

Cognitive Geographies of Catastrophe Narratives: Georeferenced Interview Transcriptions as Language Resource for Models of Forced Displacement

Annie K. Lamar^{1,2*}, Rick Castle^{1,2*}, Carissa Chappell^{1,2*}, Emmanouela Schoinoplokaki^{1,2*}, Allene M. Seet^{1,2*}, Amit Shilo^{1*}, Chloe Nahas³

¹Department of Classics, University of California, Santa Barbara

²Low-Resource Language (LOREL) Lab, University of California, Santa Barbara

³Department of Psychology, University of Montreal

Correspondence: aklamar@ucsb.edu

Abstract

We present a machine-understandable geotagged dataset of translated interviews from the Nakba Archive alongside a complete georeferenced dataset of named locations mentioned in the interviews. In a preliminary analysis of this dataset, we find that the cognitive relationship of interviewees to place and spatiality is significantly correlated with gender. Our data also shows that interviewees with birthplaces depopulated in the 1948 Nakba incorporate references to named places in their interviews in substantially different ways than other interviewees. This suggests that the status of the interviewee's birthplace may impact the way they narrate their experiences. Our work serves as a foundation for continued and expanded statistical and cognitive models of Palestinian forced displacement.

1 Introduction

The Nakba Archive ([The Nakba Archive, 2002](#)), a grassroots oral history collective, conducts and archives interviews with Palestinians forcibly displaced during the 1948 Palestinian Nakba. Only thirty of the over five hundred interviews in the Archive have been transcribed and translated into English. In this paper, we present a new language resource to enable semantic and cognitive geospatial analysis of the translated portion of the Nakba Archive.

We provide a georeferenced, machine-understandable dataset of the translated interviews from the Nakba Archive. Due to the limits of current named entity recognition (NER) models for multilingual datasets, we performed the georeferencing manually. We also offer a preliminary analysis of the interviews through the lens of cognitive geography and computational corpus linguistics.

We observe two significant outcomes in our preliminary analysis. First, we find that interviewees'

linguistic usage of named places and types of places is significantly correlated to gender (see 4.1). Second, we find that interviews with people whose birthplaces were depopulated in the 1948 Nakba contain references to fewer named places, places in a smaller geographic range, and places, on average, farther from the interviewee's birthplace than other interviews. This suggests that the status of the interviewee's hometown may impact the way they narrate their experiences (see 4.3-4).

2 Background

2.1 The Palestinian Nakba of 1948

The Nakba marks the mass exodus of more than 700,000 Palestinians in 1948, the destruction of more than five-hundred Palestinian villages, and the erasure of hundreds of years of history and local culture after the creation of the state of Israel ([Khalidi, 1992](#); [Pappé, 2006](#)). The Nakba had a long-standing impact on the population, creating a massive ongoing refugee crisis ([Masalha, 2003](#)). These refugees were refused the right to return to Palestine and to their homes despite United Nations Resolution 194 ([United Nations General Assembly, 1948](#)).

The cultural memory of the Nakba is a central element in Palestinian identity and motivates the pursuit of justice and return ([Sa'di, 2007](#); [Morris, 2004](#)). This event has left Palestinians with intergenerational trauma that continues to have an impact on mental health and collective identity ([Kassem, 2011](#); [Pappé, 2006](#)). The Nakba is both a historical event and a symbol for the ongoing process of colonization and displacement of the Palestinian people ([Finkelstein, 2003](#)). Official archives and documents from government sources reveal military plans, such as the Dalet plan, to gain control of Palestinian territories and to displace Palestinians ([Pappé, 2006](#)). Every year the Nakba is commemorated on May 15 as a reaffir-

mation of Palestinian rights and remembrance of the injustices endured by the population (Sayigh, 1994).

2.2 Cognitive and Semantic Geography

Cognitive geography is a set of methods and approaches to understanding human thought and behavior as it pertains to space, place, and environment (Mark et al., 1999; Montello, 2018). The field uses geographic features and references to infer how people perceive, conceptualize, and respond to their environments.

Cognitive geography approaches include methods from geospatial semantics, a subfield that focuses on understanding the meaning of geographic entities (Hu, 2018). Such approaches are not necessarily limited to coordinate-based geography, especially when considering regions with vague or changing geographic boundaries (Montello et al., 2014).

In our preliminary analysis, we consider both coordinate and non-coordinate characteristics of named places. We annotate and analyze, for example, the categories of named places an interviewee mentions and the density of references to named places within the transcribed text. Our approach is limited by the amount, quality, and language of the available data (see Limitations).

We do not draw conclusions about the cognition or emotions of the interviewees, but rather we offer a framework for future research into the relationship of displaced peoples and their geographic environments.

3 Data and Language Resource Creation

The Nakba Archive (The Nakba Archive, 2002) is a digital collection of interviews with Palestinians forcibly displaced during the 1948 Palestinian Nakba. Of the over five-hundred oral histories currently preserved in the Nakba Archive, only thirty have been transcribed and translated from Arabic to English. Our structured, geotagged, and georeferenced dataset created from the interviews presented by the Nakba Archive is a significant new language resource for computational historians, linguists, and activists. The dataset has been published in the Harvard Dataverse (Lamar et al., 2024).

3.1 Data Structuring and Annotation

For each of the thirty translated interviews, we construct a dataset that partitions interviewer prompts

and their respective interviewee responses. We record separately all the metadata about the interview that is provided by the Nakba Archive, including the interviewee name(s), place and date of birth, place and date of interview, interviewer name, and permalink.^{1,2} We also add metadata that includes the language of the original oral interview³ and the presumed gender⁴ of both the interviewee and interviewer. Interview text and metadata are correlated by unique identifiers.

3.2 Geotagging and Georeferencing

For each of the thirty translated interviews, we manually tag every named geographic location mentioned in the text of the interview. Note that locations mentioned by interviewees and in footnotes are tagged, but not considered in our analysis. We consider a "location" to be any instance of a named place for which we reasonably expected to find geographic coordinates. For example, in the phrase "my home" we do not tag a location, but in the phrase "my home in Haifa," we tag "Haifa" as a location.

We then create a set of all tagged named locations from the interviewee responses. There are a total of 331 unique locations in the dataset. For each named location, we manually locate the latitude and longitude. Many places in the dataset have the same names and their physical location must be determined through context; in addition, some places were only able to be located relative to other places or landmarks and by the use of historical resources and scanned maps. If the location exists as an entity on Wikipedia (Wikimedia Foundation, 2012), we use the coordinates presented there. If not, we use other mapping tools and context clues to assign a likely latitude and longitude. If such resources are used, they are cited in the *notes* variable of the dataset containing georeferenced places.

¹For the sake of data stability, we preserve the PDF files of the translated interviews as they were available through the Nakba Archive in September 2024.

²Most of the interview transcripts also include additional information and definitions relevant to the interview content in numbered footnotes. We annotated and recorded each of the footnotes in a separate data table, which is available alongside the other components of our dataset.

³Although all the interviews currently in the Nakba Archive are in Arabic, we preserve this variable to make it easier for future researchers to join this dataset with others.

⁴One interview contains responses by two interviewees, Jabr Muhammad Yunis and Khalidiya Muhammad Yunis. Because Jabr Muhammad Yunis does a large portion of the speaking in this interview, we consider the interviewee to be 'male' for the purpose of metadata.

Category	Description	Example	Count	Frequency
Camp	Refugee camp.	Shatila Camp	11	3.32%
City	Large inhabited settlement.	Nablus	65	19.64%
Continent	Name of a continent or portion that mentions continent.	Eastern Europe	8	2.42%
Country	Name of a country.	Syria	26	7.85%
Feature	Specific, named locations such as landmarks, shops, bridges, etc.	Qasmiya Bridge	50	15.11%
Moshav	Jewish agricultural settlement.	Meiron	5	1.51%
Neighborhood	A named neighborhood or district within a city.	Burj al-Barajneh	18	5.44%
Region	Large portion of a country or countries.	Upper Galilee	12	3.63%
School	School of any level, including universities.	Birzeit University	23	6.95%
Town	Inhabited settlement larger than a village, smaller than a city.	al-Nasirah	29	8.76%
Village (-1948)	Village depopulated in 1948.	al-Kabri	55	16.62%
Village (current)	Currently inhabited village.	Yirka	29	8.76%

Table 1: Location categories used in our dataset with descriptions and distribution of categories throughout the dataset of georeferenced places. The **Count** column includes the number of unique places in each category. Note that **Frequency** is the number of places within a particular category out of the list of unique places. For frequencies within the interview text, see Results and Analysis below.

3.3 Location Categorization

We also label each of the 331 locations with a location category. We include twelve possible categories. Whenever possible, we use the label provided by Wikipedia. Otherwise, we use context clues to infer the category to which a location belongs. More information about the location categories is available in Table 1.⁵

4 Results and Analysis

Of all the locations mentioned in these interviews, Palestine is by far the most frequent. Over 7% of geographic references are to Palestine. The next most common reference is to the large city of 'Akka (Acre), which represents 4.97% of references.

⁵Note that we code "Palestine" as a country for the purposes of this language resource.

4.1 Named Geographic References by Interview

There are 1,168 references to 331 unique named geographic locations in the thirty translated interviews available in the Nakba Archive. The distribution of named geographic references across the interviews is shown in Figure 1.

There is a significant correlation between the gender of the interviewee and the frequency of references to named locations ($r(28) = -.49, p < .01$).⁶ Men make 50% more references to named locations than women. The frequency of geographic references is not significantly correlated to the gender of the interviewer ($r(28) = -.14, p > .01$). Likewise, the interviewee's gender does not have a statistically significant correlation to the total length of the interview ($r(28) = -.13, p > .01$;

⁶In this study, presumed gender was coded as binary; we use 0 to represent 'male' and 1 to represent 'female'.

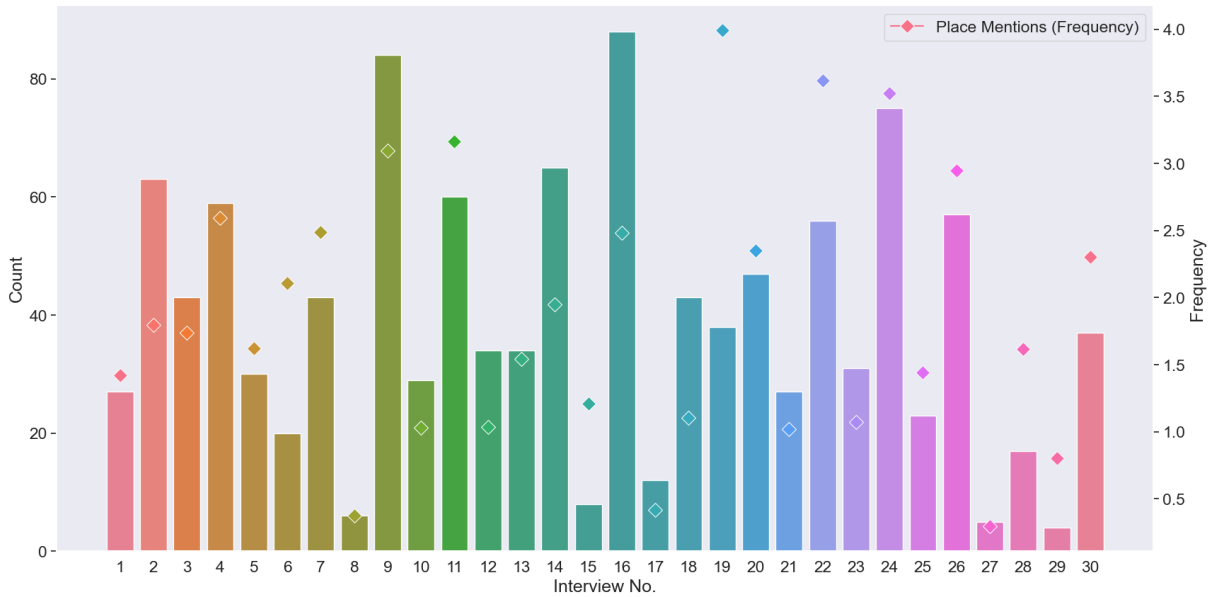


Figure 1: Distribution of named geographic references across all thirty interviews in our dataset. The distribution is presented both as **Count** (bars, left y-axis) and **Frequency** (diamonds, right y-axis). Interviews are represented as numbers on the x-axis for the sake of space; the map of interview numbers to interviewee names is included in Appendix A.

Place Type	Female	Male
Camp	-74.16	55.95
City	-32.87	24.8
Continent	-66.78	50.38
Country	-27.33	20.61
Feature	-16.13	12.17
Moshav	45.35	-34.21
Neighborhood	-64.22	48.45
Region	-41.86	31.58
School	-75.94	57.29
Town	-3.1	2.34
Village (-1948)	8.35	-6.3
Village (current)	47.06	-35.5

Table 2: Percent difference between the expected number of references to **Place Type** and the true number of references, grouped by interviewee gender. Men comprise 57% of the dataset, so the expected number of references to, for example, cities is $(0.57 * \text{total number of references to cities}) = 485$. Men are thus expected to make 277 references to cities, but in fact make 345: a 24.80% difference.

length computed as word count).

It is therefore unsurprising that when we consider named references grouped by location category, men make more than the statistically expected number of references (i.e. 57%) in almost all cat-

egories (Table 2). Male interviewees make, for example, 345 references to cities, a number about 25% higher than we would expect for a dataset comprised of 57% men (expected: 277).

There are three types of locations that women reference with a frequency greater than is expected. The first of these categories, moshavs, has an extremely limited representation in the dataset. The second category, villages depopulated in 1948, shows only a slight over-representation among female interviewees. The third category, however, is well-represented. Women make nearly 50% more references to currently inhabited named villages than is expected based on the gender demographics of the dataset. In fact, of the total 244 references to named villages, women make 125 of them.

4.2 Named Geographic References by Interview Time

We use the concept of interview time in our analysis. Interview time is based on the concept of narrative time, the time it takes the narrator to tell a story in text (Genette, 1980). We define interview time as the percent of total interview progress based on word count.⁷ This allows us to examine at what stage of the interview participants mention

⁷Note that many transcripts appear to not include the entirety of the interview. We base interview time on the transcripts provided by the Nakba Archive without consideration of video timestamps.

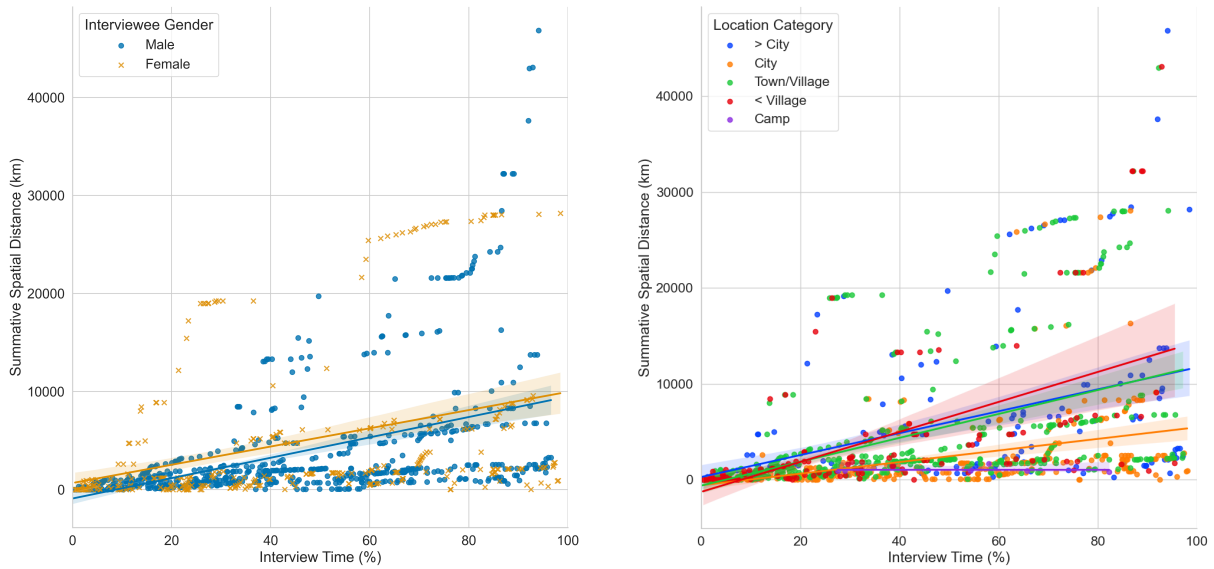


Figure 2: Scatter plots showing summative distance vs. interview time, grouped by gender of interviewee (left) and location category (right). Each dot represents the total summative distance at a specific time in the interview.

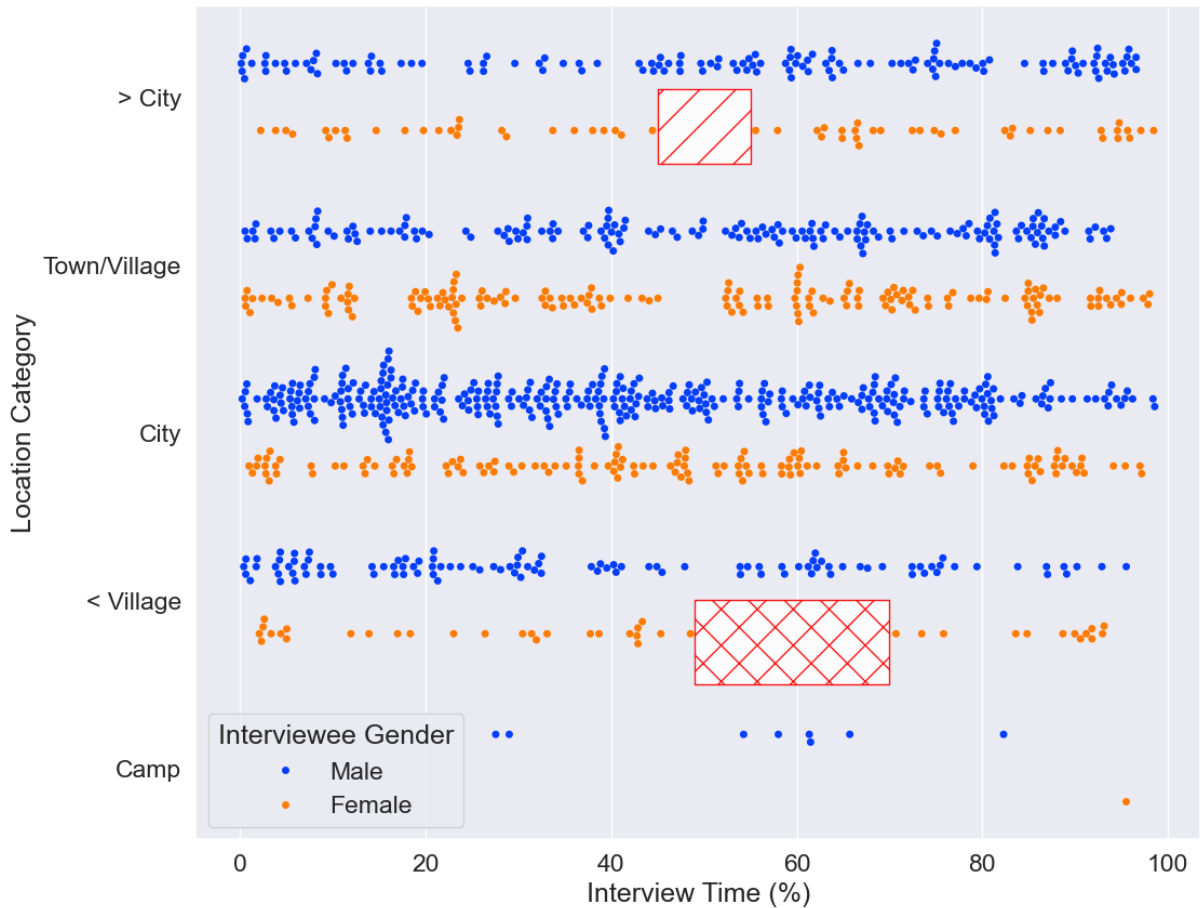


Figure 3: Swarmplot showing references to places in five broad location categories vs. interview time, grouped by gender. *> City* includes continents, countries, and regions. *Town/Village* includes towns, currently inhabited villages, and depopulated villages. *< Village* includes features, schools, neighborhoods and moshavs. The red-hatched boxes highlight two spans of interview time in which women made zero references to cities (upper box) and villages, features, and schools (lower box).

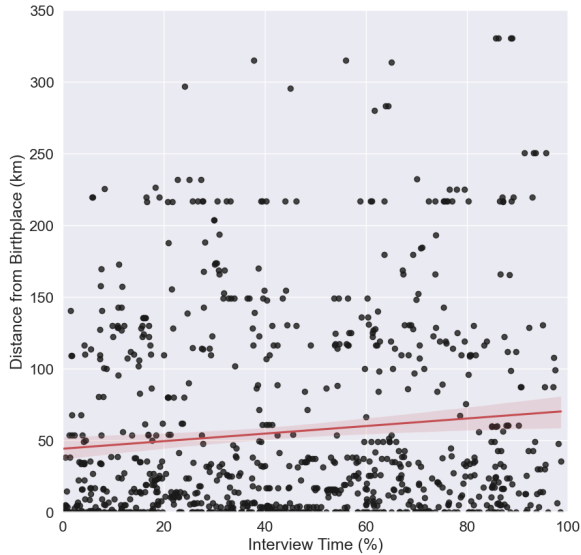


Figure 4: Scatter plot showing **Distance from Birthplace** vs. **Interview Time**. Each dot represents a single reference to a named location. The red line is a linear trend line. For readability, this scatter plot only shows references to locations with a distance from interviewee’s birthplace within 0.1 standard deviations. The complete dataset is published in the Harvard Dataverse (Lamar et al., 2024).

certain named places or categories of places while controlling for the length of the interview.

Figure 3 shows the distribution of named references to places within five broad categories. For male interviewees, we observe a slightly higher frequency of overall named geographic references earlier in the interview.

For female interviewees, we observe a fairly consistent frequency of overall named geographic references throughout the interview, with some exceptions. Near the middle of the interview, at about 45%-55% of interview time, women make no references to named geographic entities larger than cities. From 55% to 70% of interview time, women make zero references to named features, schools, or neighborhoods. During this same span of interview time, men make most of their references to refugee camps.⁸

4.3 Spatial Distance and Geographic Range

We use two metrics to evaluate the georeferenced dataset in terms of coordinate-based spatial distance: distance from birthplace and summative distance. Whenever the distance between two pairs of coordinates is computed, we use haversine distance

⁸There is only one reference to a refugee camp made by a female interviewee.

(de Mendoza y Rios, 1795). Haversine distance uses the haversine trigonometric function (Equation 1).

$$\text{haversin}(\theta) = \sin^2\left(\frac{\theta}{2}\right) \quad (1)$$

Let d be the spherical distance between two points and let r be the radius of Earth. Given two pairs of coordinates for those points (ϕ_1, ϕ_2) and (λ_1, λ_2) , we compute haversine distance as in Equation 2. We implement the Haversine Python package (Rouberol and Deniau, 2024).

$$\begin{aligned} \text{haversin}\left(\frac{d}{r}\right) = & \text{haversin}(\phi_2 - \phi_1) \\ & + \cos(\phi_1)\cos(\phi_2) \\ & \times \text{haversin}(\lambda_2 - \lambda_1) \end{aligned} \quad (2)$$

4.3.1 Distance from Birthplace

We first investigate the distance between named locations and an interviewee’s birthplace over interview time. As shown in Figure 4, when we consider the named locations with a distance from interviewee’s birthplace within 0.1 standard deviations, we observe only a slight upward trend across interview time. Further investigation with a larger dataset is required to determine if this trend is an artifact of the limited size of our dataset (see Conclusion).

4.3.2 Summative Distance

We define summative distance as the total distance between all points in the interview, in the order they are mentioned. For example, if an interviewee consecutively mentions Place A, Place B, Place A, and Place C, we sum $\text{haversin}(A - B)$, $\text{haversin}(B - A)$, $\text{haversin}(A - C)$. Therefore, summative distance will *always* increase over interview time; our analysis depends on the relative rate of increase.

We find very minimal difference in the rate of summative distance growth between male and female interviewees (Figure 2, left). It is notable that at approximately 80% through the interview, some interviewees name locations that are a relatively greater distance away than previously named locations. This result is primarily influenced by the interviews of Nicola Ziadeh and Renee Kutih, represented by interviews nine and twenty-three, respectively. Nicola Ziadeh mentions locations in England and Russia in the context of education, dramatically increasing the summative distance. Renee Kutih mentions a number of cities relatively

distant from each other in Israel and Palestine, and recalls a maid from Jizin in Saudi Arabia.

We also find that further into the interview, smaller, more specific named locations have a greater impact on summative distance (Figure 2, right). *City* has a slow rate of increase while *Village* (including features, schools, and neighborhoods) has the highest rate of increase. This might suggest that even though interviewees continue to mention a variety of distant places, they name increasingly specific locations (e.g. a specific neighborhood, rather than a city) nearer to the end of the interview. One possible explanation for this is an established vocabulary between the interviewer and interviewee; since the interviewee has provided ample context for their story by the end of the interview, the specific names of places make sense to the interviewer.

4.4 Named Geographic References by Status of Birthplace

Of the thirty interviewees,⁹ exactly half provide a birthplace that was one of the villages depopulated in 1948. We find a significant correlation between status of birthplace and distance from birthplace across all named location references ($r(1163) = .18, p < .01$) (Figure 5, left). The mean distance from birthplace mentioned by the group of interviewees with birthplaces depopulated in 1948 is 1,485.24 km vs. a mean distance from birthplace of 1,057.73 km for the other interviewees.

We also observe, however, a much slower rate of increase of summative distance for interviewees with birthplaces depopulated in 1948. This is largely due to two factors. First, interviewees with birthplaces depopulated in 1948 mention named places far less than the other interviewees. Although representing half the dataset, the former set of interviewees only make 37.8% of references to named places.

Second, interviewees with birthplaces depopulated in 1948 reference a much smaller geographic range of named places overall. We compute geographic range as the maximum distance between any two places named in a single interview. The mean geographic range for interviewees with birthplaces depopulated in 1948 is 28% smaller than that for the other interviewees.

⁹Jabr Muhammad Yunis and Khalidiya Muhammad Yunis, who are interviewed together, list the same birthplace.

Interviewees from depopulated birthplaces verbalize geographies in their narratives that are simultaneously farther from their birthplaces and closer to their birth region.

5 Conclusion

We present a geotagged machine-understandable dataset of the translated interviews from the Nakba Archive alongside a complete georeferenced dataset of named locations mentioned in the interviews. Our structured, geotagged, and georeferenced dataset created from the interviews presented by the Nakba Archive is a significant new language resource for computational historians, linguists, and activists.

We also offer a preliminary analysis as an exemplum of how this data can be used in the future. We find 1,168 references to 331 unique named geographic locations in the thirty translated interviews available in the Nakba Archive. We find a significant correlation between the gender of the interviewee and the frequency of references to geographic locations. By considering spatial distance, we find that interviewees mention places slightly farther from their birthplace the farther they are into the interview and that smaller, more specific locations have a greater impact on summative distance near the end of the interviews.

Finally, we also observe a much slower rate of increase of summative distance for interviewees with birthplaces depopulated in 1948. Investigation into the underlying data reveals that such interviewees mention fewer named places overall and present narratives with smaller geographic ranges.

We intend for these results to serve as a model for continued work and to allow for work towards a cognitive model of geospatial displacement. Information about how displaced peoples understand place and their role in it is invaluable for those working to promote peace and create opportunities for healing connections between homelands and forcibly displaced peoples.

6 Limitations

The two most significant limitations of our work are (1) our inability to work with data in the original language of Arabic and (2) our annotation of only named locations rather than all locations ("my home," "the river") in the interviews. Our analysis is therefore limited to those interviews made available by the Nakba Archive in English and to

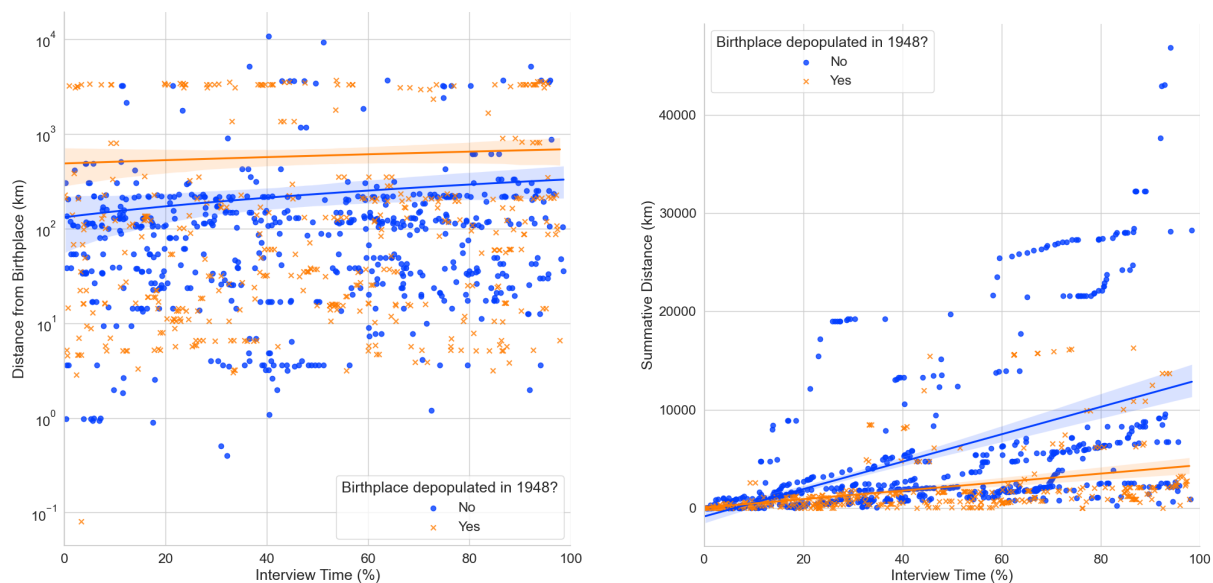


Figure 5: Scatter plots showing distance from birthplace vs. interview time (left) and summative distance vs. interview time (right), each grouped by categorization of birthplace (depopulated in 1948 or not). Note the logarithmic y-axis for the left plot.

consideration of named geographic locations. It is our hope that our method and open-source code can be used by researchers with linguistic and cultural expertise in Arabic.

Other limitations also include the availability of data, especially for small villages and towns that do not have coordinates available through WikiData or other major mapping services like Google Maps. In these cases, we relied on context clues from the interviews (e.g. "five kilometers south of Yaffa") to assign coordinates. Still, we were unable to locate coordinates for fourteen locations in the dataset, which represent 4% of the dataset of named places.

Acknowledgments

We express gratitude to the interviewees whose stories comprise the language resource presented above. We also thank the creators of the Nakba Archive and the interview translators, without whom this project would not have been possible.

References

- J. de Mendoza y Rios. 1795. *Memoria sobre algunos Métodos nuevos de calcular la Longitud por las distancias lunares: y aplicacion de su teórica á la solucion de otras problemas de navegacion*. Imprenta Real.
- Norman Finkelstein. 2003. *Image and Reality of the Israel-Palestine Conflict*. Verso Books.

Gérard Genette. 1980. *Narrative Discourse: An Essay in Method*. Cornell University Press.

Yingjie Hu. 2018. [Geospatial semantics](#). *Comprehensive Geographic Information Systems*, page 80–94.

Fatma Kassem. 2011. *Palestinian Women: Narrative Histories and Gendered Memory*. Zed Books.

Walid Khalidi. 1992. *All that Remains: The Palestinian Villages Occupied and Depopulated by Israel in 1948*. Institute for Palestine Studies.

Annie K. Lamar, Emmanouela Schoinoplokaki, Rick Castle, Carissa Chappell, and Allene M. Seet. 2024. [Georeferenced Interview Transcriptions from the Nakba Archive](#).

David M. Mark, Christian Freksa, Stephen C. Hirtle, Robert Lloyd, and Barbara Tversky. 1999. [Cognitive models of geographical space](#). *International Journal of Geographical Information Science*, 13(8):747–774.

Nur-eldeen Masalha. 2003. *The Politics of Denial: Israel and the Palestinian Refugee Problem*. Pluto Press.

Daniel Montello, Alinda Friedman, and Daniel Phillips. 2014. [Vague cognitive regions in geography and geographic information science](#). *International Journal of Geographical Information Science*, 28:1802–1820.

Daniel R. Montello. 2018. *Handbook of Behavioral and Cognitive Geography*. Edward Elgar Publishing.

Benny Morris. 2004. *The Birth of the Palestinian Refugee Problem Revisited*. Cambridge University Press.

- National Information Standards Organization (NISO). 2022. *CRedit, Contributor Roles Taxonomy*. National Information Standards Organization.
- Ilan Pappé. 2006. *The Ethnic Cleansing of Palestine*. Oneworld Publications.
- Balthazar Rouberol and Julien Deniau. 2024. Haversine. <https://pypi.org/project/haversine>. Version 2.8.1.
- Lila Sa'di, Ahmad H.; Abu-Lughod, editor. 2007. *Nakba: Palestine, 1948, and the Claims of Memory*. Columbia University Press.
- Rosemary Sayigh. 1994. *Too Many Enemies: The Palestinian Experience in Lebanon*. Zed Books.
- The Nakba Archive. 2002. [Nakba Archive](#).
- United Nations General Assembly. 1948. *Resolution 194 (III), Article 11*.
- Wikimedia Foundation. 2012. [Wikidata](#).

A Appendix: Interviewee Names and Metadata

Interview ID	Interviewee	Gender	Birthyear	Birthplace
1	Ibrahim Mahmoud Blaybil	Male	1920	Taytabah, Palestine
2	Fifi Khouri	Female	1922	Yafa, Palestine
3	Hasna Mana	Female	1931	al-Manshiyya, Palestine
4	Hamda Jumaa	Female	NA	Arab al-Zubayd, Palestine
5	Ahmad Agha	Male	1930	Tarshiha, Palestine
6	Rifaat al-Nimir	Male	1918	Nablus, Palestine
7	Husayn Mustafa Taha	Male	1921	Miar, Palestine
8	Jabr Muhammad Yunis	Male	1924	Safsaf, Palestine
8	Khalidiya Muhammad Yunis	Female	1922	Safsaf, Palestine
9	Nicola Ziadeh	Male	1907	Damascus, Syria
10	Muhammad Jamil Arabi	Male	1923	Haifa, Palestine
11	Mahmud Abu al-Hayja	Male	1928	Haifa, Palestine
12	Umar Shihada	Male	1922	Qabbaa, Palestine
13	Abd al-Rahman Saad al-Din	Male	1915	al-Zib, Palestine
14	Taliba Muhammad Fuda	Female	1929	Suhmata, Palestine
15	Amina Abd al-Karim al-Wakid	Female	NA	Aylut, Palestine
16	Ismail Shammout	Male	1930	Lydda, Palestine
17	Fatima Shaaban	Female	1928	al-Zib, Palestine
18	Amina Hasan banat	Female	1931	Shaykh Dannun, Palestine
19	Salih al-Nasir	Male	1912	Saffuriyya, Palestine
20	Husayn Lubani	Male	1937	al-Damun, Palestine
21	Kamila al-Abd Tahir	Female	1933	Saliha, Palestine
22	Kamil Ahmad Balawi	Male	1928	Shafa Amr, Palestine
23	Renee Kutih	Female	1925	Ramla, Palestine
24	Muhammad Abu Raqaba	Male	1929	Akka, Palestine
25	Subhiya Salama	Female	NA	al-Zahiriyya, Palestine
26	Anis Sayigh	Male	1931	Tabariyya, Palestine
27	Maryam Uthman	Female	1937	al-Husayniyya, Palestine
28	Fatima Abdallah	Female	NA	Sasaa, Palestine
29	Maryam Mahmud Sabha	Female	1920	al-Zib, Palestine
30	Hasan al-Husayni	Male	1925	al-Quds, Palestine

Table 3: Interviewee metadata, including name, presumed gender, birthyear and birthplace.

B Appendix: Interview Metadata

Interview ID	Location	Date	Interviewer
1	Ayn al-Hilweh camp, Sayda	2004-02-07	Mahmoud Zeidan
2	Hamra, Beirut	2004-07-06	Bushra Mughrabi
3	Ayn al-Hilweh camp, Sayda	2003-01-01	Mahmoud Zeidan
4	Qasmiya gathering north of Sur	2003-01-01	Bushra Mughrabi
5	Burj al-Barajneh, Beirut	2004-03-14	Bushra Mughrabi
6	Beirut	2003-12-11	Mahmoud Zeidan
7	Miye wa Miye, Sayda	2003-04-25	Mahmoud Zeidan
8	Ayn al-Hilweh camp, Sayda	2003-01-01	Mahmoud Zeidan
9	Beirut	2004-01-29	Mahmoud Zeidan
10	Mazbud, Iqlim al-Kharub	2003-10-12	Mahmoud Zeidan
11	Burj al-Barajneh, Beirut	2003-01-01	Mahmoud Zeidan
12	Taalabaya, al-Biqaa	2004-05-29	Mahmoud Zeidan
13	Beirut	2003-01-02	Mahmoud Zeidan
14	Not provided.	2003-10-05	Amira Ahmad Alwan
15	al-Bus camp, Sur	2003-10-15	Bushra Mughrabi
16	Malaab al-Baladi, Beirut	2003-10-11	Mahmoud Zeidan
17	al-Bus camp, Sur	2003-05-15	Jihad al-'Ali
18	Burj al-Barajneh, Beirut	2003-01-01	Bushra Mughrabi
18	Burj al-Barajneh, Beirut	2003-01-01	Mahmoud Zeidan
19	Ayn al-Hilweh camp, Sayda	2003-01-01	Mahmoud Zeidan
20	Trablus, Lebanon	2004-02-08	Mahmoud Zeidan
21	al-Murayja, Beirut	2004-07-09	Bushra Mughrabi
22	Badawi camp, North Lebanon	2003-01-01	Amira Ahmad Alwan
23	Verdun, Beirut	2011-08-17	Mahmoud Zeidan
24	Beirut	2003-11-12	Mahmoud Zeidan
25	Ayn al-Hilweh camp, Sayda	2006-04-29	Bushra Mughrabi
26	Beirut	2003-01-01	Mahmoud Zeidan
27	Burj al-Shamali, Sur	2004-05-09	Bushra Mughrabi
28	Mar Elias camp	2004-03-25	Muhammad al-Masri
29	al-Maashuq, Sur	2003-10-23	Mahmoud Zeidan
30	Verdun, Beirut	2003-12-19	Mahmoud Zeidan

Table 4: Interview metadata, including date and location of interview and name of interviewer. If only a year was provided, we supplied January 1 as the month and day.

C Contributor Roles

We use the CRediT framework to provide detailed information about the contributions of each of this paper’s authors (National Information Standards Organization (NISO), 2022).

Author	A	B	C	D	E	F	G	H	I	J	K	L	M	N
Lamar	◆	●	◆	◆	◆	◆	◆	◆	◆	◆	◆	◆	◆	◆
Castle	●	◆	-	-	●	●	●	-	-	-	-	-	●	●
Chappell	●	●	-	-	●	●	-	-	-	-	-	●	●	●
Schoinoplokaki	●	●	-	-	●	●	-	-	-	-	-	-	◆	●
Seet	●	●	-	-	●	●	-	-	-	-	-	-	●	●
Shilo	◆	-	-	-	-	●	-	-	-	-	◆	-	●	◆
Nahas	-	-	-	-	-	-	-	-	-	-	-	-	●	-

Table 5: The ◆ symbol indicates (co-)lead role in category; ● indicates contribution in category. **A: Conceptualization** (Ideas; formulation or evolution of overarching research goals and aims), **B: Data Curation** (Management activities to annotate (produce metadata), scrub data and maintain research data (including software code, where it is necessary for interpreting the data itself) for initial use and later re-use), **C: Formal Analysis** (Application of statistical, mathematical, computational, or other formal techniques to analyze or synthesize study data), **D: Funding Acquisition** (Acquisition of the financial support for the project leading to this publication), **E: Investigation** (Conducting a research and investigation process, specifically performing the experiments, or data/evidence collection), **F: Methodology** (Development or design of methodology; creation of models), **G: Project Administration** (Management and coordination responsibility for the research activity planning and execution), **H: Resources** (Provision of study materials, reagents, materials, patients, laboratory samples, animals, instrumentation, computing resources, or other analysis tools), **I: Software** (Programming, software development; designing computer programs; implementation of the computer code and supporting algorithms; testing of existing code components), **J: Supervision** (Oversight and leadership responsibility for the research activity planning and execution, including mentorship external to the core team), **K: Validation** (Verification, whether as a part of the activity or separate, of the overall replication/reproducibility of results/experiments and other research outputs), **L: Visualization** (Preparation, creation and/or presentation of the published work, specifically visualization/data presentation), **M: Writing - Original Draft** (Preparation, creation and/or presentation of the published work, specifically writing the initial draft), **N: Writing - Review & Editing** (Preparation, creation and/or presentation of the published work by those from the original research group, specifically critical review, commentary or revision – including pre- or post-publication stages)