', 'straw', 'bed', 'sleeping', 'cold', 'sleep', 'little'
Death	'cemetery', 'buried', 'grave', 'jewish', 'people', 'put', 'died', 'stone', 'family', 'find', 'mass', 'tombstones', 'graves', 'place', 'stones'
Unknown	'people', 'remember', 'jewish', 'see', 'think', 'father', 'come', 'knew', 'little', 'german', 'camp', 'day', 'mother', 'told', 'thats