

# 홀로코스트 생존자들의 감정과 기억: GPT 활용 증언 분석

나영은 (역사학과 20204362)

[요약] 이 프로젝트는 GPT를 활용하여 홀로코스트 아우슈비츠 생존자들의 증언을 분석하고, 감정 변화와 주요 경험 요소를 탐구하는 것을 목표로 하였다. 데이터 기반 접근법으로 폴 에크만의 6가지 감정을 중심으로 증언을 분석하고, 워드클라우드 시각화를 통해 공통 주제를 도출하였다.

## 1. 서론

### 1) 연구 목적

홀로코스트 생존자들의 증언은 역사적, 인문학적 가치를 지닌 중요한 기록이다. 역사 전공 인공지능인문학 학부 연구생으로서 증언 데이터를 GPT를 활용하여 분석하여 감정 변화와 주요 경험을 정밀하게 파악하는 것이 목표이다. 또, 이 연구를 통해 인공지능 기술을 인문학적 연구에 어떻게 활용할지 고민하며 접근한다.

### 2) 연구 방법론

아우슈비츠 수용소 생존자 10인의 증언 텍스트를 대상으로 감정 분석을 수행한다.

① 감정 분석 : 폴 에크만의 6가지 기본 감정 (행복, 슬픔, 두려움, 분노, 놀라움, 혐오)을 기준으로 증언 텍스트를 분석한다. GPT 모델을 활용하여 텍스트에서 나타나는 감정적 표현을 단계별로 분석하고, 각 단계에서의 감정 변화를 파악한다. 이 분석은 증언을 다섯 단계(어린 시절, 전쟁 전, 전쟁 초기, 아우슈비츠, 전후)로 나누어 진행된다.

② 단어 빈도 분석: 생존자들이 사용한 주요 단어들의 빈도를 분석하여, 증언에서 자주 언급된 경험 요소와 감정 상태를 시각적으로 표현하고자 한다.

### 3) 데이터 출처 및 분석 도구

미국 홀로코스트 기념관에서 제공하는 아우슈비츠 생존자 10인의 증언 텍스트를 활용한다. 이 텍스트 데이터는 각 생존자가 아우슈비츠에서 겪은 경험과 감정을 기록한 인터뷰 자료로 구성되어 있다.

분석 도구로는 GPT 모델을 사용하여 텍스트의 감정적 표현을 자동으로 분석하고, 각 단계별 감정 변화 추이를 탐구한다. 단어 빈도 분석 및 시각화는 Python의 기본 라이브러리를 활용한다. 이러한 도구들을 통해 정량적이고 체계적인 분석을 가능하게 하여, 생존자 증언 속에 숨겨진 감정적 흐름과 경험을 도출하고자 한다.

## 2. 본론

### 1. 활용 데이터

미국 홀로코스트 기념관에서 제공하는 아우슈비츠 수용소 생존자 10인 증언 인터뷰 텍스트로, 미국 홀로코스트 기념관에서 구술된 내용으로 그대로 적은 텍스트 pdf 파일이다.

〈표1〉 데이터 요약

1	Michael Vogel
2	Morris Kornberg
3	Hana Bruml
4	Fritzie Weiss Fritshall
5	Ruth Borsos
6	Lily Malnik
7	Leo Schneiderman
8	Bernard Pasternak
9	Madeline Deutsch
10	Boleslaw Brodecki

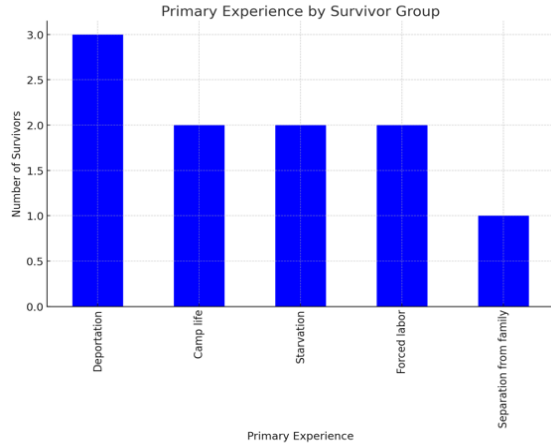
### 2. 공통 테마 분석

10명의 생존자들의 증언에서 자주 언급된 주요 경험 요소를 분석한 결과, '가족과의 이별', '강제 수용소 생활', '기아', '해방' 이 공통적으로 많이 언급되었다. 이는 생존자들이 겪은 경험 중 가장 중요한 요소임을 나타낸다.

Common Themes in Survivor Testimonies

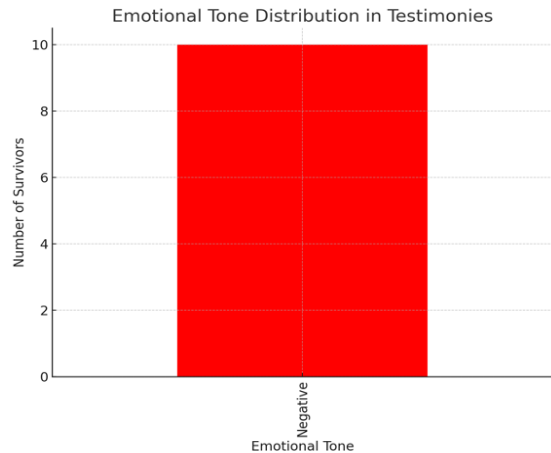
Forced labor  
Liberation  
Separation from family  
Starvation  
Camp life  
Deportation

〈그림1〉 공통 테마 분석 시각화



〈그림2〉 공통 경험 분석

감정적 톤 분석 결과, 모든 증언에서 부정적인 감정 표현(고통, 두려움, 슬픔)이 강하게 나타난 것을 확인할 수 있었다.



〈그림3〉 감정 톤 분석

3. 폴 에크만의 감정 이론 기반 감정 변화 추이 분석  
폴 에크만의 6가지 기본 감정(행복, 슬픔, 두려움, 분노, 놀라움, 혐오)을 기반으로, 홀로코스트 아우슈비츠 생존자 10인 인터뷰 텍스트를 GPT를 활용해 분석한다.

인터뷰 내용은 삶을 다섯 단계(어린 시절, 전쟁 전, 전쟁 초기, 아우슈비츠, 전후)로 나누어 각 단계에서의 주요 감정 변화를 파악하는 것을 목표로 한다.

각 단계별로 6가지 감정의 강도를 0에서 10까지의 척도로 평가하여 그래프로 시각화한다. 각 감정이 시간에 따라 어떻게 변했는지를 명확하게 파악하고자 한다.

#### 1) GPT 프롬프트 활용 메뉴얼

- 10인의 텍스트 데이터 pdf 각각의 프롬프트의 입력
- 이 텍스트 데이터는 홀로코스트 아우슈비츠 수용소 생존자 증언 인터뷰 내용이다. 폴 에크만 6가

지 감정을 기반으로 감정 분석을 하는 프로젝트이다.

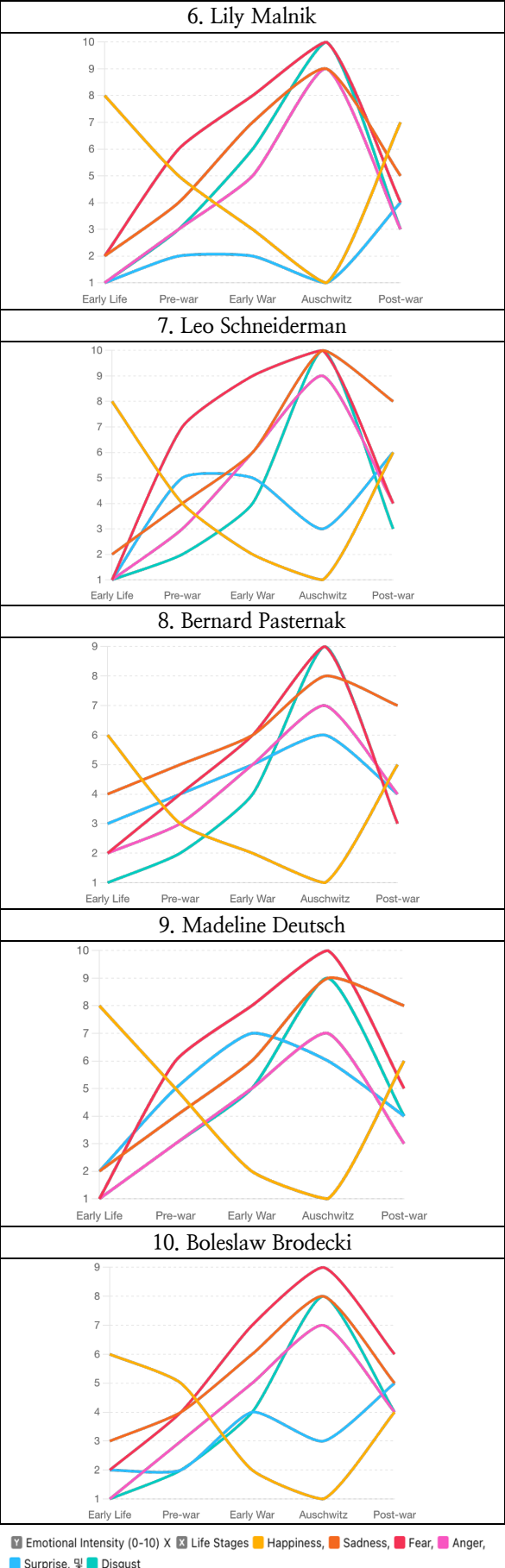
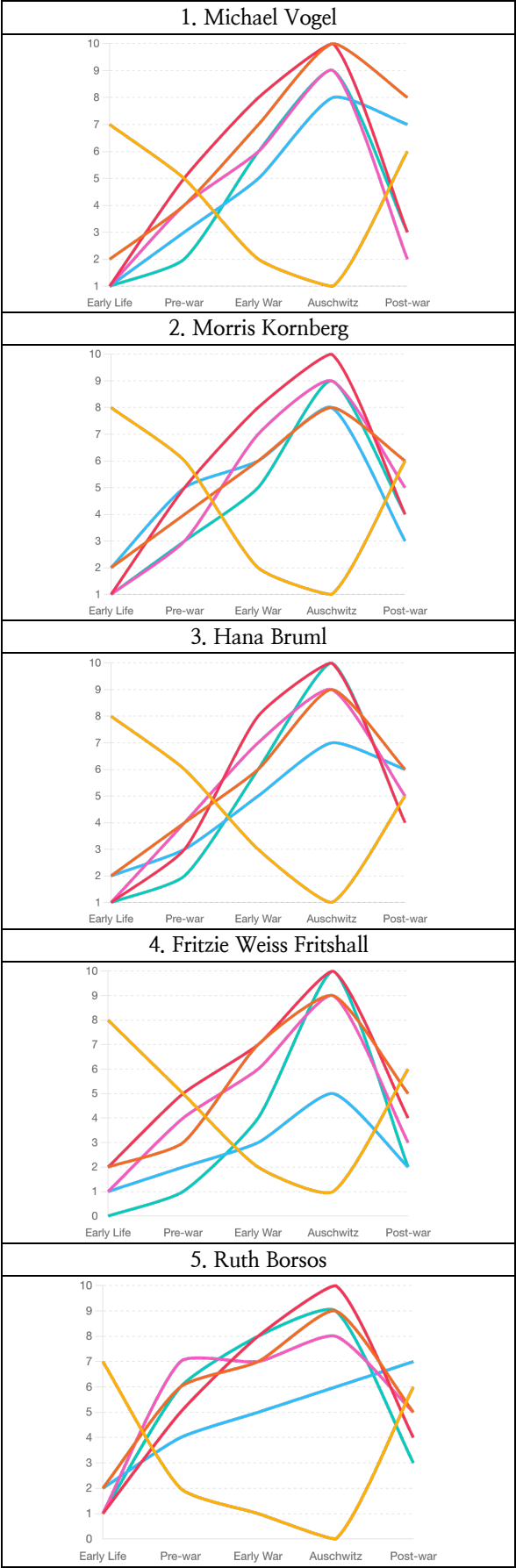
3. Early life/ pre-war/early war/ Auschwitz/ post war 5단계로 생애를 나누고 주요 감정 및 내용을 파악해라.
4. 5단계에 따른 6가지 감정 변화를 그래프로 시각화해라.
5. 감정의 강도를 0-10 상대적인 수치로 표현해서 그래프를 그려라.

#### 2) 결과

위 과정을 거쳐 아래 첨부한 <표2>의 결과를 얻었다. 10인이 비슷한 양상을 보이는 것을 확인할 수 있었다. 전쟁이 시작되면서 행복은 급격히 감소하고, 슬픔과 두려움은 점진적으로 증가하여 아우슈비츠 시기에 최고조에 달한다. 아우슈비츠 시기에서는 두려움, 분노, 슬픔, 혐오가 주요 감정으로 나타나며, 이 시기는 감정적으로 매우 극심한 변화가 일어난 시기로 볼 수 있다. 전후(Post-war) 시기에는 행복이 다시 회복되지만, 슬픔과 분노는 여전히 높은 수준으로 남아있다. 이는 전쟁과 아우슈비츠 경험이 이들의 삶에 깊은 상처를 남겼음을 시사한다.

놀라움의 정도에 대해서는 개개인이 다른 양상을 보이는데 이 부분에 대해서는 개별적인 심도 깊은 연구가 필요할 듯하다.

<표 2> 10인 감정 변화 추이 분석 그래프



#### 4. 단어 사용 빈도 및 시각화

생존자들이 인터뷰에서 사용한 단어들의 빈도를 분석하여, 그들이 인터뷰 도중 어떤 단어를 많이 사용했는지 파악하고자 했다, 이를 통해 강조된 주제나 감정을 파악하고 시각적으로 표현하는 것을 목표로 했다.

##### 1) 과정

데이터 수집 및 전처리:

인터뷰 데이터를 불러오고, 전처리를 통해 분석에 적합한 형태로 변환한다. 불필요한 문자를 제거하고, 공통적으로 사용되는 불용어(stop words)를 필터링하는 과정을 거친다.

단어 빈도 계산:

전처리된 인터뷰 데이터를 기반으로 각 단어의 빈도를 계산한다.

시각화 & 워드클라우드 생성 :

계산된 단어 빈도를 바탕으로 막대 그래프로 데이터를 시각적으로 표현한 후, 인터뷰에서 사용된 단어들의 빈도를 바탕으로 워드클라우드를 생성한다. 생존자들이 주로 언급한 단어와 감정을 한눈에 파악할 수 있게 한다.

##### 2) 결과 : 다음 페이지 표에 첨부

##### 3) 한계

인터뷰 질문의 다양성과 사건 전개 중심의 서술로 유의미한 결론을 도출하기에 어려움이 있었다. 중요 단어에 가중치를 두는 방식으로 추후 접근해 볼 필요성이 있다.

#### 5.해방 후 심리적 회복 분석

해방 당시 생존자의 나이, 건강 상태, 성별, 해방 후 정착지, 고향 복귀 여부, 가족과 재회 여부 등 여러 변수들이 해방 후 심리적 회복에 미치는 영향을 탐구하고자 10인의 데이터를 분석해보았다. 고향 복귀와 심리적 회복, 가족 재결합과 심리적 회복의 연관성을 파악하고자 했다.

##### 1) 과정

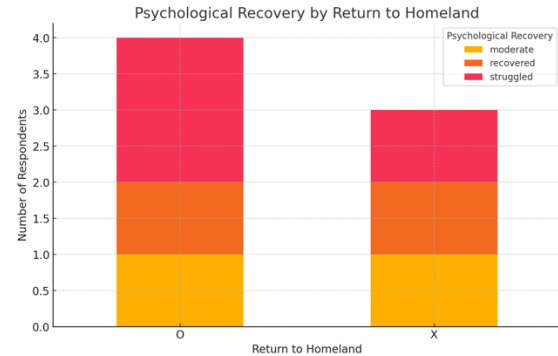
10인의 해방 당시 나이, 성별, 고향, 고향 복귀 여부, 전후 목적지, 전후 직업, 가족과 재회 여부, 수용 생활 당시 업무 및 직책, 해방 당시 건강 상태, 심리적 회복의 상대적 정도를 파악하여 데이터프레임을 완성한다.

〈표 2〉 10인 해방 후 심리적 회복 분석 위한 데이터 프레임

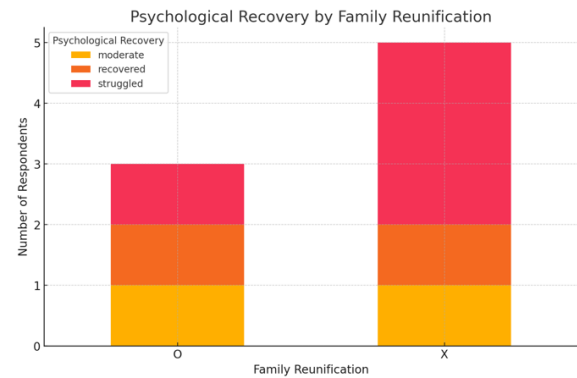
	age (At lib)	gender	homeland	return to h	post-war d	occupation	family reun	camp role	post war su	health state	Psychologic
1	21	male	Czechoslovak	X	USA		X	forced labor	BAD	struggled	
2	27	male	Poland		USA			forced labor	BAD	moderate	
3	23	female	Czechoslovak	O		Nurse	X		BAD	struggled	
4	18	female	Czechoslovak	O	USA		O	forced labor	BAD	moderate	
5	22	female	Germany	X	USA		O	tailor	BAD	recovered	
6	17	female	Belgium	O	USA		O	chef assist	BAD	struggled	
7	24	male	Poland	X	USA	Jewish Centr	X	tailor	O	BAD	moderate
8	25	male	Romania		Italy			forced labor	BAD	struggled	
9	15	female	Czechoslovak	O	USA	tailor	X	forced labor	O	BAD	recovered
10	24	male	Poland				X	forced labor	BAD	struggled	

심리적 회복의 상대적 정도를 moderate, recovered, struggled로 나누고 원인 변수를 찾고자 하였다.

##### 2) 결과



〈그림4〉 고향 복귀 여부와 심리적 회복 정도 관계 분석



〈그림5〉 가족 재회 여부와 심리적 회복 정도 관계 분석

분석 결과, 고향 복귀 여부와 가족 재회 여부 두 변수와 심리적 회복의 정도 간의 상관관계를 파악할 수 있었다.

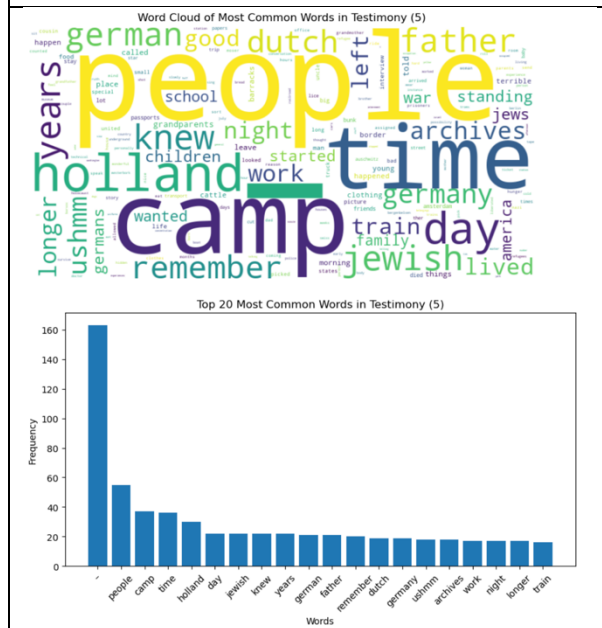
##### 3) 한계

작은 샘플의 크기, 누락된 데이터, 주관적인 변수 등의 여러 이유로 유의미한 결론을 도출하기에 어려움이 있다. 더 크고 다양한 샘플과 더 포괄적인 데이터를 사용하여 미래 연구를 진행할 수 있을 것 같고, 이를 통해 더 강력하고 일반화 가능한 결론을 도출할 수 있을 것으로 기대한다.

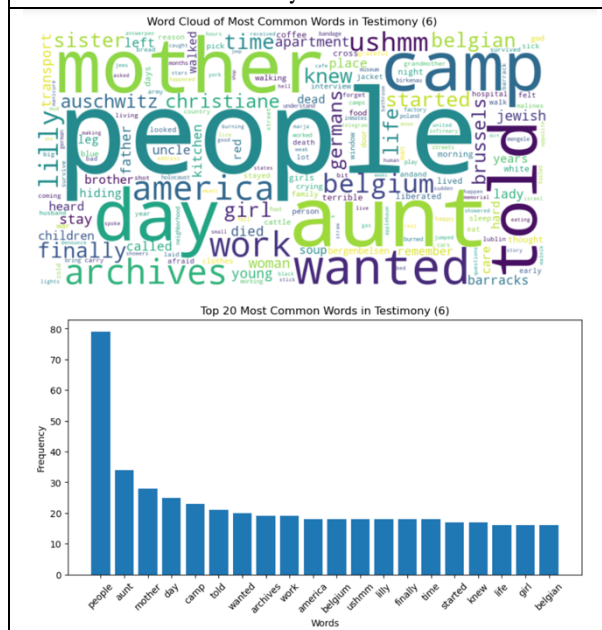




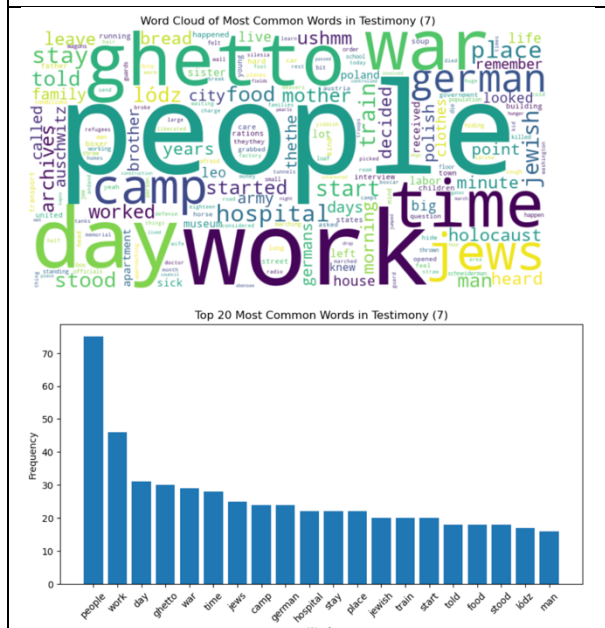
### 5. Ruth Borsos



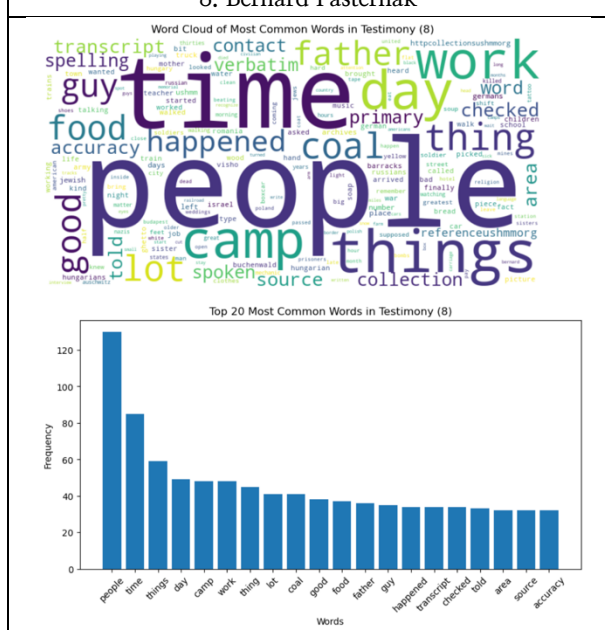
### 6. Lily Malnik



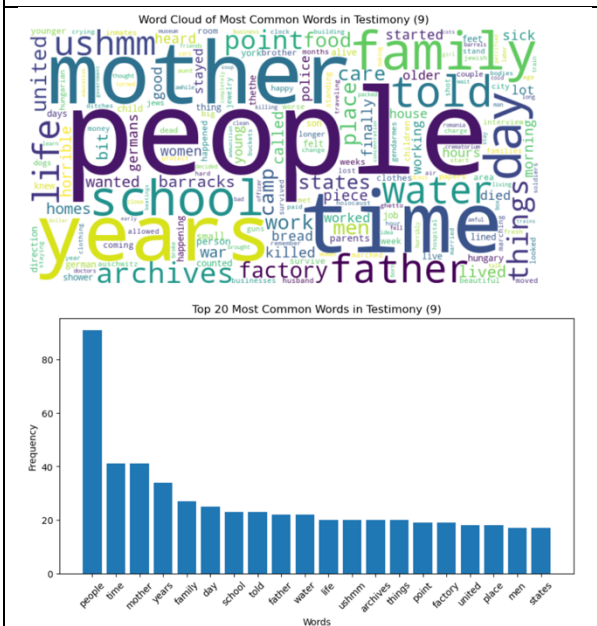
### 7. Leo Schneiderman



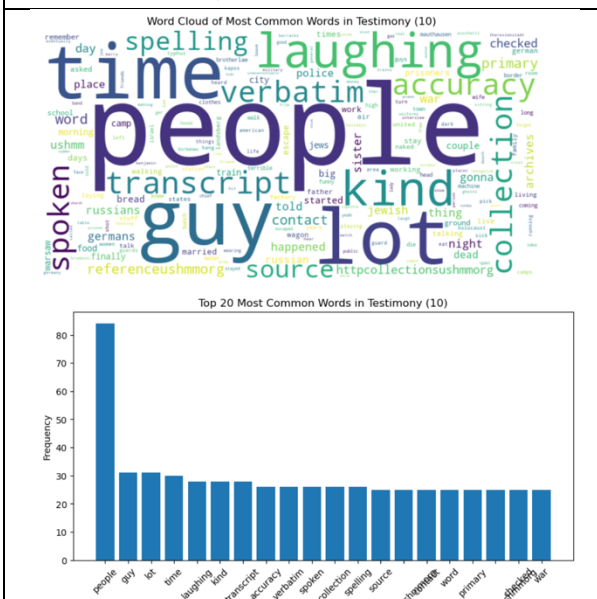
### 8. Bernard Pasternak



## 9. Madeline Deutsch



## 10. Boleslaw Brodecki



분석 기법을 적용하여 감정 변화와 경험을 보다 다각적으로 분석할 수 있을 것으로 예상된다. 감정 변화의 정량적 분석을 더욱 정교화하고, 다양한 변수와의 상관관계를 연구하여 더 깊이 있는 결과를 도출할 계획이다.

### 3. 결 론

연구를 통해 아우슈비츠 생존자들의 증언에서 감정 변화가 특정 패턴을 따르고 있다는 점을 발견했다. 특히, 폴 에크만의 6가지 감정 중 슬픔과 두려움이 아우슈비츠 시기에 특히 두드러졌으며, 해방 이후에도 그 여파가 지속되었다는 점이 확인되었다.

증언에서 '가족과의 이별', '강제 수용소 생활', '기아', '해방' 등이 반복적으로 언급되며, 이들이 생존자들의 감정에 깊은 영향을 미쳤음을 확인할 수 있었다.

다양한 방법으로 홀로코스트 생존자들의 감정과 기억을 탐구하고자 하였고 추후 여러 기술적인 한계들을 보완하여 깊이 있게 탐구해보고 싶다. 일제강점기 수용소 생존자 증언과 비교분석을 해보는 연구로도 이어 나갈 수 있을 듯하다.

## 참고 문헌

- 1) 미국 홀로코스트 기념관,  
<https://www.ushmm.org/ko>
- 2) Paul Ekman, “An Argument for Basic Emotion”,  
Cognition and Emotion, 1992
- 3) 최호근, 〈기념의 미래〉, 고려대학교 출판 문화원,  
2019
- 4) 최호근, 〈서양현대사의 블랙박스: 나치대학살〉, 푸른  
역사, 2006
- 5) 로버트 팩스틴, 〈파시즘: 열정과 광기의 정치 혁명〉,  
교양인, 2005

## 6. 연구 한계 및 제언

전체적으로 연구에 사용된 생존자 증언의 수가 제한적이었으며, 이로 인해 연구 결과의 일반화 가능성이 제한되었다. 또, 일부 증언은 감정적 표현이 제한적이거나 생략된 부분이 있어, 분석에 어려움이 있었다. 감정 변화의 평가에 있어 주관적 해석이 개입될 수 있음을 인지하고 있으며, 이는 결과에 영향을 미칠 수 있다는 사실이 한계라고 생각한다.

더 큰 데이터셋을 활용하여 향후 연구에서는 분석의 신뢰성을 높일 필요가 있고 다양한 텍스트