Zero-shot Trajectory Mapping in Holocaust Testimonies

Eitan Wagner[†] **Renana Keydar**[‡] **Omri Abend**[†]
[†] Department of Computer Science [‡] Faculty of Law and Digital Humanities
Hebrew University of Jerusalem
{first_name}.{last_name}@mail.huji.ac.il

Abstract

This work presents the task of Zero-shot Trajectory Mapping, which focuses on the spatial dimension of narratives. The task consists of two parts: (1) creating a "map" with all the locations mentioned in a set of texts, and (2) extracting a trajectory from a single testimony and positioning it within the map. Following recent advances in context length capabilities of large language models, we propose a pipeline for this task in a completely unsupervised manner, without the requirement of any type of labels. We demonstrate the pipeline on a set of ≈ 75 testimonies and present the resulting map and samples of the trajectory. We conclude that current long-range models succeed in generating meaningful maps and trajectories. Other than the visualization and indexing, we propose future directions for adaptation of the task as a step for dividing testimony sets into clusters and for alignment between parallel parts of different testimonies.

Keywords: Mapping, Trajectory, Testimonies, Holocaust

1. Introduction

The location trajectory, i.e., the sequence of locations in which the story takes place, is an essential aspect of a story. The significance of location in a story is crucial, as placing a story in a specific setting is often seen as a defining characteristic that sets narrative texts apart from other types of writing (Piper and Bagga, 2022).

However, despite the abundance of Natural Language Processing (NLP) research on describing locations in texts, few efforts have been made to extract the progression or sequence of locations from a narrative story (Wagner et al., 2023). As a structured prediction task with a large class set, the ability to obtain data that is sufficient for generalization is very limited.

In this work, we present the task of zero-shot trajectory mapping and design a pipeline for it with long-context large language models. Zero-shot trajectory mapping involves both the extraction of the locations for each document (as a "trajectory") and the identification of the relationship between the locations (creating a "map"). We have no prior list of locations and the map is constructed based on the given texts only. Thus, the task is unsupervised in two ways – the set of locations must be inferred from a set of unannotated texts, and the trajectory of each text must be extracted without supervision.

Our research primarily centers on transcribed Holocaust survivor testimonies, which are provided in English. The significance of this dataset in the examination and remembrance of the Holocaust cannot be emphasized enough. As the last surviving witnesses inevitably pass away, there is an urgent necessity to find new approaches to engage with the extensive collection of Holocaust testimonies housed in records. Utilizing NLP technology for the analysis of these testimonies has recently been strongly recommended (Artstein et al., 2016; Wagner et al., 2022). By leveraging NLP, researchers can extract valuable insights from the vast array of testimonies (comprising tens of thousands) instead of limiting themselves to small-scale, predominantly manual studies. Additionally, we assert that Holocaust testimonies hold distinct value for NLP due to the combination of a multitude of accounts within a relatively confined domain of topics and locations. This quality sets them apart from typical narrative datasets (Sultana et al., 2022).

We describe and run a full pipeline for zero-shot trajectory mapping, using GPT4. [1] We show the resulting map on ≈ 75 testimonies and provide examples of the trajectories on this map. Based on the resulting maps, we describe future directions for alignment between testimonies.

Trajectory extraction is valuable for visualization and trajectory clustering (Bian et al., 2018). Characterizing a story by a sequence of locations is also beneficial as a backbone for alignment between different stories–an important task in its own right (see, e.g., Ernst et al., 2022). In general, successful location extraction indicates aspects of long-range narrative understanding, which is a highly active field in NLP (Yao et al., 2022; Bertsch et al., 2024).

2. Previous Work

Narrative Analysis. Narrative schema analysis aims to capture the core of event sequences, providing a condensed sequential timeline of a lengthy story. This overview helps in aligning relevant parts

[1] https://openai.com/research/gpt-4

The First Workshop on Holocaust Testimonies as Language Resources @LREC-COLING-2024, pages 63–70
20-25 May, 2024.

and identifying common topic paths, as demonstrated by Antoniak et al. (2019) in their study on birth stories using segment-wise topic modeling.

To extract an interpretable sequential progression it was assumed necessary to divide the long story into shorter segments (Wagner et al., 2023). However, recent advances in NLP introduced significant increases in context lengths of models (Wang et al., 2024), allowing the extraction of sequences as an end-to-end task.

Recent studies have highlighted the importance of event locations in narrative analysis. Piper et al. (2021) provided a definition of narratives that included a focus on event locations. Soni et al. (2023) introduced a task involving grounding characters in specific locations. Kumar and Singh (2019) extracted event locations from individual events, such as those found in tweets. Wagner et al. (2023) expanded on this concept by examining trajectories of locations throughout entire narratives, utilizing a predetermined set of coarse-grained categories.

Trajectory Modeling in Transportation. Some works seek to extract document-level trajectories in transportation. Mathew et al. (2012) applied Hidden Markov Models (HMM) to human location trajectories. Sassi et al. (2019) utilized convolutional neural networks on location embeddings as an alternative to HMMs. Lui et al. (2021) employed LSTM-based models for predicting pedestrian trajectories. These works focus on locations given as coordinates and not as natural-text descriptions, which allow for a more thematic level of representation and comparison (Wagner et al., 2023).

Narrative Cartography Many works investigated the mapping of narratives. Reuschel and Hurni (2011) presented methods for the visualization of location maps. Their methods show differences between the maps in fiction and non-fiction. Mai et al. (2022) develop toolboxes for enrichment of geographic data, based on knowledge graphs.

These works are primarily based on a location ontology, thus limiting the scope to domains with sufficient prior knowledge. In our work, we propose a completely unsupervised method, allowing its application without any prior knowledge.

3. Task Definition

Our setting is the following: given a set of texts $\mathbf{x}^1, \mathbf{x}^2, \ldots \mathbf{x}^k$, each divided into initial segments, $\mathbf{x}^i = x^i_1, x^i_2, \ldots, x^i_n$, we wish to predict: (1) one directed graph $G = (V, E)$, where the vertices V are all the locations (name+type) in the set of texts and the edges E are the relationships between them (e.g., New York is in the United States); (2) for each $\mathbf{x}^i$, a path on the graph G, describing the trajectory in this text. The path should have additional vertex labels for the indices within the text of this location (e.g., segments 17-21) and edge labels for the method of transportation, if applicable (e.g., “by foot”, “by plane” etc.). Roughly, we can say that the first part of the task corresponds to the creation of a “map” and the second part corresponds to the action of “mapping” within it.

It is instructive to compare this task to traditional Named Entity Recognition (NER) for location categories. NER is a prediction task at the phrase level that ignores the relationship between different locations or even between mentions of the same locations. Therefore, the first part of out task can be seen as a combination of NER and Entity Relation Extraction (focusing on the containment relation). The second part of our task is completely different as it requires a structured sequence as an output. Prediction is at the document level, with possible dependencies throughout the entire document. This property requires strong long-context capabilities which were not necessary for traditional NER.

3.1. Data

Our main data consists of 1000 Holocaust survivor testimonies, received from the Shoah Foundation (SF).[2] All interviews were conducted orally by an interviewer, recorded on video, and transcribed as time-stamped text. The lengths of the testimonies range from 2609 to 88105 words, with a mean length of 23536 words.

We note that the SF testimonies are divided into segments and contain highly detailed labels. Due to the extremely large set of labels we opted to use the text only and attempt zero-shot inference only. We arbitrarily chose a set of 74 testimonies and run them through our pipeline.

4. Zero-shot Trajectory Extraction

Recent advances in LLMs lead to a substantial increase in the context window that can be inputted into the models [3]. This makes it possible to input a whole testimony and perform location tracking as an end-to-end task.

For this we used GPT4-turbo-preview, which has a context length of $128K$ tokens [4]. The price for the experiment was ≈ 60 \$.

We remark that the end-to-end task differs from supervised location tracking (Wagner et al., 2023) in multiple aspects: (1) Zero-shot extraction is not limited by granularity – it extracts countries, cities,

[2] https://sfi.usc.edu/

[3] https://www.anthropic.com/news/claude-2-1

[4] https://platform.openai.com/docs/models/gpt-4-and-gpt-4-turbo

and also different types of locations (like "the forest") (2) Zero-shot extraction considers only locations that are mentioned in the text. This is also a limitation since different texts might be more specific with the names that are mentioned, leading to a longer trajectory.

4.1. Pipeline

Our pipeline is constructed of 4 steps: per-testimony location-graph extraction, per-testimony path extraction, combining all graphs, and visualizing each path in the combined graph.

Here we describe the details for each step.

Per-testimony location-graph extraction. For each testimony, we first extract a graph of the mentioned locations and their relationships.

We used gpt-4-turbo-preview with highly detailed instructions. The prompt was the following:

> I'll give you a Holocaust testimony.
> I want you to give me a JSON representing the graph of the mentioned locations (proper and common) and any known relations between them. Locations can be GPEs (like country or city) or significant facilities (like army camps, ghettos, concentration camps and death camps).
> Some important points:
> 1. Make sure the nodes contain locations only and not anything else (no nodes for events or people).
> 2. Give the nodes a type based on the type of location. The types should include: City, Country, Village, Ghetto, Army Camp, Concentration Camp, and Death Camp.
> 3. Keep the graph as full as possible, so, for example, if a place in a city in country is mentioned, there should be nodes for the place, the city, and the country. Separate a district from a city description into two nodes.
> 4. The graph should include relations between locations (i.e., A is in B). Make sure that the direction of an edge is that of inclusion if relevant (that is, if A is in B then the edge should be from A to B).
> 5. Make sure to avoid double entries.
> 6. Give me the graph as JSON dictionary, with a the "nodes" field indicating a list of nodes and "edges" indicating a list of edges. These nodes and edges should be in a format that can be create a python networkx graph. Make sure the nodes are given as a list of tuples, in which the first value is the name and the second is a dictionary with the type (as described above) The edges should be in a list of tuples, each containing two names (see example).
>
> Here is an example (from a different testimony):
> “‘json
>
> "nodes": <Here we provide an example list of locations>,
> "edges": <Here we provide an example relations between the locations>
>
> “‘
>
> This should all be based on the text.
>
> Testimony: <Here we add the testimony divided into numbered segments>

Per-testimony trajectory extraction. Following the answer about the locations, another request is made with the following prompt:

> Now, can you give a graph with the trajectory of the witness' movements? That is, give a list of location where he is. All location nodes should be nodes from the networkx graph you gave before. The nodes should have a field noting the sentence number in the text in which the witness was in that location.
> The edges should be between each adjacent node by order of the testimony. For each edge, add the method of transportation can be inferred from the text. Methods include: By foot, By car, By train, By plane. If the method is unknown give Unknown.
> Give me a graph in JSON format (like in the example).
>
> For example:
> “‘json
> "nodes": <Here we provide an example list of locations with their place in the testimony>,
> "edges": <Here we provide an example relations between adjacent locations, with the method of transportation>
> “‘

Combining the graphs into a map. To combine the obtained graphs into one global map, we first need to make sure that each location has only one label. Once we have one name per node, we can use the name as the identifier and create a

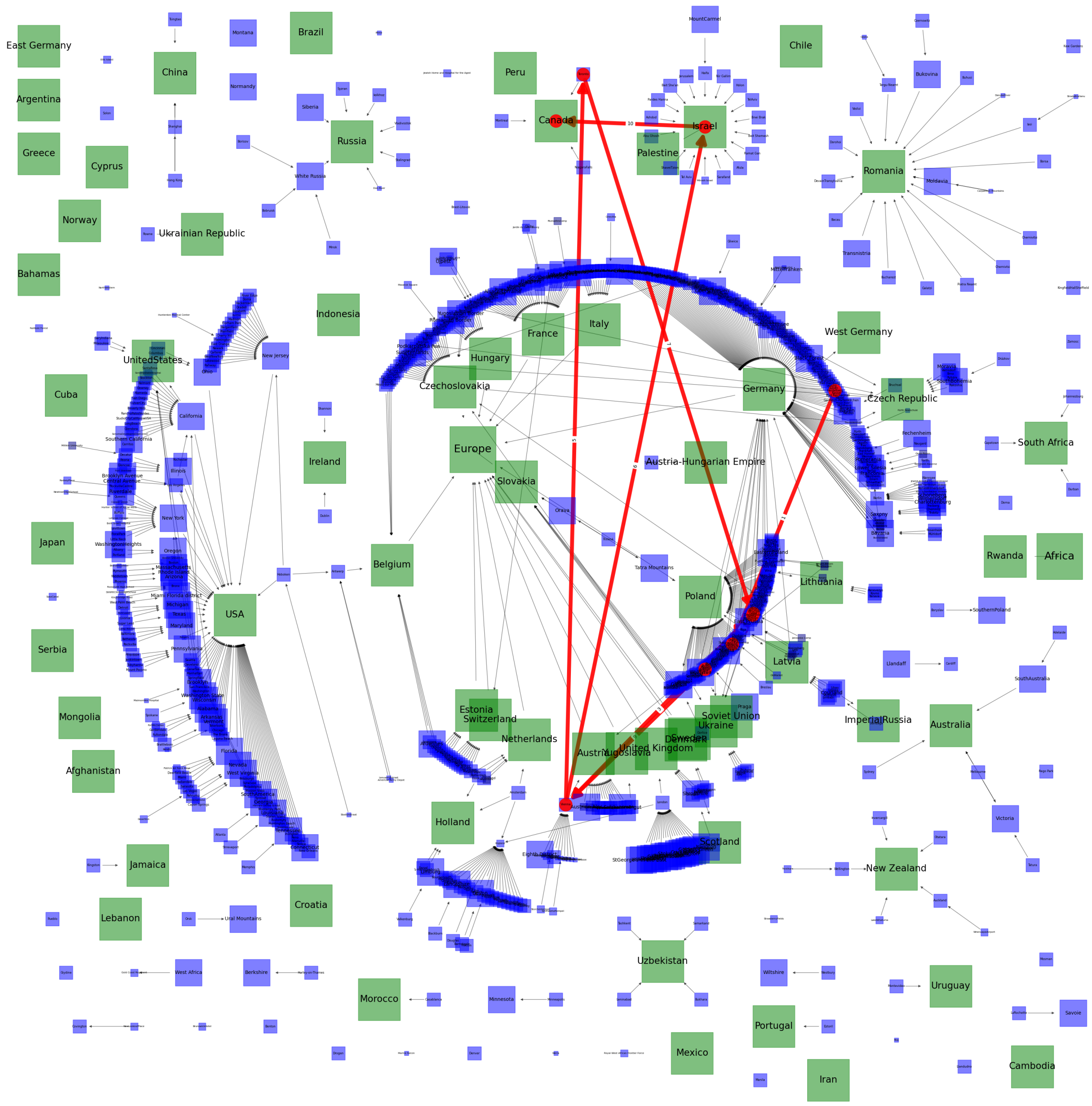

Figure 1: An example location map with a path extracted from a single testimony. Green nodes represent countries and blue nodes represent smaller locations. The path is in red.

graph with the new set of names and with all edges (removing doubles).

To create a list of double names, we again used GPT4.

We used the following prompt:

> I'll give you (in JSON format) a list of place names. I want you to see if there are any places that appear twice but with different names.
> Give me a JSON with a list of lists, where the inner list is the multiple names that describe the same place (and both appear in the input). No need to return unique names (i.e., lists with one element).
> Convert names only if you are positive that they are the same, e.g., different spellings or a longer description of the same place (like US, USA, America etc.).
> Make sure to maintain the exact spelling that appeared, including special characters. Make sure to give only the JSON format with no additional text.
>
> For example, if the input is:
> “‘json

<Here come some examples of lists of names describing the same place> “‘

Here is the input:
<Here comes a sorted list of the locations>

We manually proofed the resulting list leading to minor changes.

After aligning the node names, all nodes and edges were used to create a large map. We note that we applied some simple heuristic rules to sparsify the edges – we discarded edges between nodes from the same type (e.g., no edge from country to country) and edges that went against the type hierarchy (i.e., we discarded edges from country to city or from continent to country).

Plotting the maps and paths With the graphs and paths that we obtained, we used the Networkx[5] package for visualization.

4.2. Results

Here we present the statistics of the outputs and some examples of the resulting maps and paths.

We ran the pipeline on 74 testimonies from the Shoah Foundation. The average number of locations extracted from each testimony was ≈ 23 nodes and the average number of relationship edges was ≈ 17. The resulting graph had 883 nodes and 838 edges. The average length of the trajectories was 11.

In Figure 1 we display a view of the full map and a trajectory on it. Countries were enlarged for readability. In Figures 2 and 3 we show snippets around specific countries.

5. Future Work

The outputs from our pipeline can be useful on their own, such as for visualization or indexing. Moreover, the obtained map has theoretical qualities that can be further developed for additional uses.

For a pair of locations, we can define meaningful similarity measures that are based on the graph. For example, we use the distance from a common ancestor (so that two towns in Poland will be closer to each other than to a city in USA). In addition, since we extracted the types of locations, we might want to put special emphasis on Holocaust-specific locations (like ghettos and camps).

Provided with a point-wise distance measure (i.e., the distance between two locations) we can derive a trajectory-wise distance. For example, we can use Dynamic Time Warping (Vintsyuk, 1968) built upon the point-wise distance. This type of

[5]https://networkx.org/

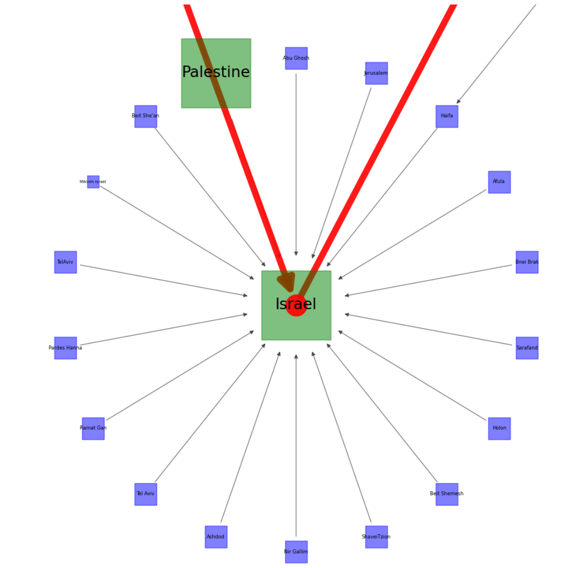

Figure 2: Snippet from the map that included Israel and locations within it. The path is from a trajectory going through Israel.

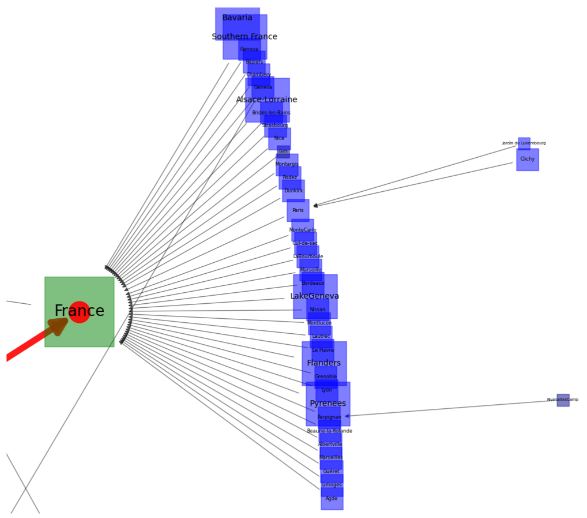

Figure 3: Snippet from the map that included France and locations within it. The path is from a trajectory concluding in France.

measure has the benefit of generating an optimal alignment between the trajectories, which in itself can be highly beneficial for Holocaust studies.

Providing a distance measure also allows us to perform unsupervised clustering based on the trajectories. The ability to cluster and align between testimonies has important implications for Holocaust research.

6. Conclusion

We presented and defined the task of zero-shot trajectory extraction. We built and demonstrated a pipeline for the task, based on GPT4. Our demonstration shows that the new models are capable of extracting meaningful trajectories from full testimonies without the necessity to break them into segments. These results suggest new ideas both for computational narrative analysis and specifically for Holocaust research.

Ethical Considerations

We followed the guidelines given by the archive. Although so the testimonies were not given anonymously, no identifying details will be included in our analysis. Our codebase and scripts will be released, but they will not contain any data from the archives. The data and trained models used in our work will not be shared with third parties without the archives' consent. To browse and research the testimonies, permission can be requested from the SF archive.

Acknowledgments

The authors acknowledge the USC Shoah Foundation - The Institute for Visual History and Education for its support of this research. This research was supported by grants from the Israeli Ministry of Science and Technology and the Council for Higher Education and the Alfred Landecker Foundation.

7. Bibliographical References

Arwa I. Alhussain and Aqil M. Azmi. 2021. Automatic story generation: A survey of approaches. *ACM Comput. Surv.*, 54(5).

Maria Antoniak, David Mimno, and Karen Levy. 2019. Narrative paths and negotiation of power in birth stories. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–27.

Ron Artstein, Alesia Gainer, Kallirroi Georgila, Anton Leuski, Ari Shapiro, and David Traum. 2016. New dimensions in testimony demonstration. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*, pages 32–36, San Diego, California. Association for Computational Linguistics.

Amanda Bertsch, Uri Alon, Graham Neubig, and Matthew Gormley. 2024. Unlimiformer: Long-range transformers with unlimited length input. *Advances in Neural Information Processing Systems*, 36.

Jiang Bian, Dayong Tian, Yuanyan Tang, and Dacheng Tao. 2018. A survey on trajectory clustering analysis.

Nicola De Cao, Gautier Izacard, Sebastian Riedel, and Fabio Petroni. 2021. Autoregressive entity retrieval.

Fred J. Damerau. 1964. A technique for computer detection and correction of spelling errors. *Commun. ACM*, 7(3):171–176.

Federico Dassereto, Laura Di Rocco, Giovanna Guerrini, and Michela Bertolotto. 2020. Evaluating the effectiveness of embeddings in representing the structure of geospatial ontologies. In *Geospatial Technologies for Local and Regional Development: Proceedings of the 22nd AGILE Conference on Geographic Information Science 22*, pages 41–57. Springer.

Ori Ernst, Avi Caciularu, Ori Shapira, Ramakanth Pasunuru, Mohit Bansal, Jacob Goldberger, and Ido Dagan. 2022. Proposition-level clustering for multi-document summarization. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1765–1779, Seattle, United States. Association for Computational Linguistics.

Evelyn Gius and Michael Vauth. 2022. Towards an event based plot model. a computational narratology approach. *Journal of Computational Literary Studies*, 1(1).

Pengcheng He, Xiaodong Liu, Jianfeng Gao, and Weizhu Chen. 2021. Deberta: Decoding-enhanced bert with disentangled attention.

Andreas Hotho, Fotis Jannidis, Lucas Kaufmann, Markus Krug, Frank Puppe, Nils Reiter, and Annekea Schreiber. 2021. Shared task on scene segmentation@ konvens 2021.

Xuke Hu, Yeran Sun, Jens Kersten, Zhiyong Zhou, Friederike Klan, and Hongchao Fan. 2023. How can voting mechanisms improve the robustness and generalizability of toponym disambiguation? *International Journal of Applied Earth Observation and Geoinformation*, 117:103191.

Xuke Hu, Zhiyong Zhou, Hao Li, Yingjie Hu, Fuqiang Gu, Jens Kersten, Hongchao Fan, and Friederike Klan. 2022. Location reference recognition from texts: A survey and comparison. *arXiv preprint arXiv:2207.01683*.

Ehsan Kamalloo and Davood Rafiei. 2018. A coherent unsupervised model for toponym resolution. In *Proceedings of the 2018 World Wide Web Conference*, WWW '18, page 1287–1296, Republic and Canton of Geneva, CHE. International World Wide Web Conferences Steering Committee.

Mayank Kejriwal and Pedro Szekely. 2017. Neural embeddings for populated geonames locations. In *The Semantic Web–ISWC 2017: 16th International Semantic Web Conference, Vienna, Austria, October 21-25, 2017, Proceedings, Part II 16*, pages 139–146. Springer.

Abhinav Kumar and Jyoti Prakash Singh. 2019. Location reference identification from tweets during emergencies: A deep learning approach. *International Journal of Disaster Risk Reduction*, 33:365–375.

John Lafferty, Andrew McCallum, and Fernando CN Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data.

Quoc Le and Tomas Mikolov. 2014. Distributed representations of sentences and documents. In *International conference on machine learning*, pages 1188–1196. PMLR.

Andrew Kwok-Fai Lui, Yin-Hei Chan, and Man-Fai Leung. 2021. Modelling of destinations for data-driven pedestrian trajectory prediction in public buildings. In *2021 IEEE International Conference on Big Data (Big Data)*, pages 1709–1717.

Gengchen Mai, Weiming Huang, Ling Cai, Rui Zhu, and Ni Lao. 2022. Narrative cartography with knowledge graphs. *Journal of Geovisualization and Spatial Analysis*, 6(1):4.

Wesley Mathew, Ruben Raposo, and Bruno Martins. 2012. Predicting future locations with hidden markov models. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, UbiComp '12, page 911–918, New York, NY, USA. Association for Computing Machinery.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'13, page 3111–3119, Red Hook, NY, USA. Curran Associates Inc.

Semi Min and Juyong Park. 2019. Modeling narrative structure and dynamics with networks, sentiment analysis, and topic modeling. *PloS one*, 14(12):e0226025.

Pinelopi Papalampidi, Kris Cao, and Tomas Kocisky. 2022. Towards coherent and consistent use of entities in narrative generation. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 17278–17294. PMLR.

Andrew Piper and Sunyam Bagga. 2022. Toward a data-driven theory of narrativity. *New Literary History*, 54(1):879–901.

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.

John W Ratcliff and David E Metzener. 1988. Pattern-matching-the gestalt approach. *Dr Dobbs Journal*, 13(7):46.

Anne-Kathrin Reuschel and Lorenz Hurni. 2011. Mapping literature: Visualisation of spatial uncertainty in fiction. *The Cartographic Journal*, 48(4):293–308.

Abdessamed Sassi, Mohammed Brahimi, Walid Bechkit, and Abdelmalik Bachir. 2019. Location embedding and deep convolutional neural networks for next location prediction. In *2019 IEEE 44th LCN Symposium on Emerging Topics in Networking (LCN Symposium)*, pages 149–157.

Özge Sevgili, Artem Shelmanov, Mikhail Arkhipov, Alexander Panchenko, and Chris Biemann. 2022. Neural entity linking: A survey of models based on deep learning. *Semantic Web*, (Preprint):1–44.

Sandeep Soni, Amanpreet Sihra, Elizabeth F. Evans, Matthew Wilkens, and David Bamman. 2023. Grounding characters and places in narrative texts.

Sharifa Sultana, Renwen Zhang, Hajin Lim, and Maria Antoniak. 2022. Narrative datasets through the lenses of nlp and hci.

Chenyu Tian, Yuchun Zhang, Zefeng Weng, Xiusen Gu, and Wai Kin Chan. 2022. Learning fine-grained location embedding from human mobility with graph neural networks. In *2022 International Joint Conference on Neural Networks (IJCNN)*. IEEE.

Taras K Vintsyuk. 1968. Speech discrimination by dynamic programming. *Cybernetics*, 4(1):52–57.

Eitan Wagner, Renana Keydar, and Omri Abend. 2023. Event-location tracking in narratives: A case study on holocaust testimonies. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 8789–8805, Singapore. 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 *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 6809–6821, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Jimin Wang and Yingjie Hu. 2019. Enhancing spatial and textual analysis with eupeg: An extensible and unified platform for evaluating geoparsers. *Transactions in GIS*, 23.

Xindi Wang, Mahsa Salmani, Parsa Omidi, Xiangyu Ren, Mehdi Rezagholizadeh, and Armaghan Eshaghi. 2024. Beyond the limits: A survey of techniques to extend the context length in large language models.

Bingsheng Yao, Ethan Joseph, Julian Lioanag, and Mei Si. 2022. A corpus for commonsense inference in story cloze test. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 3500–3508, Marseille, France. European Language Resources Association.

Albin Zehe, Leonard Konle, Lea Katharina Dümpelmann, Evelyn Gius, Andreas Hotho, Fotis Jannidis, Lucas Kaufmann, Markus Krug, Frank Puppe, Nils Reiter, Annekea Schreiber, and Nathalie Wiedmer. 2021. Detecting scenes in fiction: A new segmentation task. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 3167–3177, Online. Association for Computational Linguistics.

Zeyu Zhang and Steven Bethard. 2023. Improving toponym resolution with better candidate generation, transformer-based reranking, and two-stage resolution.

8. Language Resource References