Digging into Human Rights Violations:
Anaphora Resolution and Emergent Witnesses

A Digging into Data Challenge Project

Project Whitepaper
2015

1. TABLE OF CONTENTS

2. EXECUTIVE SUMMARY

Digging into Human Rights Violations was an international effort linking scholarly and industry investigations into the application of Natural Language Processing (NLP) techniques and tools to advance current human rights violation research. Focusing broadly on global crises, the project began in 2012 to expand models for information extraction from witness statements and government reports, for visualizing that data to facilitate better event understanding, for understanding how researchers and investigators used computational methods in their human rights work, and for developing methods to model speakers' expressed certainty of his or her statements. The results of this project appeared in publications and presentations ranging from IEEE Big Data to Digital Humanities to the Association for Computational Linguistics to the American Comparative Literature Association. By focusing on the subdomain of anaphora resolution know as cross-document co-reference, this project led to better tools for navigating large corpora by finding similar story elements throughout a corpus. This correlation of elements across a corpus helped piece together multiple fragmented narratives that enable the identification of emergent witnesses through the creation of transversal narratives.

To assist in the correlation of events through cross-document co-reference, or the recognition that a person mentioned in one report is the same as a person from another report, the project created and implemented a novel visualization technique entitled Storygraph. This technique helps investigators trace individual actors' movements over time throughout a corpus. This visualization technique developed in concert with Storygram, a tri-gram visualization that facilitated similarity comparisons of narrative elements of time, location, and person. The visualization research led to methods for comparisons of narrative fragments using graph similarity techniques, the results of which was shared via venues like the Computational Models of Narrative Workshop, the Association for Computational Linguistics, Digital Humanities.

Further investigations into the expression of individual narratives within violation reports also required research into an aspect of attribution known as speaker veridicality. Every measurement tool includes some margin of error: speech, more so than most. As witness statements move into the more quantitative realm of data, having metrics to indicate the "error," or certainty of a statement would allow investigators to better appreciate what they see. This area of further study led to the development and testing of a model for assessing the veridicality of witness statements.

Working across international borders, the project personnel met with human rights organizations and researchers in Switzerland, the Netherlands, Canada, Central America, South America, and the United States. The greatest impact of this project lay in its potential to introduce computational research methods to human rights investigations. Specifically, the project personnel were able to offer guidance on methodologies for navigating large corpora of witness testimony, mining those texts for relevant information using existing NLP techniques, and visualizing that information for further analysis.

Over the three years of this project, many undergraduate and graduate student researchers contributed to the project's outcomes. These students, many of whom have now gone on to more advanced degrees and positions, contributed skills from a variety of disciplines: computer science, applied linguistics, computational linguistics, interactive computing, English, Spanish, communication studies, political science.

3. INTRODUCTION, GOALS, OUTCOMES

3.1. Introduction

A survivor, at some point after the event that nearly took his or her life, speaks. After being recorded, transcribed, and collected, the resulting media are key portals for the historical and human understanding of memory, trauma, historiography, and violations of human rights both personal and massive. Since the Federal Writers Project collected 2,300 Ex-Slave Narratives in the 19030s, these collections have been seen as necessary and, as such, have grown. Some, like the Shoah Foundation video testimonies of Holocaust survivors, range into the many hundreds of thousands of hours. Others, like the very granular event and organizational history captured by the World Trade Center Task Force Interviews, are relatively constrained at 1.6 million words. Between those two extremes lay global collections documenting the worst of our history, a history made opaque because of the scale of the collections produced to preserve and redress it.

Human rights violation records are both essential and problematic to studying the events and memories they represent. These texts contain information that is key to understanding the human rights implications of past and ongoing events and can assist in identifying current and future human rights violations in various contexts. These records, however, are contained within heterogeneous documentation such as field reports, witness testimony, and court documents by various entities including state organizations, NGOs, and individual witnesses, perpetrators, and victims. The heterogeneous nature of this documentation and its documentation standards by various entities makes the correlation of multiple narratives relating to a given event difficult; but, all of these individual narratives are necessary for constructing a more holistic collective memory of the event.

Developing natural language processing (NLP) tools and techniques for the reading at scale of individual heterogeneous documents within a given human rights violation corpora facilitates the recognition of stories and individuals that run across a collection. The premise for this mode of reading is simple. In every population data set, there are duplicate records. Likewise, in every collection of testimonies that is dense relative to the temporal and geographic constraints of the event, there are individuals and places that appear multiple times. Sometimes, these individuals are properly named such as when multiple members of a family each give their own testimony. Other times, their appearance is marked only by a pronoun. This pronominal reference, known as an anaphoric reference because of the way specific meaning is carried forward, backwards, or across documents by a general word, might serve as the only marker of an individual who did not survive to tell his or her own story. Transversal reading across a collection might be able to stitch together the stories of those that either did not survive or have chosen for various reasons to not reveal their story.

To that end, this project developed and implemented a set of NLP tools to identify, extract, and group narrative elements of time, person, place, and event and resolve the anaphoric references that occur across documents. Once identified and resolved, these elements can be correlated from across multiple documents throughout a given corpus to reveal a group narrative. In this process, cross-document co-references can be extended from the document level to that of the corpus by correlating entities across multiple narratives. However, without a robust visualization technique, the resulting data would be as opaque as the original collections and the goal of transversal reading would fail. The ability to visualize this disparate data as connected lines that show the story of individual entities and the provenance of that story was a key component of this work.

This project was driven by four core research questions; (1) how can one mine large text archives of witnessing to identify hidden victims and unnamed perpetrators, (2) how can a reading workbench facilitate the coherence of meaning using related fragments drawn from large volumes of free text, (3) how can computational linguistics and Natural Language Processing (NLP) help in this identification and coherence, and (4) how can the results of text-mining best be visualized to reveal genealogical connections amongst reports. These questions drove the development of methods to help organize obscure fragments of text. NLP and visual analytic techniques, informed by the survey and questionnaire work of the project, helped identify and present material so as to reveal emergent, formerly silent witnesses to violations.

3.2. Goals

The core goals that initiated this project grew as it became more apparent that certain kinds of language understanding were necessary. These goals are as follows.

- Understand how researchers and investigators use computational methods to facilitate their work with human right violations data
- Identify related fragments of text within documents from across an archival collection relevant to human rights research and truth and reconciliation contexts
- Contribute to the current analysis of human rights violation reports by developing and augmenting techniques for sparse or imprecise information correlation
- Improve upon existing anaphora resolution techniques in NLP for cross-document co-reference
- Develop an automated reader for large text archives of human rights abuses that will reconstruct stories from fragments scattered across a collection
- Design an interface for navigating the stories within the human rights violation corpora using the automated reader
- Facilitate visualizations of the results of text-mining to reveal genealogical connections across large numbers of documents as well as patterns over time and location in corpora.
- Identify phrases indicative of a speaker's level of certainty and incorporate values representing that certainty into the text's metadata for analysis
- Connect researchers in the computational social sciences and humanities concerned with human rights research and investigation

3.3. Outcomes

Throughout this project, our team reached many of our original goals as well as additional relevant outcomes across various disciplines. These outcomes are as follows.

- Augmented big data, graph similarity, and computational narratology methods for use within a human rights domain
- Facilitated the connection of human rights organizations and scholars focusing on text mining and data analysis
- Collected and interpreted data from surveys and interviews with researchers and investigators concerned with human rights violations
- Developed a corpora-level open-source data visualization tool (Storygraph) for rendering individual data sets of person at time at location

- Developed a tri-gram visualization (Storygram) for suggesting links between narrative elements and facilitating the comparison of those elements
- Established a working model for comparing narrative elements using graph similarity methods and techniques
- Produced a pipeline for using existing NLP tools for the identification and extraction of narrative elements
- Develop a list of features to identify the verticality of statements relative to a human rights violation context within a given text
- Contribute to relevant scholarship in a variety of disciplines based on project results including computer science, computational linguistics, applied linguistics, literary studies, and communications

4. PROJECT TEAM, US AND CANADA

4.1. Faculty

Ben Miller (Overall Project Lead), Assistant Professor, Georgia State University
Lu Xiao (Canadian Team Lead), Assistant Professor, Western University
Karthikeyan Umapathy, Associate Professor, University of North Florida
Fuxin Li, Research Scientist, Georgia Institute of Technology
George Pullman, Associate Professor, Georgia State University
Ben Kiernan, A. Whitney Griswold Professor of History, Yale University
Nancy Marrelli, Archivist Emerita, Concordia University
Dan Jurafsky, Professor, Stanford University
Mary Beth Rosson, Professor, Pennsylvania State University
Ágnes Sándor, Project Leader, Xerox Research Centre Europe (XRCE)

4.2. Graduate and Postdoctoral Research Assistants

Lindsay Baker, MIS candidate, Western University
Patrick Crenshaw, MA student, Georgia State University
Jill R. Kavanaugh, MLIS candidate, Western University
Kristopher Kyle, PhD candidate, Georgia State University
Shakthidhar Gopavaram, PhD Student, University of Indiana - Bloomington
Jennifer Olive, PhD student, Georgia State University
Ayush Shrestha, PhD, Georgia State University
Nicholas Subtirelu, PhD student, Georgia State University
Tatiana Vashchilko, Research Associate, Western University
Vicky Yan, MLIS candidate, Western University
Jin Zhao, PhD, Georgia State University
Yanjun Zhao, PhD, Georgia State University

4.3. Industry Partners

Benetech Human Rights Project (Palo Alto, CA)
Human Rights Data Analysis Group (San Francisco, CA)
Invisible Children (Los Angeles, CA)
The Resolve (Washington, D.C.)

5. WHAT IS HUMAN RIGHTS VIOLATION DATA?

5.1. Why Human Rights Violation Data?

Information surrounding human rights violations is often fragmented, scattered, and hidden within the many narratives that describe them. From various witness perspectives to various document types and formats, the corpora of heterogeneous data regarding human rights violations make it difficult to effectively stitch together related pieces of information. From the size of these heterogeneous corpora to their inherent aversion to more rudimentary search practices such as keyword-based searches, human rights violation corpora present new and interesting challenges to existing textual and linguistic analysis methodologies. Yet, interaction with these challenges promises to yield new insights not only to existing scholarship on textual analysis but also in the identification and analysis of human rights violations and those individuals involved.

This project sought to address these concerns regarding large corpora of heterogeneous human rights violations data by providing an automated pipeline for identifying and connecting related fragments of information across multiple documents throughout a given corpus. Using existing tools from natural language processing (NLP), methodologies from computer science, and theoretical frameworks from narratology, this project expands on previous human rights and textual analysis scholarship to provide a way of *reading* rather than *searching* large corpora of heterogeneous documents. Specifically focusing on the application of anaphora resolution and cross-document co-referencing NLP techniques, this project proposes a method of horizontal reading that identifies and connects information from narrative elements across multiple documents within a given corpus. This method encourages data from individual narratives to be seen as pieces of the larger corpora, a perspective that enables the potential composition of collective memories and identification of emergent witnesses. The results of this reading can then be fed into a data visualization program developed for this project entitled *Storygraph* for visualization and analysis by members of the academic community as well as other human rights activists in various roles.

5.2. Obtaining Corpora

For this project, we used several corpora from different known human rights violation events in order to produce a methodology that would work with shared features of human rights violations documentation. Some of the corpora used by this project were publically available, and some were provided to this project by various custodian organizations. In cases where the data was confidential, it was retained in an encrypted, secure manner. The diversity of documentation was necessary to develop language models that were sensitive to each type of report. Additionally, the variation in language patterns helped the project recognize the significance of automated veridicality, or certainty measurements. The goals, procedures, and historicity of a trial transcript or organizational history are vastly different, for example, from those of an individual witness statement collected as part of an outreach effort.

The corpora used for this project included the following.

- Lord's Resistance Army NGO reports and statements (~3,000 reports and documents)

- Extraordinary Chambers in the Courts of Cambodia (2,672 court documents and transcripts)
- South African Truth and Reconciliation Commission (2,004 court documents and transcripts)
- World Trade Center Task Force Interviews (503 interviews)
- Bosnian Historical Memories (84 life stories)

5.3. Heterogeneous Materials

The corpora used often contained heterogeneous documents related to their respective human rights violation situations. These corpora were likely to contain ranges of documentation types including court documents; survivor, witness, and perpetrator testimonies; field reports; police reports; interviews; etc. These documents had to be reviewed and evaluated for material that could be used for the purposes of the project. For example, court documents relating to the procedures of the court were not relevant to our analysis and, thus, needed to be removed from consideration.

In addition to the heterogeneous document purposes, we also encountered various document types within a given corpora including text documents, PDF, DOC, and XLS. To feed these documents into our pipeline, each of these documents had to be converted into UTF 8 encoded text files. For some of this documentation, the documents could be converted using a script; however, some of the documents that included special characters or were stored as images needed extra processing such as OCR. The diverse formats indicated the diverse time periods captured by these documents, their geographically diverse production, and the diverse language groups represented. Although the project was limited to English and Spanish language documentation, Cambodian and South African names of people and places densely populated their respective corpora.

5.4. Narrative Elements in Witness Testimony

The corpora used in this project are composed of individual narratives from victims, perpetrators, and/or witnesses of potential or confirmed human rights violations. Traumatic memories, as Cathy Caruth describes them, are experiences that were, paradoxically, too immediate to be fully incorporated into an individual's understanding of a given situation, which leads to belated and recurring attempts to incorporate it into one's understanding (1995). The act of composing a narrative is one such exercise in bearing witness to the "impossible history" within the individual. The narrative itself functions as an organization and reference mechanism for the traumatic memory in these situations by using language as a tool.

Framing for the data extracted and grouped for analysis within these corpora consisted of narrative fabular elements in varying detail. This division of narrative elements finds its roots in the Russian formalist tradition as recounted by Mieke Bal in her seminal work *Narratology*. This understanding of narrative divides its contents into three layers (text, story, and fabula), which are each composed of its own elements and aspects. The fabula layer of the narrative deals with the higher-level concepts of narrative content whereas the text and story levels primarily deal with the organization and representation of those fabular elements. Bal identifies the fabular elements as person/actor, time, location, and event, which should remain the same in a given narrative despite the multiple possible manifestations of the text.

6. A PIPELINE FOR PROCESSING SPATIO-TEMPORAL, ENTITY, AND EVENT LANGUAGE

The major goal of the project is to facilitate transversal reading of human rights violations documentation so as to discover stories implicit within the corpora. To that end, cross-document co-reference techniques for fuzzy reference matching were required. Those techniques extract the named and temporal entities from multiple documents within the corpus and correlate them. Correlating the entities not only help to identify the victims and perpetrators across documents but also provide a better understanding of *what* and *when* events were said to take place.

The first phase of the pipeline involves cleaning and preprocessing the documents. This phase ensures that the documents are digitized, in the correct format for the code to process, and contain no characters that would violate the standard character sets. The steps in this phase contain the most critical and time consuming portion in the pipeline and require a decent amount of manual work.

The second phase includes feeding these sets of documents through the temporal tagger, named entity recogznizer and event parser. These are three independent pieces of software running in parallel. The named entity recognizer tags all the names of people, locations and organizations mentioned in the document that it can find, and the temporal tagger marks all instance of absolute time reference. The output from the NER further feeds into anaphora resolver. The anaphora resolves all the pronouns with the found proper nouns. The event tagger marks all the 'events' in the document.

In the third phase, the program gathers and merges the data from these three different sources and stores it in the database.

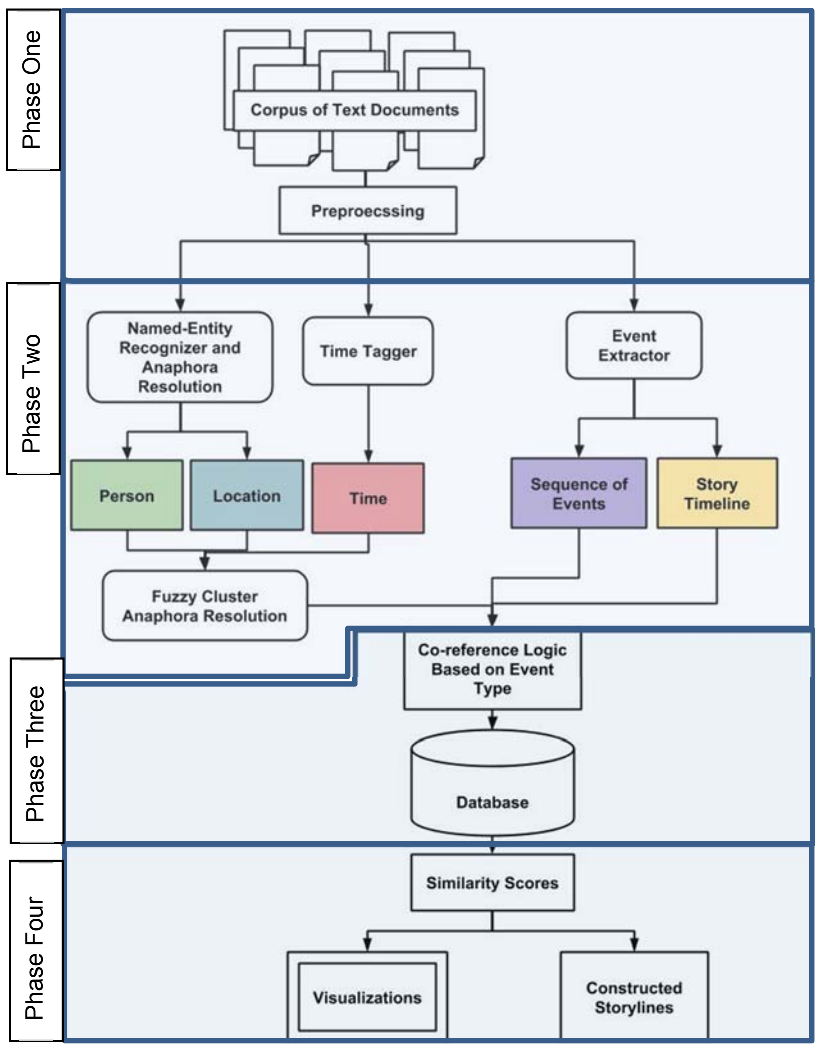

Figure 1: Language processing pipeline for resolution of cross-document co-reference via fuzzy clusters

After all the documents have been stored in the database, a document is selected and checked against the remaining documents for event similarity. An event is a representation of an entity at a location at a given time. In case the events are deemed similar, the characters and temporal tags are then checked to see if they are similar. Finally, a similarity score is produced for two characters. This process is repeated iteratively and exhaustively till all the similarity scores are computed between all the characters. Finally, the characters beyond an adjustable threshold are suggested as being the same person. Ultimately, this tool only makes suggestions to human investigators; it narrows the information field, rather than authoritatively concluding that individuals in a given corpus are self-similar.

At the end of the pipeline and visualization processes, the project was able to explore the movements of characters across the time and space of a given event context as delimited by the stories describing that context. Described in detail below in section 8, the project's visualization tool, Storygraph, yielded examples like Figure 2. Although difficult to interpret at first, this method allows for people to see, for example, the chaotic movements and chaotic storytelling that predominated the early morning of September 11, the stabilization of the scene as triage facilities were established and first responders worked to rescue and reassure people, and occasionally, the tragedy as life lines were cut short by the falling towers. Each line represents the pathway of one individual through the story. When the line ends, sometimes, that's when their part in the story ends. Other times, it represents something more tragic.

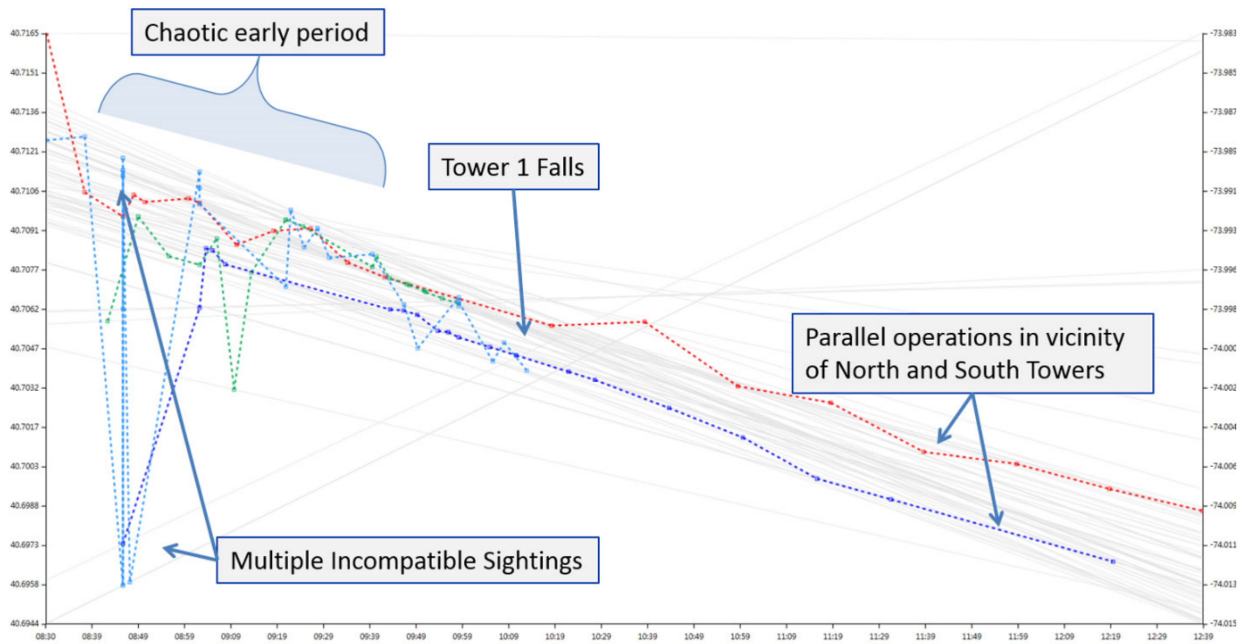

Figure 2: Storylines data from 503 first responder testimonies to the World Trade Center attacks of September 11, 2001, showing individual pathways through the spatiotemporal frame described by all interviews irrespective of the document in which the person appears

7. MEASURING CERTAINTY: WITNESS ATTRIBUTION AND VERIDICALITY

The overall goal of this piece of the overall project is to create a program that can assign a classification with regards to interpreted speaker certainty or veridicality. Events will be identified using EVITA (Saurí et al 2005). The classifier will then categorize each event into one of nine categories. One set of categories will be for positive or non-negated events: Certain+, Probable+, and Possible+. Another will be for negative or negated events: Certain-, Probable-, and Possible-. The final set will deal with events for which certainty is not sufficiently specified. A positive event is one a witness says happened; a negative event is one a witness said did not happen. The classifier will be trained using data from a selection of publicly available documents describing rights violations with judgments on that data collected from Amazon Mechanical Turk (mTurk) users. The machine learning approach will use semantic-syntactic features derived from computational linguistic tools (e.g., negation and tense) as predictors.

7.1. Background

The problem that we are addressing involves the linguistic concept of event certainty or veridicality. This concept arises from the observation that besides communicating propositional information, such as who did what to whom, language users routinely communicate other information about their propositions such as their attitudes toward the propositional information (Hyland, 2005). In particular, while every description of an event contains some combination of propositional information, specifying actors, acts, time, and location, language users make use of various linguistic mechanisms that allow them to commit more or less strongly to the information they present in their utterances.

For example, Hyland distinguishes between hedges and boosters. Speakers and writers use hedges to communicate that they are not fully certain about a proposition. Hedges are commonly communicated through features like modal verbs (e.g., *might* and *may*) or adverbs (e.g., *perhaps* and *possibly*). In contrast, boosters allow the speaker to more fully commit to being certain about a proposition. Commonly, boosters are adverbs such as *definitely*. These examples, however, are far from exhaustive and past computational research (Saurí and Pustejovsky, 2009; de Marneffe et al., 2012) suggests that a large inventory of linguistic features is necessary to even begin approaching a comprehensive description of how language users communicate veridicality.

More specifically for our purposes, veridicality presents an issue for research projects attempting to use natural language processing to make quantitative generalizations about the contents of corpora, a common research process in the digital humanities. For example, our own research attempts to derive summative analyses of human rights violation events by automatically processing the testimony of many eye-witnesses stored in digitized corpora. One potential threat to the validity of these analyses is the fact that the tools we use, such as named event recognition (using EVITA), merely identify potentially relevant elements in discourse without regard for whether the speaker is certain about them or not. Thus, such tools may not make any distinction between the events represented by the verb *kill* in (1) and (2) below, although, for our purposes, the two need to be separated.

(1) He definitely did **kill** that man.
(2) I'm pretty sure he didn't **kill** that man.

In processing our corpora, we would want to mark (1) as a report of a human rights violation, while (2) should not be. Automatically distinguishing between these two requires a natural language processing tool that is able to make judgments about veridicality. Past research has already been undertaken on this subject (Saurí and Pustejovsky 2009; de Marneffe et al. 2012), but the text types used have been highly constrained: in both cases, newspaper reports. Our data, transcribed oral interviews, differ greatly. This project, then, extends past research by attempting to use already-established computational approaches in a new context.

7.2. Questionnaire

Following the example of Saurí and Pustejovsky (2009) and de Marneffe et al. (2012), a questionnaire was created to gather judgments of certainty from mTurk users.

First, sentences were extracted from different corpora of interviews from different contexts: the South African Truth and Reconciliation Commission, the Cambodian Khmer Rouge Tribunal, interviews with survivors of the Holocaust, statements from survivors of the Rwandan genocide, and interviews with survivors of ethnic cleansing in the former Yugoslavia. Each sentence was tokenized using Stanford Core NLP. Using EVITA, events in each sentence were tagged. Only sentences containing OCCURRENCE or STATE event tags were retained as candidates for the questionnaire. A random sample of sentences contains about 800 events (many sentences contain multiple events) was taken from each of the five corpora leading to a total of approximately 4000 events.

Second, sentences were pre-processed so that each event inside of them was identified and a different iteration of the sentence was created to highlight (through **bolding**) the individual event. In this manner, a sentence that contained two events like (3) below would be rendered as two items in our questionnaire: (4) and (5).

(3) He ran up the street and jumped on the train.
(4) He **ran** up the street and jumped on the train.
(5) He ran up the street and **jumped** on the train.

As previously stated, a total of 4000 items similar to (4) and (5) above were contained in our overall bank of items for our questionnaires.

Third, a series of different versions of the questionnaire were created that first trained mTurk users in how to make judgments about certainty in the sentences and then asked them to place items like (4) and (5) above in one of nine categories presented in Table 1 below.

Table 1: Veridicality categories

Positive or non-negated	

	1. Certain +	According to the speaker, it is certainly the case that…
	2. Probable +	According to the speaker, it is probably the case that…
	3. Possible +	According to the speaker, it is possibly the case that…
Negative or negated		
	4. Certain -	According to the speaker, it is certainly <u>not</u> the case that…
	5. Probable -	According to the speaker, it is probably <u>not</u> the case that…
	6. Possible -	According to the speaker, it is possibly <u>not</u> the case that…
Under-specified		
	7. Certain but under-specified	The speaker knows but does not fully communicate whether or not it is the case that…
	8. Uncertain	The speaker does not know or does not commit to whether or not it is the case that…
Error		
	9. NA / Error	There is something wrong with the sentence.

Fourth, the questionnaire items were presented to mTurk users. 50 versions of the questionnaire were created each containing (1) several training questions, (2) a randomly selected set of 50 items from the bank of items described above, and (3) items designed to screen out users not completing the questionnaire in good faith. 10 mTurk users completed each version of the questionnaire. If a user was found not to have completed the questionnaire in good faith (they did not specify the correct answer to items that contained their own answers “Please select *x* as the answer”), then their responses were removed and replaced with an additional participant.

7.3. A Model of Event Veridicality in Witness Statements

The previously described steps have been completed, and we are now in the process of analyzing the data. Preliminary analyses suggest that the data we have gathered so far does not provide numerous examples of each category. Table 2 presents a summary of classifications made by mTurk users. The totals presented for each category represent the number of items for which 6 or more of 10 mTurk users agreed on a particular classification for an item.

Table 2: Category frequencies (agreement from 6 or more mTurk users)

Certain+	1083

Probable+	370
Possible+	21
Certain-	18
Probable-	5
Possible-	0
Certain but Underspecified	0
Uncertain	0
Error	0

We are currently developing a program that will process each item used in the questionnaire in order to detect the presence of a number of linguistic features associated with veridicality in past research (Saurí and Pustejovsky, 2009; de Marneffe et al., 2012), for example, negation, modality, and tense. We are also considering collecting more data strategically to gather more instances of more weakly represented categories especially those in the negative or negated categories.

Ultimately, we will use a machine learning approach to classify items into certainty or veridicality categories according to the linguistic features found in the sentence.

7.4. Crowdsourcing Measurements

We used Amazon’s Mechanical Turk (mTurk), a web service that provides us with on- demand workforce, for hosting our survey to collect crowdsourced judgments on the level of certainty readers attributed to different phrases. We chose mTurk because it supplied us with the framework and the tools necessary to collect quality data while preserving anonymity of the participants. Section 7.4.1 briefly describes the key concepts and terminology required to understand mTurk, and Section 7.4.2 describes the design of the survey.

7.4.1. Key Concepts

This section describes the key concepts and terminology of mTurk.

Requester: A requester is a person or organization that creates and hosts the surveys or tasks for the workers on Amazon to perform. A requester has access to a software application to host surveys, options for reviewing and retrieving results, and a built-in payment scheme.

Worker: A worker is a person who takes the survey or performs the tasks posted by the requester. Amazon’s mTurk preserves the privacy of the participants by using worker ID’s rather than the workers name on the submitted tasks.

Human Intelligence Task (HIT): A HIT is a task or a survey that a requester hosts on mTurk for the worker to complete. Each HIT has a lifetime, specified by the requester, which determines how long the HIT is available to workers. A HIT also has an assignment duration, which is the amount of time a worker has to complete a HIT after accepting it. Each HIT also has limited number of assignments, which determines the maximum number of workers that can submit completed work.

Assignment: An assignment represents the work submitted by a worker. Each assignment belongs exclusively to a single worker.

Reward: It is the compensation a requester pays the workers for successfully completing the HIT. A requester is given the opportunity to review the work submitted by the Worker and approve or reject it. The worker gets paid only if his work has been approved. For more information about mTurk please visit http://aws.amazon.com/documentation/mturk/.

Known Answer Question (KAQ): A KAQ is a test question with a known answer designed to test whether workers are reading the items before responding. A KAQ will instruct a worker to "please select *x* as the answer". For these questions, the worker should simply follow the instructions and select *x*. If a worker fails to do so, then his/her assignment is rejected.

7.4.2. Survey

This section describes the design of the survey.

The survey consists of fifty HITs and each HIT consists of four pages. The first page is a preview page, which gives the workers a brief description about the work that needs to be done in order to successfully complete the HIT. After going over the preview page, if a worker decides to work on the HIT, he or she can accept the HIT. Once a worker accepts the HIT, he/she is redirected to the Consent page which displays the terms and conditions of the survey. After reading the terms and conditions, if the worker choses to volunteer for the survey, he/she can select the "I Agree" option and click on the "continue" button to move on to the next page. Otherwise, he/she can return the HIT.

The next page, Instructions and Training page, displays the instructions along with five training questions, which will prepare the workers for answering the survey questions. Once a worker answers all the training questions, he/she can move onto the survey questions, displayed in the Survey page, by clicking on the "continue" button.

The Survey page consists of forty nine survey questions and one known answer question (KAQ). After answering all fifty questions, the worker can click on the "submit" button to submit his/her work (also known as an assignment). mTurk then retrieves the submitted assignments and reviews the results. During the review, it checks the answers for the KAQ. If a worker fails to answer the KAQ correctly then his/her assignment is rejected. Otherwise, it is approved.

In addition to these parameters, each HIT has a maximum of ten accepted assignments, and a worker cannot work on the same HIT more than once.

8. REPORT ON INTERVIEW AND QUESTIONNAIRE STUDIES

8.1. Introduction

Human rights researchers use a wide range of data analysis techniques, from statistical analysis and modeling to examining correlations among different factors of human rights violation to studying the impact of human rights violation events ethnographically. Researchers also interact with relevant data in various formats (e.g., textual, audio, and video) and conduct both qualitative and quantitative content analyses. These data are increasingly accessible as Internet-based resources; prominent examples include the Center for Human Rights Documentation and Research at Columbia University, the Human Rights Documentation Initiative (HRDI) at the University of Texas at Austin, and the University of Connecticut Human Rights film collection.

While these trends in the generation and access of human rights violation data offer more data analysis opportunities to human rights researchers, they also present challenges in the qualitative data analysis because of the ever-growing size of the data resources. Computer scientists and digital humanities scholars have started to explore computational approaches to address these challenges (e.g., Miller, Li, Shrestha, and Umapathy, 2013). However, there is limited understanding about human rights researchers' data analysis practices.

Addressing this gap, we interviewed human rights researchers and conducted an online questionnaire to reach out to more human rights researchers to understand the characteristics of the data they analyze and their data analysis and management practices such as their experiences with data analysis software programs. We also explored their expectations with respect to a qualitative data analysis software program.

8.2. Research Methods

8.2.1. Interview

Our interview had four sections. The first section was about the interviewees' research background and the data that they have worked with in the research. For example, we asked interviewees to describe the data they often analyze such as the format (photos, text, video, etc.), the size, the type (primary or secondary data), and the source (publicly available or private). In the second section, we asked the interviewees to think of a concrete example of data analysis in their projects and describe the analysis process. We also asked them to provide details like what they were looking for in the data, whether they used software programs for analysis, and, if so, the name of the programs and the software program features that they found most useful in the analysis. Interested in how researchers coordinate on shared data during the analysis process, we asked them how many people are generally involved in data analysis projects, and, if there are indeed multiple analysts, we asked how they coordinate with each other in the process of analyzing data and integrating the results. In the last section, we sought their expectations on a software program that would support qualitative analysis of human rights data, either small or large scale. We asked for concerns or issues that should be addressed, three most important design features desired in such a program, and features they considered ineffective and should be excluded (based on their experiences).

Through purposive and snowball sampling techniques, we interviewed 13 North American academic and non-academic researchers of human rights violation research. For example, we emailed the authors of the reviewed papers who have either Canadian or American contact information, and emailed the researchers whom we know have been doing human rights violation research. Some interviewees provided additional contacts. All interviews were conducted in a semi-structured format. The interviews were 60 – 90 minutes long except one interview which lasted 25 minutes because the interviewee's background and experiences were different from what we were looking for. One interview was via phone, one via Skype, and the rest were conducted face-to-face at locations of the interviewees' choices (e.g., their offices). The first author and a research assistant were both present in the interviews taking the roles of asking questions and taking notes. All interviews were audio recorded. Interviewees were compensated CAD $25 or USD $25 for their time. One researcher kindly refused to accept the compensation after being interviewed, which we understood as a voluntary support of our study.

We analyzed the interviews through thematic analysis. Specifically, we identified salient themes in individual transcripts and searched for repetitions within and across narratives and field notes (Ryan & Bernard, 2003).

8.2.2. Online Questionnaire

Our questionnaire had 19 questions that reflected the four sections of the interview: background, data, analysis, and QDA software programs. We used several resources to reach out to potential respondents. First, we sent the email invitation (with the web link to the online questionnaire) to all authors who have articles published in *International Journal of Human Rights* (publications years 2010 – March 2013), *Human Rights Law Review* (2005 - March 2013), *Human Rights Review* (2000 – June 2013), *Journal of Human Rights* (2010 – April 2013), and *Journal of Human Rights Practice* (2009 - March 2013). Due to a limited number of responses we received from this approach, we contacted non-profit organizations in human rights research or human rights activities. One major resource we used to identify these organizations is the Human Rights Web Archive, an initiative by Columbia University Libraries and Information Services (CUL), and its Center for Human Rights Documentation & Research (CHRDR). We selected the organizations that communicate in English or French (as our questionnaire has English and French versions) and sent our email invitation with the web link.

As indicated in the information letter (that is the first page of the online questionnaire), we donated CAD$2 to United Way for each completed questionnaire. In total, we sent out 772 email invitations (this did not include two reminder emails for each non-responded invitation) and received 145 responses. We discarded 3 that did not consent to the study (as no data was provided), and 29 that gave their consent to the study but did not respond to any questionnaire item. Of the 118 responses, 73 were complete.

8.3. **Findings**

8.3.1. Human Rights Researchers' Academic Background and Experiences

Our interview and questionnaire results illustrate that researchers working on human rights related issues have diverse academic backgrounds. The thirteen interviewees, including two non-academic researchers, covered eight disciplines such as political science, statistics, and sociology. Among our questionnaire respondents, 54% are academic researchers, 18% non-academic researchers, and 27% selected "other" and provided their profession (e.g., lawyer, independent consultant, and program director). The respondents represented close to 30 fields with the most common fields being law, political science, international studies, sociology, and anthropology.

The researchers we interviewed have a wide range of experience, ranging from a few years to over 35 years. When asking for this information in the questionnaire, we provided several options from “0-5 years” to “20+ years”. The median was 6-10 years. 18 respondents had 11-15 years of experience, and 14 respondents had over 20 years of experience. We consider the questionnaire respondents to be modestly experienced in general in performing human rights research.

8.3.2. Data Characteristics in Human Rights Research

All but one interviewee has dealt with primary data, with some being publicly accessible, some required special access. These data varied in formats and include text, number, audio/video, and artifact (e.g., images, museum objects). They came from various sources, such as police records, court decisions, U.S. Department of State Human Rights Reports, and interviews conducted by the researchers themselves. The size of the data in the qualitative analysis is within hundreds of pages (e.g., 40 interviews, documents of 20 countries, etc.).

Our questionnaire result is somewhat similar, whereby out of 77 respondents, 48% reported to have used both primary and secondary data, 26% used primary data only, and 25% only used secondary data. About 56.8% of the respondents reported to have used public data, 30.5% have used private data, and 12.7% have used semi-public data. Respondents used text, images, videos, audios, and artifacts. Some provided additional details about the data, e.g., discussions, focus group or interviews, site visits, and even participatory research where victims doing their own research.

Hafner-Burton and Ron (2009) posited that most scholars of human rights draw on two key sources of cross-national data: the Political Terror Scale (PTS) and the Cingranelli and Richards Index (CIRI), noting that they are the most complete cross-national sources for violations of personal integrity rights data for many countries. However, our questionnaire results show that there was only one response that noted the use of PTS datasets. In fact one respondent explained that he/she would not use PTS because their processes are not reliable and have validity issues. Additionally, the results suggest that the researchers used various data resources. 51 of 78 respondents (65%) used data from Amnesty International, 39 of 78 respondents (50%) used private interviews, and 27 of 78 respondents (34%) used data from the U.S. Department of State Country Reports on Human Rights Practices. Respondents who used publicly available interviews were 26 of 78 (33%), while 19 of 78 respondents (24%) used data from Truth and Reconciliation Commission Reports. Also, 16 of 78 respondents (20%) used CIRI Human Rights Dataset (20%), and 6 respondents out of the 78 who responded to this question (7.6%) used The WomanStats Project. In addition to the cases of data sources that we provided in the questionnaire, respondents were also given the option of listing ‘other’ data sources that they use for their human rights research projects. From this, 53 of 78 respondents (68%) noted that they used other data sources, and listed examples such as - academic literature, fact descriptions in case law, journal articles; HRW reports; legal decisions, government statistics, media reports, NGO reports, newspapers and newspaper articles, budgets, world bank reports, social media, email communications, in person interviews, research missions, country reports outside US, UN reports, shadow reports, on-line data bases assembled by the World Bank, IMF, and the UN, data bases on grave violations against children's rights, court records and government statistics (where available), survey data, on-line administrative data, research and news publications, etc. A respondent also noted that they draw on research that comes in through the equality and human rights helpline and Citizens advice Bureau, which helps them to monitor emerging trends, while another remarked that they used data from Anti-bullying Alliance, Article 19, British and Irish Legal Information Institute, British Institute of Human Rights, Care Quality Commission, Children’s Commissioner, Crown Prosecution Service, Human Rights Watch, Joseph Rowntree Foundation, Justice, Law Commission, Law Society, Legal Action Group, LexisNexis, Local Government Ombudsman, Low Pay Commission, National Housing Federation, Nursing Standard, Open Democracy, UK Human Rights Blog, We are Spartacus, Westlaw etc. These ‘other’ sources were categorized into academic literature or publications, media or news sources, international institution data (e.g. World Bank, UN), legal documentations and proceedings, databases, and government publications.

Original data collected by the researcher was another common response that our questionnaire respondents stated they used for their data analysis human rights projects, which were not listed in our instrument.

With respect to the size of the data, we asked “How many items do you analyze in a project? Please select all that apply”. Out of the 79 researchers who responded to this question, 34 (43%) reported to that they analyzed less than 50 items in a single project, 25 (31.6%) analyzed between 50 to 100 items in a project, 12 people (15.2%) analyzed 101 – 500 items, 8 people (10.1%) analyzed between 501 to 1000, and 10 (12.6%) analyzed more than 1000 items in a project. Note here that instead of the term *document,* we used *item* because we expected that not all researchers analyzed documents in the project. For example, some may analyze artifacts such as images or videos, or some may analyze numerical data. Some researchers specified the items used in their projects. These include cases, interviews, videos, reports, survey responses, judgments or decisions, information on a violation that happened or is alleged to have happened, indicators such as the percentage of children that are malnourished, and additional specific responses. In our next questionnaire item, we asked the respondents to define what an item meant in their study. The respondents gave varying definitions. To some, data items are a primary document or a video. For others, data items are “cases” or “index value per year and country.” One respondent noted that, “the ‘*item’ for me is every evidence or indication that can provide information on the question of participation of Congolese civil society to mechanisms of human rights at local, national and international level*.” Another researcher stated that items are predominantly aggregates of human rights measures, with datasets often having more than 3000 data points to them. Furthermore, in response to what defines an item some respondents expressed that even the term ‘item’ was ambiguous so they could not specify (e.g., “*it could be any piece of information*” one respondent noted,”), some noted that items in they analyzed in their researches included cases, academic analysis, NGO reports, government reports and human rights organization reports.

We asked the researchers to indicate whether they analyzed primary data, secondary data, or both primary and secondary data in their human rights research projects. 21 of 77 respondents

(27.3%) noted that they analyzed primary data only, while 19 (24.7%) analyzed only secondary data. 37 (48%) respondents analyzed both primary and secondary data in their research projects.

We also asked the researchers to select if they considered their data sources to be public or private. From the response to this question, 67 of 77 respondents (87%) selected that their data sources were private, 36 of 77 respondents (46.8%) selected public, and 15 of 77 respondents (19.5%) selected semi-public. Table 1 shows the examples of data sources reported by the respondents, which they categorized into public, private and semi public

Table 3: Categories of data sources reported by respondent

Public data sources	**Private data sources**	**Semi-public data sources**
Online availability	Interviews and witness testimonials (individual confidentiality)	Academic literature/publications
Open datasets (government and other institutions)	Subscription	Media/news sources
Judgments/legal decisions	Researcher rights/ownership	International institutions' data (World Bank, UN)
Government and institution-published data (UN, World Bank, IMF)	Unpublished/internal information	Legal documentation/proceedings
	Personal collection	Databases
Published literature/articles		
		Government publications
Public reports (including news sources)		
		Data collected by researcher

8.4. Data Analysis Practices in Human Rights Research

8.4.1. Methods

The interviewees reported the use of quantitative and qualitative analysis techniques, with the majority of the interviewees focused on the qualitative analysis approach; however, for the questionnaire, about 47% of the 74 researchers who responded to this question reported that they used both qualitative and quantitative approaches in analyzing their human rights violation data, while 5 of 74 respondents (6.7%) and 34 of 74 (46%) researchers used only statistical and qualitative methods respectively.

8.4.2. Data Analysis Software Programs

A variety of software programs have been developed to facilitate the qualitative data analysis process. Both the interview and questionnaire results suggest that the use of qualitative data analysis software is not common in this particular research community. The interviewees acknowledged the lengthy process of reading, annotating and analyzing the textual data. For example, one interviewee said that he/she had to read all of a specific country's reports to judge the country's human rights violation situation based on the reading. Despite acknowledging the time required for these efforts, no interviewee used any QDA software (e.g., NVivo) to facilitate the process. In fact, many interviewees were not even familiar with QDA software. Instead, they used traditional and low-technical methods to analyze textual data, such as reading through PDFs, Word documents, or working almost exclusively with print documents. One interviewee explained that he/she used Word processing and spreadsheet software to store excerpts that had been extracted from the textual documents. Several interviewees created their own databases for storing and accessing their documents, and one interviewee created their own open source software program for researchers to share audio/video interviews within the research team.

In the questionnaire results, 69 of 74 respondents expressed that they used qualitative data analysis for their projects. Yet only 21 of these respondents (28.4%) acknowledged that they used software programs in analyzing their qualitative data. The researchers gave varied reasons for not using the qualitative analysis software programs. Majority of the respondents (31.4%) reported that were not aware such software existed, 11 of 51 people (21.6%) selected that they were aware that such software existed, but did not have the time or resources to learn the program, while 6 people (11.7%) recorded they preferred analyzing documents in print. In addition 18 of 51 people selected 'other' reasons for not using qualitative data analysis tools. These 'other' reasons given were that they were either unaware of its existence, saw no value in its use, or that their analysis required statistical software.

Eight researchers who used QDA software used NVivo, accounting for 47%. Two researchers used OpenEvsys and Martus Bulletin System (MBS) respectively; one used CDS/ISIS, Palantir, and Stories respectively, while eight researchers used other software programs. The other qualitative analysis programs listed by the researchers were ATLAS.ti, Excel's non-statistical functions, R, and a respondent noted that they developed their own qualitative analysis tool.

8.4.3. Data Analysis Processes – Collaborative or Individual

The interview study suggested that a collaborative or cooperative data analysis process is common. One reoccurring theme is that the data analysis process is staged from judging/filtering textual data to synthesizing and publishing the results, and multiple people are involved in different stages. Also, those who perform the initial reading/judging/filtering step are often undergraduate or graduate students as well as volunteers with certain background that meets the need of the research (e.g., cultural or language background).

Several interviewees used collaborative software programs such as Dropbox and Basecamp to aid the sharing and management of their data or documents for data analysis. One interviewee whose work was done mainly through ethnographic studies explained that he/she used Facebook to communicate with people from the studied countries and the analytical process was "a collaborative interpretation (process)". Another interviewee discussed how the online project management software Basecamp was used for collaborating their interviews of victims of human rights violations. In this project, there were multiple working groups performing these interviews and the interviewers' reflections were highly regarded. The interviewee said, "*Building a reflexology into methodology is so crucial, especially in this project based research*". As a result, every interviewer and videographer wrote a reflection blog within 24 hours of interviewing an individual. These reflection blogs were stored on the research team's private Basecamp site, and were shared among the research team (which included several hundred members in this particularly large project).

When asked "who analyzes the dataset in qualitative/non-statistical analysis of human rights violation data?" in the questionnaire, 35 of 73 respondents (48%) reported that they did the qualitative analysis of human rights violation data by themselves. 35 noted that they analyzed the data with others, and to illustrate who the "others" were in their projects, they listed colleague/peer/co-researcher, "team" (unspecified), research assistants, respondents, and organizational staff as exemplification of people that they analyzed their qualitative data with. Specifically, eight of thirty-five people (22.9%) reported colleagues, 7 people (20%) entered research assistants, 6 people (17.1%) noted research team members, and 4 of 35 (11.4%) noted co-authors.

8.5. Discussion

8.5.1. Primary Data or Secondary Data?

Interestingly, we learned from both the interviews and questionnaire results that a researcher's understanding of the difference between primary and secondary data could be obscure. We speculate that this is due to the wide variety of primary and secondary data that the researchers deal with in their practices. Working on an Inter-American course project, one interviewee examines advisory opinions and supervisory decisions by analyzing government archives and the archives of human rights organizations, in addition to the oral history life stories of human rights activists. We consider these data as primary data since they were firsthand records of advisory opinions and supervisory decisions. In a different project, the interviewee analyzed the Columbian newspapers published in the last fifty years in order to understand patterns of violence. These newspapers appear to be secondary data to us for the research focus, as they are not from direct observers of the violence. Nonetheless, the interviewee referred to all of the data he/she has worked with as primary data. Another interviewee analyzes the human rights reports from U.S. Department of State (http://www.state.gov/j/drl/rls/hrrpt/) which are secondary data to us, but when the interviewer asked, "*is it fair for say that your data are mainly secondary data?*", the interviewee said, "*I don't know… it's something I struggled with you know. …we do quite a bit of reliability testing to make sure that the coding guidelines are tight. See that's why I don't know if that's secondary…we're using secondary information, but the measure is very original…*" Two other interviewees who also do statistical modeling with human rights violation data worked with primary data such as court reports, surveillance cards, and victim interviews. To convert the textual primary data into numeric data for statistics, they also recruited people to read through these primary data and enter certain information to a database based on a form. The information is then converted to numeric scale, and they use the information for quantitative analysis and statistical modeling.

The unclear perception of the idea of primary and/or secondary data was also evident in the questionnaire data. In the examples of *primary* and *secondary* data that the respondents gave, similar examples were identified as both primary and secondary sources of data. For instance, the examples of primary data listed by the respondents include court cases, judgments, case reports, UN reports, case law judgments, interviews with people who have had their rights compromised, legal documents, victim and witness testimony, health records. And the listed secondary data are government reports, human rights NGO reports, Political Terror Scale and Cingranelli-Richards datasets, reports of other organizations, books, journal articles, newspapers, field reports, legislative review, incident reports, transcripts, legal literature, interviews with present and former clients, legal aid lawyers, judges, and court officials, existing data sets, demographic data, HR reports, legal documents, etc. As shown from these two lists, there is vagueness and inconsistency in what the researchers consider to be primary and secondary sources of data, and this is noteworthy.

Another interesting point related to this issue arose from the interview with a researcher who researches how countries treat poverty issues through studying constitutions. Despite that his/her main approach is quantitative methods with secondary data or public documentations, she/he commented, *"I do know that our biggest frustration is that getting data, getting information from those places…Really for people who work in poverty getting disaggregated data, very on the ground data, instead of just averages, is crucially important".*

8.5.2. Terms that Cross or Do Not Cross Disciplines

One thing we learned from this interview study is that it is important to consider whether certain terminologies cross or do not cross disciplines when doing research *about* the research practices, including data collection and analysis techniques, of other researchers. Some terms such as "coding" and "qualitative" may not cross disciplines because there are inconsistent understandings of a term's meaning. We found that although the interviewer and the interviewee were talking about the *coding process,* it actually could mean something different to the interviewee. In an interview with a political science professor,, *coding* meant judging the severity of a country's human rights violation situation based on the textual data, and giving a numeric code to rank the country (e.g., in the *coding* scale of 1 to 3, 1 could mean *no concern on human rights violation*, 2 *some concerns/issues,* and 3 *severe situation in the country*). We did not realize this different interpretation of *coding* until the interviewee showed us the *coding schema* and the related instruction guide. They are essentially evaluation criteria of different aspects of human rights violation and explanations of what each numeric code means. With this documentation the evaluators (the interviewee referred to them as *coders*) hopefully have shared understanding on the meaning of each numeric code and their results are comparable then (e.g., the rating of 3 should mean the same level of severity among the evaluators). Whereas in our interview with a history professor who analyzes news articles and interviews, *coding* meant open-theme reading and interpreting textual data and extracting segments as evidence to support the interpretation.

Also, although the qualitative analysis is similar to thematic analysis according to our knowledge of research methods, the term "thematic analysis" does not seem to cross disciplines well. One interviewee explained how he/she reads the course case documents and generate themes to interpret his/her readings but did not refer to this as thematic analysis – "*I take the leading principles from these cases, and I would then summarize where the law is going by reading from these documents…So it's an intuitive reasoning based on all my experiences…*" The different interpretation of coding between the interviewer and the interviewee was also captured in the same interview: at one point, the interviewer tried to make the connections between the qualitative analysis that she is familiar with and this interviewee's work, "*if I come up with this (the results), I will need to follow some content analysis, or coding analysis…*" and the interviewee responded, "*we don't do that*". Further discussion revealed that the interviewee was only considering coding in the quantitative sense. When the interviewer explained the term thematic analysis, the interviewee acknowledged that it was very similar to what he/she has practiced in the research.

Similarly, not every discipline refers to qualitative analysis by using the word “qualitative”, even though that is the kind of analysis they are performing. One interview showed the interviewee’s awareness of this terminology issue across disciplines. When we asked an interviewee if he/she has done any work with qualitative analysis approach, the interviewee started with a description of a paper co-authored with his/her colleagues, and then paused, “*I am not sure if this would count…Can you give me a definition of what you mean by qualitative?*” We responded by asking how his/her research group generated the previously mentioned database based on written documents – constitutions. It turned out that it was an example of qualitative analysis: he/she had a “*coding sheet*” that detailed what kind of provisions can be considered as enforceable laws and what cannot, and he/she and the graduate student would then read the constitutions and identify and judge the provisions.

8.5.3. Awareness of Data Availability, Data Pre-Processing, and Data Access

Because of the sensitivity of their type of research, we were curious if there were additional requirements besides the ethics protocol that the researchers needed to satisfy in order to access the data, and if so, what they were. From the respondents’ answers, we identified the following requirements: confidentiality agreements; registration – e.g. “*like CIRI dataset, I had to register and log in in order to have an access*”; general human-subject protocol, permissions to quote/limit of direct quotation; "*official approval*"; and legal adherence. However, many of the respondents noted that this question was not applicable or there were no specific requirements, such as “*no barriers to data access*”, and “*no, as I am using publicly available data, there are no further restrictions*”.

One interviewee who worked extensively in the process of obtaining primary human rights violation data commented that it was very important for the researchers to realize that the human rights violation data was often affected by hidden factors, like the witness’ availability and background, and whether the victim reports were filtered by the agency that took the victim’s statements. We were curious whether the researchers were aware of such processes, so we asked “Some pre-collected data have been processed before being available to researchers (e.g. certain data are stripped for privacy concerns). Are you aware of such data pre-processing protocols in your research? How has this impacted your research? Please explain”. A majority of the respondents (16 of 31) noted that they were not aware of any such preprocessing. While 12 respondents acknowledged their awareness of such processes, they expressed that this had no impact on their research. 13 of the 31 respondents said that they were aware and that it did have some impact on their research, for instance “*I am aware that this happens, but I have not been aware of what has been stripped from which data sets. This means I haven’t adjusted for these processes*”. And such processes could affect the selection of datasets and databases, as one respondent explains, “*It makes me a much more critical user of some data sets. For example, I would not use PTS or Freedom House as their processes are not reliable and have validity issues as well*”, and/or make it difficult to find certain information.

8.6. Toward the Development of Software Programs to Support Qualitative Data Analysis

Our study also contributes to the development of the software programs that support such processes. As Peters and Wester (2007) stated, in spite of the growing attention for qualitative analysis, there is a problematic link between procedures of qualitative analysis and software programs that support it. To help novice qualitative researchers understand what tool to use for which stage of analysis, Peters and Wester (2007) emphasized the importance of providing specific and detailed instructions to link computer tools to research methodology in order to help researchers better understand the tools and support their research. They also advocate that more effort should be put into illustrating how computer programs may support the methodological quality of procedures used in various approaches of qualitative data analysis.

Taking a user-centered development perspective, it is critical to elicit the design requirements of such software programs through requirements analysis. One important component that helps designers construct the design requirements is the understanding of the existing practices in qualitative data analysis and the important factors. There are studies that help us these aspects. For example, to understand the meanings attached to qualitative research practice and the perceived challenges posed by contemporary innovations in data management, access, and analysis through electronic archives, Broom, Cheshire and Emmison (2009) conducted six focus groups with 37 Australian qualitative researchers and revealed that the researcher has a special relationship with the data which prevents anyone else from analyzing them in their original context and which leaves the data 'disembodied' when archived. They also identified other important issues such as the concerns over research ethics and data ownership.

The literature in Qualitative Data Analysis (QDA) software has also identified important design requirements. For instance, one major concern is the potential loss of data, due to either technological obsolescence or technological failure. Cligget (2013) pointed out that with advances in technology, QDA software programs must have "robust exporting options for durable formats" (Cliggett 2013, p. 7). The author suggested that an important feature to have in QDA software is the ability to anonymize individuals and key identifiers while preserving the context and relationships, and the software must strive to preserve metadata in a variety of file formats, so future researchers using the same data can consistently "identify relationships between interviewees, images and spatial data, or regional relationships such as village residence and agricultural fields identified in maps, or topical themes addressed in different data types, such as interviews and field notes" (Cliggett, 2013, p. 8). This emphasis on assisting future qualitative human rights researchers is also highlighted by Wesley (2014).

Prior studies have identified important analytic tasks that occur in qualitative research. Gilbert (2014) identifies four broad steps, with more specific analytic tasks, that are essential to the qualitative data analysis process: organizing data, exploring data, interpreting and reflecting, and integrating data. Gilbert also highlights the work of Lewins and Silver (2007), who devised a list of required QDA software functionalities based on the following analytic tasks: project planning and management; analytic memo writing; reading, marking and commenting on data; searching data; developing coding schemes as well as the actual coding, retrieval of coded sections, and recoding; data organization, linking, and mapping; searching the data and codes; and reporting features (Lewins and Silver, p. 9).

In our view, it is also important to understand the researchers' experiences with the existing software programs in their qualitative data analysis, such as the features they disliked, the features that would have helped, etc. So we asked for the researchers' experiences with the existing QDA software, their expectations of a new QDA program, as well as the concerns regarding the software.

8.6.1. Experiences with the Existing QDA Software

The researchers that responded to our questionnaire used the software to map relationships between different human rights violation (12 of 18), to compare data from many different sources (10 of 18), to organize and share human rights violation data with other (9 of 18), to collect and input human rights violation data (8 of 18), to analyze complex datasets (7 of 18) and to securely store human rights violation data (7 of 18). Also, 5 of them explained that they used it because it saved time and/or costs associated with pre-processing data. The researchers also provided other reasons for using the qualitative software programs, e.g., it helped them explore new areas of rights and procedures, to analyze my interview data, to reduce biases in during interpretation, or to share their data and results with their project partners and for presentations to their donors.

We also asked our respondents about the features that were available in the qualitative data analysis software programs they had used. Only 16 researchers responded to this question. The features that at least 8 researchers noted were: the ability to search through textual data by keyword matching, the ability to search by concepts, the ability to generate reports about the analysis results, the ability to sort analysis results by different criteria such as date, author, and source, and the ability to allow user to annotate the documents without tampering the original data. Features like the ability to search through annotations, to share analysis results with other users, to visualize analysis results, to search by sentiments, to identify trends and of violations with algorithms, and to save search histories were noted by a few researchers. Although it is expected that identifying victims and/or perpetrators from the data would be highly desired by the researchers, only 1 respondent noted that the qualitative data analysis software program provided such feature. In addition to the features we provided in the multiple-choice question, the respondents noted two additional features: to map to a geographic location, and to map against other kinetic and non-kinetic factors.

8.6.2. Expectation of a New QDA Software Program

With regards to the expectations of qualitative data analysis software, the interviewees presented a wide range of characteristics that would be useful for their research purposes. Two interviewees would like a program that would recognize patterns and meaning in the text being analyzed. One suggested, "*create a more effective, searchable database with all of the documents held by the Inter-American Court, and again, that would prompt the user to look for certain types of patterns and certain types of information that could be organic, be changing. I wouldn't want it to just be a Google search of a bunch of documents*". When asked to define a "*pattern*," the interviewee provided an example of how the court's decisions often do not mention violence against women. For example, certain cases "*may in fact be cases of violence against women, or they may in fact contain within them a story about patterns of violence against women that are not being treated by the court as such, so the court is talking about this* *individual being disappeared or murdered*". Another interviewee was also interested in a program that would locate patterns, stating that such a program would "*allow me to find patterns that I wouldn't be able to find otherwise, and also quickly, which is great, and of large interview collections*".

Interestingly, one interviewee would like to see a program that not only assists in the analysis of data, but also the management of data, in order to assist in the collaborative efforts of qualitative data analysis. The interviewee, who supervises students analyzing U.S. Department of State Human Rights Reports, stated, "*one of the things we struggled with was how to have the students manage the massive amounts of data…and be able to find things quickly*." Other requested features included:

- offering multi-language support,
- reducing the workload of collecting disaggregated survey data,
- enabling transcription of the audio files, and
- enabling adding and searching detailed metadata

In the questionnaire, sixty-three people responded to the question concerning their expectations of a qualitative data analysis software program. Out of which they identified that it is important for the software program to have the ability to search through textual data by keyword matching (85.7%), to allow user to annotate the documents without tampering the original data (74.6%), to search by concepts (73%), to generate reports about the analysis results (66.7%), to visualize analysis results (66.7%), to sort analysis results by different criteria such as date, author, and source (66.7%), to search through annotations (61.9%), to identify trends and patterns of violations with algorithms (60.3%), to share analysis results (57.1%), to refine analysis results (57.1%), to share documents with other users (53.9%), to save search histories (49.2%), to relate self-generated results (46%), to search by sentiments (30.1%), and to identify victims and/or perpetrators with algorithms (12.7%).

From our questionnaire results, we see that there is a correspondence with the features of the qualitative data analysis tools that the researchers reported to be available in the programs that they have used, and the features that they think should be included in qualitative analysis programs. For instance, it appears that the ability to identify human rights violation victims was noted by only one researcher as a feature in the existing analysis software, and as an identified expected feature by the respondents it also had the smallest percentage of the responses compared to the other expected features. Another example is the ability to search through textual data. It was well recognized both as a function provided by the existing software, and as a feature that should be included in qualitative software programs. In addition, the ability for the software to be able to generate reports and search by concepts seem to be quite important too, as more than half of the respondents reported that they have used this feature, and noted that this was a feature that should be included in the new software. This is in accord with the three reasons with the highest frequencies given by the respondents for using a software program for qualitative data analysis - to map relationships between different human rights violation, to compare data from many different sources, and to organize and share human rights violation data with other individuals. The important features identified by the researchers should enable them carry out these tasks.

There are features that a majority of the respondents expected from the new software but were not noted in the existing programs, suggesting that such features are additional user requirements that have not been met in the existing programs. Examples of such features are: to search through annotations, to identify trends and patterns of violations with algorithms, to visualize analysis results, and to share the analysis results.

8.6.3. Concerns regarding the QDA software

In the questionnaire, we asked, "What are the concerns or issues that you think should be addressed in qualitative/non-statistical analysis software?" This was a multiple-choice item with eight options provided plus an "other" choice to allow for textual input. Our questionnaire results reveal that the respondents have the concerns about using the qualitative analysis software programs in these aspects (N = 62): security of the data (75.8%), that the tool might not properly interpret the nuances/context of non-statistical data (58%), data security during the sharing (51.6%), the navigational complexity and data storage (46.8%), the inflexibility of technology in terms of importing and exporting data files (38.7%). In addition, 17 of 62 people (27.4%) had concerns related to the user access (different levels of access for different, and the data storage (after completion of the project). The respondents also noted some other concerns including the ability to share their data with the people they are researching, the cost of software, ownership, privacy issues, reduction of investigator bias, reliability of analysis, replication, and internal as well as external validity.

8.7. **Conclusion**

As the importance of reporting human rights violations by governmental and non-governmental organizations rises, there is a quick emergence of primary and secondary sources with rich textual data to offer additional details about human rights violation events. The growing interest of preserving human rights data and making them publicly accessible has also increased the access to human rights violation data. While these trends in the generation and access of human rights violation data offer more data analysis opportunities to human rights researchers, the increasing amount of data also presents a challenge in the data analysis process. This challenge is particularly severe in qualitative data analysis, as the researchers often need to identify, extract, and interpret meaningful or relevant information from the datasets. Conversely, the advancement of computational technologies for identifying and visualizing patterns from large amounts of textual data has offered a new direction for addressing the above challenge; however, it demands more than just the computational algorithms to develop user friendly tools that support the human rights researchers' data analysis processes. A thorough understanding of the researchers' data analysis and management practices is required.

Recognizing that there is lack of such understanding in the literature, we interviewed human rights researchers of different disciplines and research areas to understand their data analysis and management practices. We also surveyed the tools that researchers have used or were aware of for facilitating the qualitative analysis process. Our results provide insights about the human rights research data – data sources, types of data, and format of the data, and data analysis and management practices – data access issues, analysis methods and processes, and the use of software programs to facilitate data analysis. We also probed the researchers' experiences with the existing software for qualitative data analysis, their expectations and concerns regarding such a software program.

8.8. Acknowledgement

We thank all the interviewees and questionnaire respondents for their time in this study.

9. SPATIO-TEMPORAL DATA VISUALIZATION WITH STORYGRAPH

9.1. Storygraph

One of the major goals of spatio-temporal data analysis is to discover the relationships and patterns among scattered spatio-temporal events. However, it is difficult to construct a visualization that properly integrates spatial, temporal, and other data dimensions. For this project, we developed an interactive 2D visualization technique called Storygraph to address this problem. A Storygraph is composed of two parallel vertical axes representing geographic coordinates like a latitude-longitude pair and a horizontal axis representing a timeline. Each spatial coordinate or location is represented as a location line connecting its latitude and longitude on the two vertical axes on the Storygraph. Events occurring at that location are plotted as markers on the location line. Other data dimensions beside space and time are represented by different shapes, sizes, and colors of the markers. For this work, we assume that each event has a location, time, and characters with the first two being mandatory. For a dataset with one or more characters, a storyline is a line chart that connects the events associated with a character. Storylines show the movement of characters across different locations over time and how different characters cross path with each other and meet at a particular event. An interactive, linked 2D map view allows users to correlate location lines on the chart with location points on a map. This technique allows for readers to see the changing significance over time of a given location and the movements of people in the geographic and temporal context of a document-event space. A document-event space is the temporal and spatial limits of an event as described by one or more documents.

9.1.1. Model

Storygraph is a 2D diagram consisting of two parallel vertical axes $V_{latitude} \subset \Re$ and $V_{longitude} \subset \Re$ and an orthogonal horizontal axis $T \subset \Re$. All three of the axes, as in Cartesian graphs, are unbounded at both ends. The values in the axes are ordered in ascending order: from left to right in horizontal axis and bottom to top in vertical axes. The vertical axes $V_{latitude}$ and $V_{longitude}$ represent geographic *latitude* and *longitude* or Northing and Easting on a map. They can also be generalized to represent x and y coordinates of a point in a Euclidean plane. The horizontal axis, T, represents time. Thus a point plotted on Storygraph, which shall be referred to as *event* in the rest of the proposal will have at least three dimensions: latitude, longitude, and timestamp.

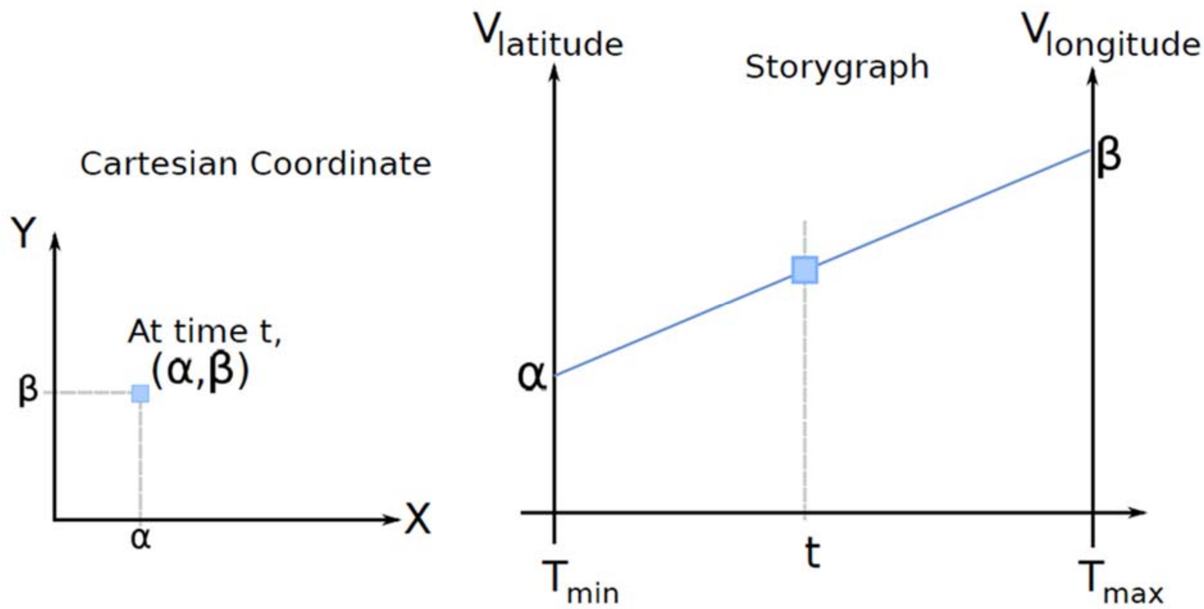

Figure 3: Left: A point in the Cartesian coordinate system such as a location on a map at time t, Right: Same point represented in a Storygraph.

For any event occurring at (α, β) in time t as shown in Figure 3, our algorithm first draws a *location line* by connecting the points on the two axes, $\alpha \in V_{latitude}$ and $\beta \in V_{longitude}$. The algorithm then returns the point on this line at time t.

The function $f(\alpha, \beta, t) \rightarrow (x_{storygraph}, y_{storygraph})$ which maps an event to the 2D Storygraph plane can be formally written as follows:

Equation 1

$$y_{storygraph} = \frac{(\beta - \alpha)(x_{storygraph} - T_{min})}{T_{max} - T_{min}} + \alpha$$

Equation 2

$$x_{storygraph} = t$$

where T_{min} and T_{max} are the maximum and minimum timestamps within the dataset.

Figure 3 illustrates how a location on a regular 2D map, coded with a Cartesian coordinate, is presented in the Storygraph. Equation 1 and 2 is used to convert a location on a regular map to a Storygraph plane, and vice versa.

Assuming $T_{min} = 0$ and $T_{max} = T$ without loss of generality, Equation 1 simplifies to, Equation 3

$$y_{storygraph} = \frac{(\beta - \alpha)(x_{storygraph})}{T} + \alpha$$

Lemma 1 *A right rectangular region at time t on a Euclidean plane can be represented by a vertical line segment on the Storygraph at t.*

Proof Given a right rectangular region bounded by four corners ($\alpha1, \beta1$), ($\alpha1, \beta2$), ($\alpha2, \beta2$) and ($\alpha2, \beta1$) at any time, *t*, from (Figure 4) we can see that since all the events occur at time *t*, result in a vertical line segment on the Storygraph. In addition, since $\forall\alpha$: $\alpha1 \leq \alpha \leq \alpha2$ and $\forall\,\beta$: $\beta1 \leq \beta \leq \beta2$, this line segment consists of all the vertices.

Figure 4 illustrates Lemma 1.

Theorem 1 *An enclosed area of arbitrary shape at time t on a Euclidean plane can be represented by a vertical line segment on a Storygraph at t.*

Proof Construct a right rectangle bounding the shape. From Lemma 1, we get a vertical line on the Storygraph.

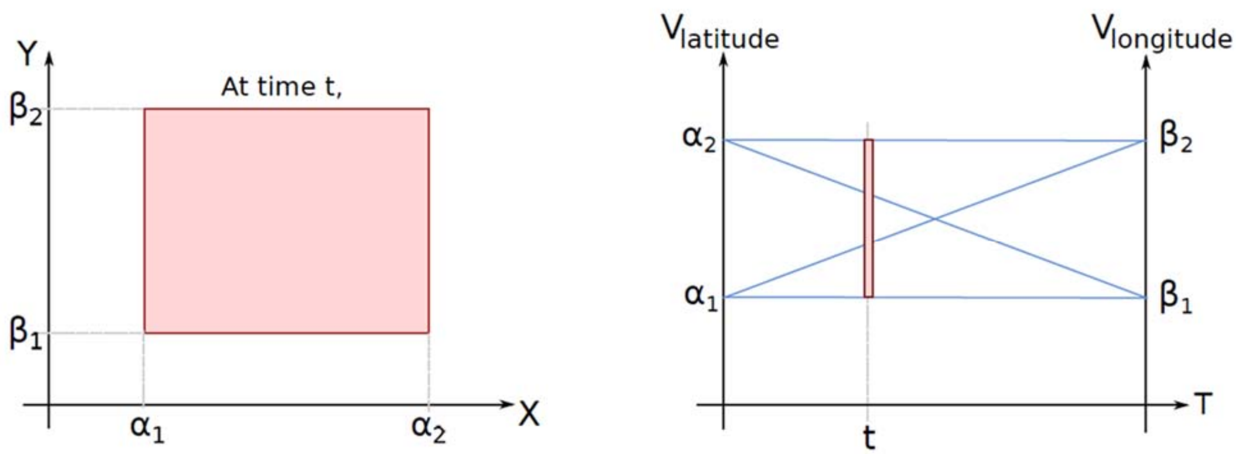

Figure 4: Left: An area filled with events in the Cartesian coordinate at time *t*, Right: Same area represented in a Storygraph.

9.1.2. Location Lines

Location lines are the line segments connecting the two vertical axes in the Storygraph. Each location line represents one real world location or a geographic coordinate and vice versa. However, in some cases, location lines might result in over-plotting. To avoid this issue, the implementation of storygraph can be made interactive so that users can hide/show location lines.

In absence of locations lines, the mapping function f is not one-to-one. Thus, the Storygraph has the following properties:

Lemma 2 *Without location lines, a point on a Storygraph at time t corresponds to a line segment on a map.*

Proof We can rewrite Equation 3 as

Equation 4

$$\beta = (1 - \frac{T}{x})\alpha - \frac{yT}{x}$$

Thus a fixed point (x, y) on the Storygraph corresponds to many points (α, β) on the Cartesian map at time $t = x$: those $\alpha_{min} \leq \alpha \leq \alpha_{max}$ and $\beta_{min} \leq \beta \leq \beta_{max}$ satisfying Equation 4. Plotting these values of (α, β) results in a line segment with non-positive slope since $x \leq T$ as illustrated in Figure 5.

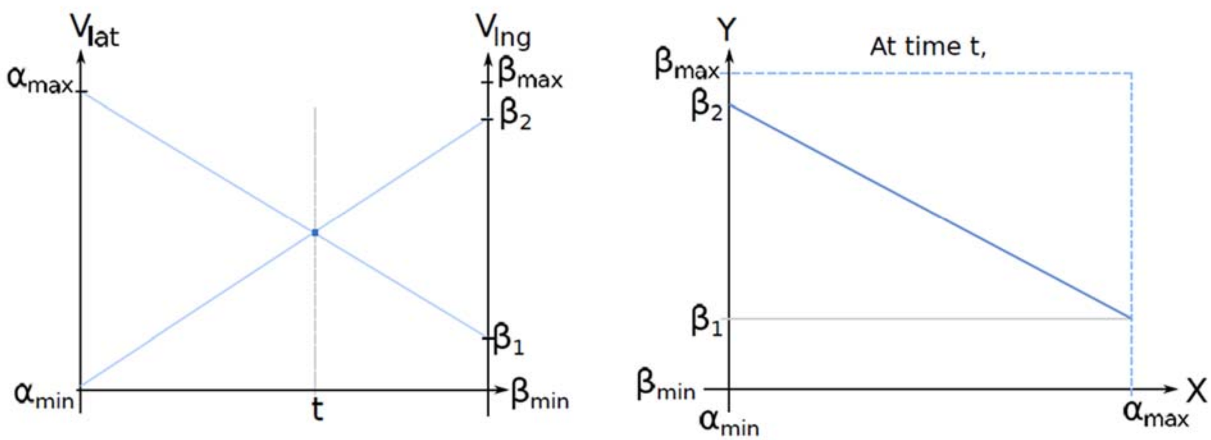

Figure 5: Left: A point in the Storygraph at time t and the corresponding location lines the point can belong to shaded, Right: The line segment generated in the Cartesian coordinate by mapping the point.

Lemma 3 *Without location lines, a vertical line segment at time t on a Storygraph corresponds to an area on a map.*

Proof Consider a vertical line segment, with end coordinates (x, y_1) and (x, y_2), y1 ≤ y2. Using Lemma 3, these extremes of the line segment in Figure 6 we get two straight line equations

Equation 5

$$\beta = (1 - \frac{T}{x})\alpha - \frac{y_1 T}{x}$$

Equation 6

$$\beta = (1 - \frac{T}{x})\alpha - \frac{y_2 T}{x}$$

Hence the vertical line segment between (x, y1) and (x, y2) on the Storygraph corresponds to an area between two parallel lines from Equation 5 to Equation 6 on the Cartesian map. As in Lemma 2, this area is also bounded by the maximum and minimum values of α and β – this results in a polygon as illustrated in Figure 6.

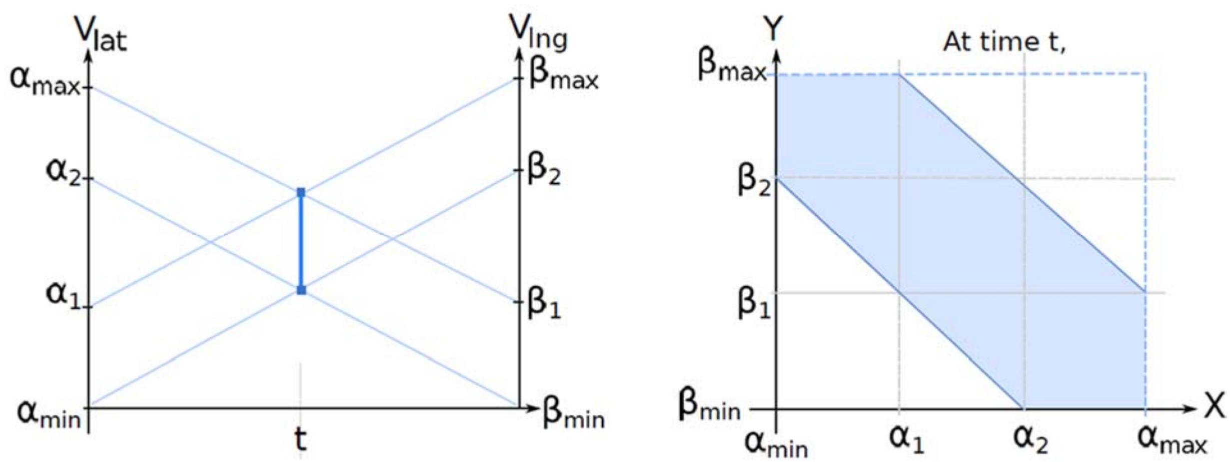

Figure 6: Left: A vertical line in Storygraph at time *t* and the corresponding location lines the line segment can belong to shaded, Right: The bounded region generated in the Cartesian coordinate by mapping the line segment.

Lemma 4 *Without location lines, a vertical line segment at t on the Storygraph corresponds to a projected area,* $A_{storygraph} \geq A_{actual}$*, the actual area on a Euclidean plane at t.*

Proof If the area on the plane is bounded by right rectangle, since $\forall \alpha : \alpha 1 \leq \alpha \leq \alpha 2$ and $\forall \beta : \beta 1 \leq \beta \leq \beta 2$, $A_{storygraph} = A_{actual}$. For any other shape, the vertical line segment in the Storygraph represents a rectangular bounding box (from Theorem 1). Thus, $\exists\ \alpha : \alpha \in A_{storygraph} - A_{actual}$. Hence, $A_{storygraph} > A_{actual}$.

Note: The point of intersection of the location lines does not have any meaning or significance. However, the crossing of the lines denotes that the two locations are diagonally apart. In addition especially in cases of interactive Storygraph, the two events may or may not be close. Closeness is a relative measure and it is up to the users of the Storygraph to decide when to call two locations close or far apart. Figure 7 shows two figures to illustrate this fact. In the left figure, two locations appear to be close in the Storygraph but their location lines are far apart relative to the figure on the right.

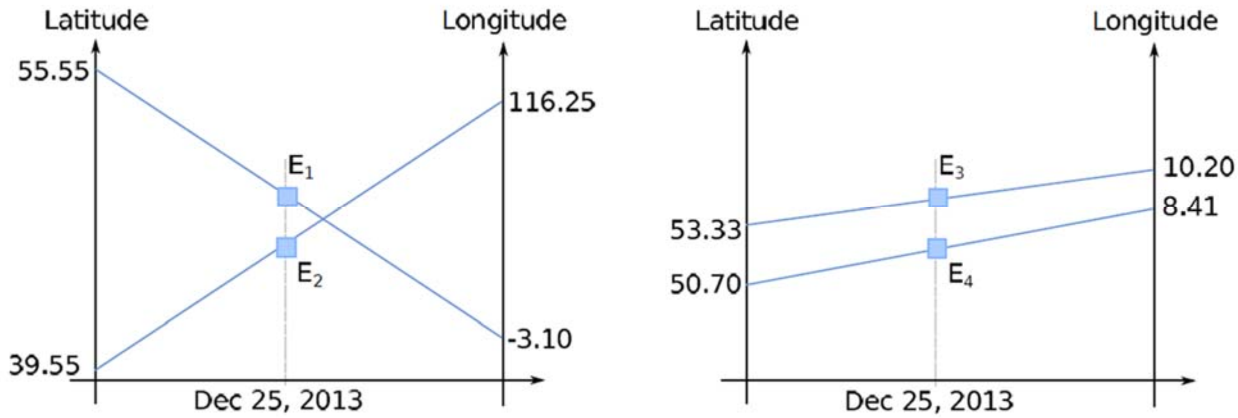

Figure 7: Left: Two events seemingly close in the Storygraph but far apart due to their location lines, Right: Two events close to each other in the Storygraph as well as geographically.

The closeness of the events depends on location lines and closeness of location lines in turn depend upon the scale of the graph. If the latitude axis in the right figure was stretched so that the max and min values in the axis were 53.33 and 50.70, then the generated location lines will have different slopes—one positive and one negative. Moreover, it would seem like the same two events occurred in far apart locations. Hence, it is often important to note the scales in all the axes during interpretation.

9.1.3. Storyline

Storyline is a series of events associated with a person or a group of people. In the Storygraph and rest of the text, people associated with event are called *actors*. Hence, drawing storylines require datasets having actors in addition to latitude, longitude, and timestamps. Storylines are drawn on Storygraph by connecting all the events associated with an actor sequentially resulting in a polyline. If there is more than one actor, a storyline is created for each of these actors as shown in Figure 8.

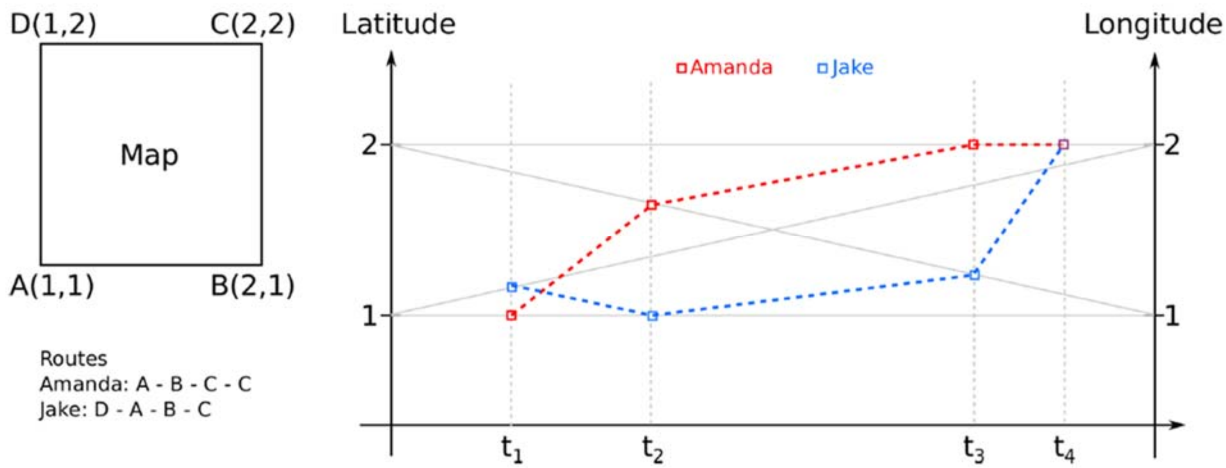

Figure 8: Storylines of two characters Jack and Amanda plotted on Storygraph. Amanda travelling from A to C and staying at C, Jake travelling from D to C via A and B.

Figure 8 shows the storylines of two fictional actors *Jake* and *Amanda*. It also shows a map on the left side with locations *A*, *B*, *C* and *D*. Amanda starts from point *A* at t_1 and then reaches *B* at t_2, *C* at t_3 and stays at *C*. Jake starts from *D* and then reaches *C* via *A* and *B*. The storylines plotted on the Storygraph shows the same information. From this figure, users can not only trace the pattern of the movement but also instances where two or more characters might have met: *C,* in this case. Hence, it allows users to visualize the relationship and interactions between the actors, which are often difficult to extract from the textual description. Storylines are especially helpful in visualizing movement as in Demsar and Virrantaus (2010), Murray et al. (2012), and Orellana et al. (2012) because a user can see how different actors meet at certain events and then move on to different directions.

Storylines can be compared to well-established space-time paths from timegeography (Lenntorp and Hort, 1976; Kraak, 2003; Thrift, 1977). A direct side-by-side comparison of storylines to space-time paths is shown in Figure 9. The figure shows three co-location phenomenons namely: co-location in time, co-location in space, and co-locations in time and space. Co-location in time means that two actors might have been in the different places but both of them were at the same place for a given period of time. Co-location in space means that both actors were at the same spot but during different times. Co-location in time and space means that both actors were at the same location at the same time period.

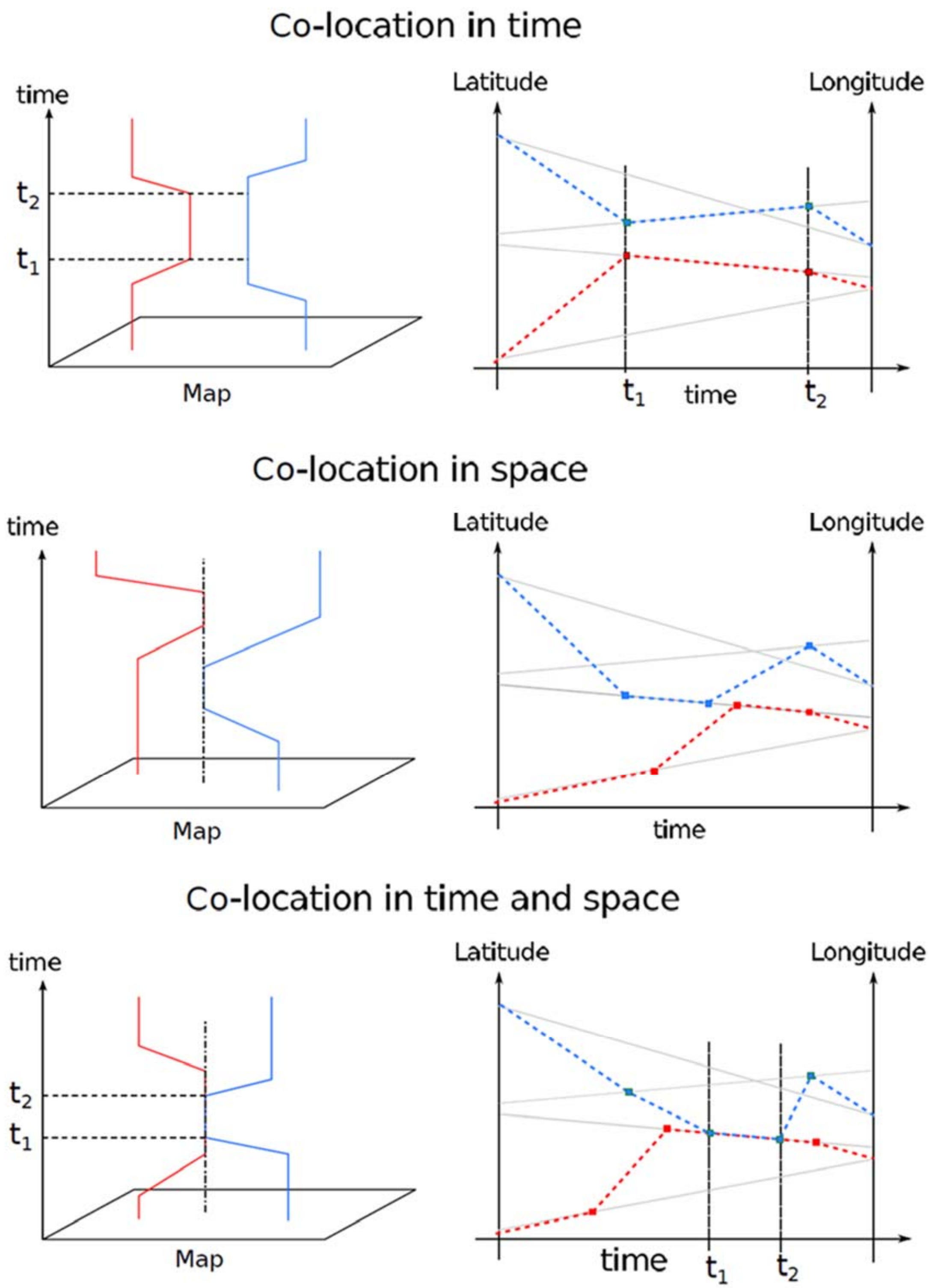

Figure 9: Side-by-side comparison of storylines to space-time paths. All the left figures show different concepts in time geography using space time paths (adapted from Shaw and Yu (2009)). All the figure in right show the information using storylines in Storygraph.

9.2. Extending Storygraph: Visualizing Spatio-Temporal Uncertainty

In the previous sub-sections, we introduced the model of Storygraph and demonstrated the benefits of using it on datasets containing precise geolocations and time. However, when applying this method to spatio-temporal data extracted from witness testimonies and field reports, we encountered problems of uncertainty in space and time. For example, our study of 511 interviews with first responders during the attack on World Trade Center (WTC) on September 11, 2001, showed that the narratives of these interviewees, who were trained to report incidences, still contained a fair amount of uncertainty in their descriptions of locations and times.

To address these issues, we extended Storygraph to visualize uncertainty. We first begin by categorizing uncertainty into two categories: (1) event uncertainty and (2) between-event uncertainty. We designed our method to distinguish and visualize these two types of uncertainty. Event uncertainty is the spatio-temporal uncertainty about the event itself, including events with poorly specified spatial and/or temporal attributes. Between-event uncertainty is the uncertainty between two precisely recorded events, which we call *key events*. This concept is in part influenced by Hägerstrand's Time Geography (Hägerstrand, 1989; Thrift, 1977; Miller, 2005). After specifying the key events, the between-event uncertainties are visualized as space-time prisms between the key events. Through this process, our visualization technique can be used to study the interactions between people (or characters) in both space and time.

9.2.1. Classification of Uncertainty

Different classifications of uncertainty have been proposed (Potter, Rosen, and Johnson, 2012; Abul, Bonchi, and Nanni, 2008); however, most of these classifications are about uncertainties introduced in scientific experiments or probabilistic models. In our case, uncertainties are introduced in narratives. Thus, we classify this kind of uncertainty into three categories:

- **Uncertainty about time and/or location of the event**. This type of uncertainty is characterized by the presence of phrases denoting uncertainty before temporal or spatial description. An example is "I got there maybe around 11 am." The phrase 'maybe around' adds uncertainty to '11 am' in this example. Such uncertainties may also arise from ambiguity in language. For example, in "I was in Brooklyn when the plane hit the building," the word 'Brooklyn' does not give a precise location. We call these types of uncertainties as *event uncertainty* which can be further divided into three sub-categories:
    - *Spatial uncertainty*. This category includes events that have precise time stamps but uncertain location.
    - *Temporal uncertainty.* In addition to uncertain phrases (e.g. maybe, about), temporal uncertainty may come from the language itself. For example, in "I was at the station in the all day," the phrase "all day" without any modifier can refer to a wide range of time introducing uncertainty.
    - *Spatio-temporal uncertainty*. This category includes events that have uncertainty in both time and location. For example, in "It was in the afternoon, I was heading south." The words 'afternoon' and 'south' are uncertain.
- **Uncertainty between two events**. In "It was 8 in the morning I was at home. As soon as I heard about it, I reached the site at 10", the first event ("at home") and the second event

("reached the site") are both certain. However, what happened between the two events is unknown. We call this type of uncertainty between-even uncertainty

- **Uncertainty about the even taking place**. In the WTC corpus, we often encounter sentences like "I think Chief pulled me back." The word 'think' indicates an uncertainty about whether the event has ever happened. Detecting this type of uncertainty is difficult and beyond the scope of this paper. Instead, we focus only on visualizing event uncertainty and between-event uncertainty.

9.2.2. Event Uncertainty

Figure 10: Three kinds of glyphs used to represent spatial, temporal, and spatio-temporal uncertainty. Left: Dashed I-beam is used to represent spatial uncertainty. The slope of the top and bottom of the beam disambiguates the range of locations in the geographical space. Middle: parallel lines are used to denote the temporal uncertainty. Right: Box showing spatio-temporal uncertainty. The slope of the edges of the box maps to a fixed geographical area within a certain time.

In this sub-section, we discuss the extraction and visualization of event uncertainty. To extract event uncertainty, we compiled a list of English words that may indicate location uncertainty such as "around," "near," "close to," "maybe," "perhaps," etc. We then gave each word an uncertainty score in the range of $1 - 100$ (Erman et al, 1980; Marneffe, Manning, and Potts, 2011); Sauri and Pustejovsky, 2009). The same process was repeated for temporal information. We extracted the named entities from WTC corpus using Stanford NER (Finkle, Grenager, and Manning, 2005) and time using SuTime (Chang and Manning, 2012). TARSQI (Verhagen et al, 2005) was used to extract the temporal sequence of the events, and locations were geocoded using Google Maps API. The results were then verified and corrected.

In the WTC corpus, we observed all three types of event uncertainties: spatial, temporal, and spatio-temporal. Some key events with precise spatio-temporal information were used as anchor events. These include the first and second plane hitting the tower, and the plane crashing into the pentagon. These events were chosen as key events because all of the interviews described more local events in reference to these global events. Examples include "When the second plane hit the tower, I was running towards Vesey," and "I was at the station when the news about the first explosion was on TV." When considering these key events in the context of the first example, the time is certain but the location is uncertain. Additionally, in a sentence that references no key events like "When the EMS arrived at the scene, I began heading south", both location and time would be considered uncertain.

For each event, latitude, longitude, date/time, color, spatial uncertainty, and temporal uncertainty were fed to the visualization program, which then visualized the uncertainty information along with other information.

- **Spatial Uncertainty**. Spatial uncertainty is visualized as a vertical dashed I-beam. From Section 9.1, we know that an area on a map corresponds to a line in Storygraph. The length of the I-beam is proportional to the area of possible locations. More importantly, the top and the bottom of the beam disambiguate the range of locations in geographical space. This representation is shown by the left sub-figure in Figure 10.
- **Temporal Uncertainty**. We use sloped double lines to represent temporal uncertainty. Each double line is drawn along the location line for the corresponding event, which can be seen in the middle sub-figure in Figure 10. A double line indicates that the event happens at a particular location within a certain time frame. In contrast, a single solid line along the location line means that the character stayed at the specified location for a period of time. Through these representations, the two cases are visually distinct.
- **Spatio-temporal Uncertainty**. We use a semi-transparent box to visualize spatio-temporal uncertainty, which means both location and time are uncertain. The sloped top and bottom sides of the box indicate the range of locations while the vertical sides of the boxes shows the temporal bound. The box is drawn as semi-transparent to prevent glyph occlusion. This representation is shown by the right sub-figure in Figure 10.

Figure 15 shows this concept applied to the events extracted from WTC corpus.

9.2.3. Between-event uncertainty

The purpose of visualizing between-event uncertainty is to display the space-time constraints between two key events. Any activity takes place within a certain span of time and a certain geographical region. Individuals participating in these activities have to trade time for space or vice versa. For example, during a workday lunch hour a person could walk to a nearby restaurant for a longer meal or drive to distant restaurant for a shorter meal. Visualizing between-event uncertainty can assist planning, scheduling, and analyzing possible overlapping in people's activities.

Our between-event visualization technique is partially based on Hägerstrand's Time Geography, a conceptual framework that focuses on constraints and trade-offs in the allocation of time among activities in space (Miller, 2005). However, Time Geography is a map-based 3D visualization. Therefore, it suffers from the typical problems associated with 3D visualizations such as 3D occlusion and difficulty of navigation. Besides, space and time are not well integrated in Time Geography. Our work is an attempt to address these issues.

9.2.3.1. Space-time paths and space-time prisms

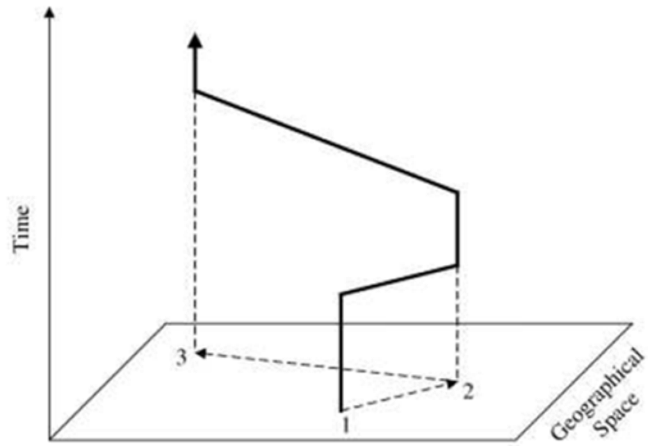

Figure 11: An example of space time path adapted from Miller (2005). Space time paths trace the movement of an individual moving from one location to another. Space-time paths also show the amount of time spent at a location by the individual before moving to the next location.

We adapted two important concepts from Time Geography: space-time paths and space-time prisms. Space-time path traces the movement of a character in space and time. Figure 11 shows an example of a space time path adapted from Miller (2005). The base plane is the geographical space and the orthogonal axis is time. In this example, an individual travels from location 1 to 2, spends some time at 2 and then moves on to 3. The time and location of the starting or end point are known as *control points* or *key events*. The straight line segments connecting two control points are known as path segments. Path segments are represented by straight line segments for simplicity (Pfoser and Jensen, 1999; Moreira, Riberio, and Saglio, 1999). In our earlier work, we adapted the concept of space-time paths in Storygraph using storylines (Shrestha et al., 2013). Here, Storylines become space-time paths, connecting two consecutive key events via dotted line segment.

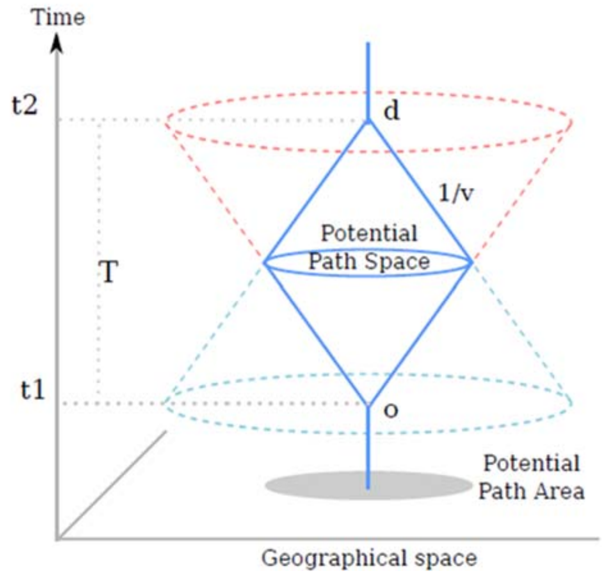

Figure 12: Space-time prism. In this figure, the individual is at location *o* at *t1* and needs to be at the same location *d* at *t2* (*o* origin of travel). (S)he has the time budget of *T*. The red dotted cone shows the possible path space starting from *o* with the maximum velocity, *v*. Similarly the blue dotted cone shows the path space towards *d*. The intersection of these two cones gives the potential path space under the given time budget *T*. The potential path area is shown by the gray area on the geographic space.

Space-time prisms extend space time paths to create a 3D space consisting of all the possible routes an individual can take while moving from one point to another. This space is known as the *potential path space*. The prism between *t*1 and *t*2 in Figure 12 demonstrates this concept. The slope of the edges of this prism is determined by the inverse of maximum velocity. That is, the possible paths are constrained by the maximum velocity of the individual, a fixed time frame, and fixed destinations. In our implementation, the maximum velocity is set by the user.

If an individual is at origin, *o*, at t_o and needs to reach destination, *d*, at t_d, the time budget is $T = t_d - t_o$. The path space from the origin under the time budget is shown by the red inverted dotted cone. This space shows all the possible paths and all the possible locations that can be reached within the time budget with maximum velocity *v*. Let this region be denoted by $R_o(T)$. Similarly, the blue dotted cone shows the path space towards *d* under the time budget. This 3D space gives all the locations from where *d* can be reached under time *T*. Let this region be $R_d(T)$. The intersection of these cones gives the potential path space for individual traveling from *o* to *d* (Neutens et al, 2008). Hence,

Equation 7

$$R_{od}(T) = R_o(T) \cap R_d(T)$$

The projection of the space time prism on the geographical space, as shown by a gray circle in the figure, shows all the possible locations that the user can reach. This area is called the *potential path area*.

Given all the control points within a specific time window, τ , the construction of space-time prism requires the destination d to lie within the $R_o(T)$ and vice versa. Stating it formally,

Equation 8

$$\forall o, d \in \phi_\tau : (o \cap R_d(T) \neg \emptyset) \wedge (d \cap R_o(T) \neg \emptyset)$$

In Time Geography, space-time paths and space-time prisms are generally drawn inside a 3D space-time cube (Kraak, 2003) (Figure 12). In our work, space-time paths and space-time prisms are drawn on Storygraph in a 2D view.

9.2.3.2. Visualizing between-event uncertainty

Storygraph draws space-time prisms based on Equations 7 and 8. From Section 9.1, we know that an area in the geographical space is mapped to a line in Storygraph. Thus, starting from a location, $o(\alpha, \beta)$, at $t0$ and taking a snapshots of the potential path area at each time step, we get a set of areas sequentially increasing at the rate of the velocity. The top sub-figure in Figure 13 shows an individual at point (α, β) at $t0$ and his/her possible path area after each time step $t1-t5$. The figure simplifies the drawing of the potential path areas by representing them with squares rather than circles. The bounding of the actual potential path by squares introduces some uncertainty itself (Brodlie, Osorio, and Lopes, 2012) but greatly simplifies the drawing and the calculations.

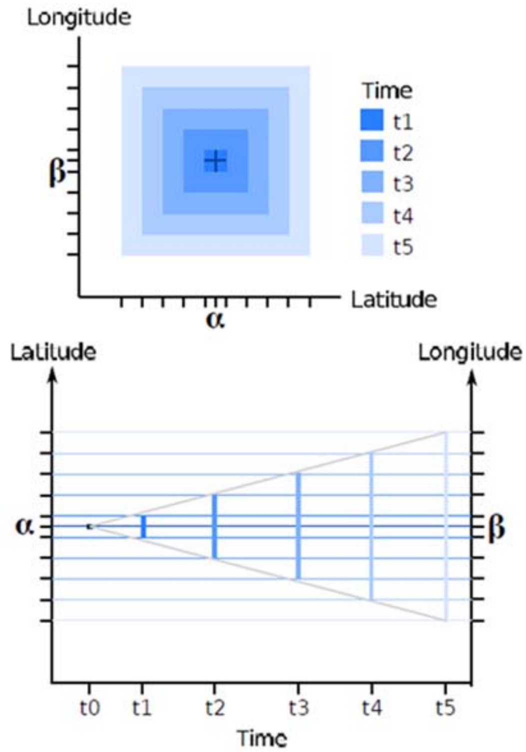

Figure 13: Above: Starting from the origin of travel, α, β, at $t0$, the potential path areas in each time step $t1 - t5$ (assuming a certain velocity and regular time intervals). Below: The same data plotted in Storygraph. Each potential path area is mapped to a line segments in Storygraph. For continuous time, this representation would result in an area enveloped by two gray lines. The slope of the gray lines is equal to the maximum velocity.

Hence, if the time step, $\Delta t \to 0$, then conical region $R_o(T)$ would be reduced to a triangular region in Storygraph. This region is shown by the area enveloped by two gray lines in the bottom sub-figure in Figure 13.

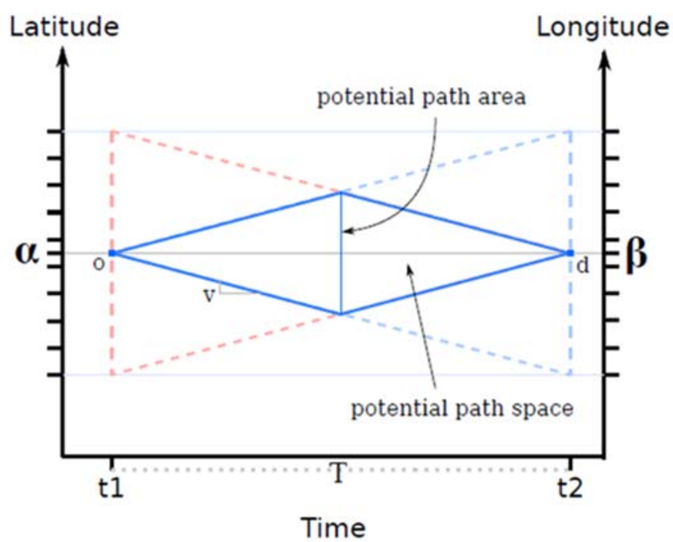

Figure 14: The space-time prism shown in Figure 12 drawn in Storygraph.

Figure 14 shows the result of mapping the space-time prism in Figure 12 in Storygraph. The mapping process resembles the drawing of the space-time prism inside the space-time cube. Given two control points, maximum velocity, and a time budget, these parameters are plugged into Equation 8 to check whether the control points satisfy these criteria. If the criteria are satisfied, we compute the extents (lat_{max}, lng_{max}) and (lat_{min}, lng_{min}) of the $R_o(T)$ with the following sets of equations,

Equation 9

$$lat_{max} = max_{lat_r}[\sqrt{(lat_r - lat_o)^2 + (lng_r - lng_o)^2} = vT]$$

Equation 10

$$lng_{max} = max_{lng_r}[\sqrt{(lat_r - lat_o)^2 + (lng_r - lng_o)^2} = vT]$$

Equation 11

$$lat_{min} = min_{lat_r}[\sqrt{(lat_r - lat_o)^2 + (lng_r - lng_o)^2} = vT]$$

Equation 12

$$lng_{min} = min_{lng_r}[\sqrt{(lat_r - lat_o)^2 + (lng_r - lng_o)^2} = vT]$$

Similarly, the extent for the $R_d(T)$ is calculated. Finally, $R_{od}(T)$ is obtained from the intersection of these regions.

9.2.3.3. Intersections of prisms in Storygraph

Space-time paths and prims are both based on the movement data of characters. Given a dataset containing the movement data of two or more individuals, it is likely that the space-time prisms will overlap. However, since Storygraph does not preserve event proximity (Section 9.1), it is important to note that these overlaps may not necessarily mean that these prisms intersect in geographical space.

Hence, given a point p and a prism $R_a(T)$ in the Storygraph, we first establish the conditions for a valid point-prism intersection. Building on this, we then present the validity of intersection between two prisms.

Let $R_a(T) : R_a(T) = R_o(T) \cup R_d(T)$, be all the possible locations that the individual can travel within the time budget T with a maximum velocity v. Then, the following cases for point-prism intersection could arise:

1. The point is not inside the prism, but the location line is inside $R_a(T)$. This case implies that the event occurred within the geographical bounds but the individual may not have been involved in the event due to the travel constraints.
2. The point is inside the prism, but the location line is not inside $R_a(T)$. This case implies that the event occurred within the time span, T, but at some other location, $\notin R_a(T)$.
3. The point is inside the prism, and the location line is inside $R_a(T)$. This case is the only case where the individual could have been involved in the event.

Theorem 2 *For a valid point-prism intersection, the point should be inside the prism and the location line should lie inside $R_a(T)$.*

Proof Assume that this is not a valid intersection. It means that the point representing the event is either spatially or temporally incorrect. It is temporally incorrect because for a point to lie inside the prism, it has to occur within the time budget. It is spatially incorrect since $R_a(T)$ defines the maximum distance an individual can travel at within a time T.

Hence, given two prisms, $P1$ and $P2$, the prism-prism intersection is only valid if there exists a point p on location line l such that $l \in R_a^{P1}(T) \wedge 1 \in R_a^{P2}(T) \wedge p \in P1 \cap P2$.

9.2.3.4. Case Study: WTC 9/11

In the immediate aftermath of the attacks in New York on September 11, 2001, the NYC Fire Department convened a task force to interview first responders to the affected areas. These 511 interviews, conducted in the two months following the attacks, were later released by *The New York Times*. Each interview was conducted by staff from the New York Fire Department assigned to the task force and ran anywhere from 8-20 minutes with the aim to elicit from first responders their activities on September 11. The language of the reports is typical of event interviews and oral histories. Despite having a population with high area knowledge and normalized reporting practices, locations and times were predominately referred to referentially. Known individuals seen by the interviewee are named, but most are either not named or referred to solely by rank. The primary reason to visualize this data is to enable historians and investigators to identify accurate and inaccurate information and to allow for recognition of corroborating evidence more readily. When viewed as a corpus rather than separate interviews, it becomes possible to identify overlaps in the reported events of the witnesses. The challenge posed to this task by the referential language usage of the witnesses is pervasive in oral history and other investigatory work reliant on interviewing.

Event Uncertainty Visualization. Time, location, and characters (or people) in this corpus were extracted using Java code and the aforementioned natural language processing tools. Each event was given an uncertainty score using the method described in Section 9.2.2. We first drew a Storygraph without uncertainty information. In that figure, key events – such as when the first and second plane hit and when the towers collapsed – are shown by $t1 - t4$. There are many co-occurring events around 15:00 hours, but the causes of these patterns are not yet clear.

Next, we plotted the storylines of four fire fighters before the South Tower collapsed with event uncertainty (Figure 15). It should be noted that two storylines crossing does not necessarily mean the two characters encounter each other; rather, it only means that two people were moving in directions diagonal to each other. One limitation of using uncertainty glyphs is that they might result in occlusion and ambiguity for large datasets. When the dataset is large, the bigger glyphs (e.g. the ones representing spatio-temporal uncertainty) could occlude the smaller ones.

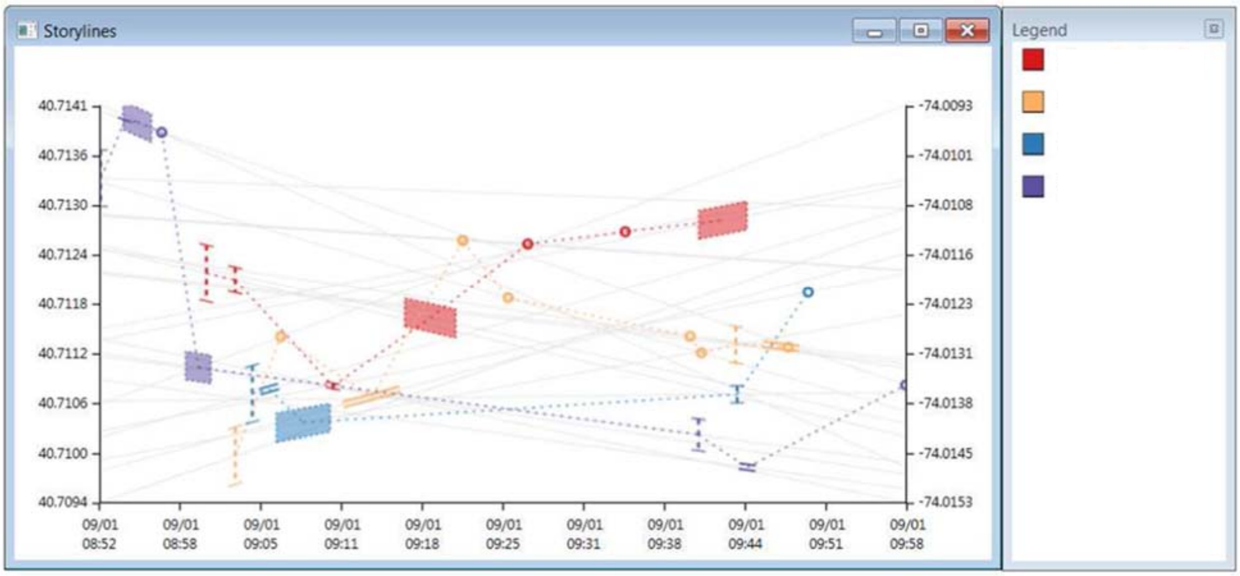

Figure 15: Storylines of four firefighters before the second tower collapsed along with event uncertainty. The dashed vertical I-beam shows the spatial uncertainty. The slope of the top and bottom portion of the beam shows the possible range of locations. The parallel lines show temporal uncertainty. The boxes represent spatio-temporal uncertainty and the circles show certain events.

Between-event Uncertainty Visualization. Figure 16 visualizes the between-event uncertainty for two victims: Father Judge and Chief Ganci. The space-time prisms in Storygraph enable users to see the possibilities of individuals encountering each other between key events. There are two patterns in this figure: (1) the prisms are only present between some key events, and (2) some prisms overlap. The first pattern indicates that locations of the two events are too far apart in that it would be impossible for a person to cover that distance at the maximum velocity. It does not necessarily mean that part of the story is false; rather, it may be the result of missing information between two events or uncertainty in the events themselves. From overlapping prisms (from Theorem 2), we can also deduce that Chief Ganci and Father Judge might have encountered each other within that time frame and region.

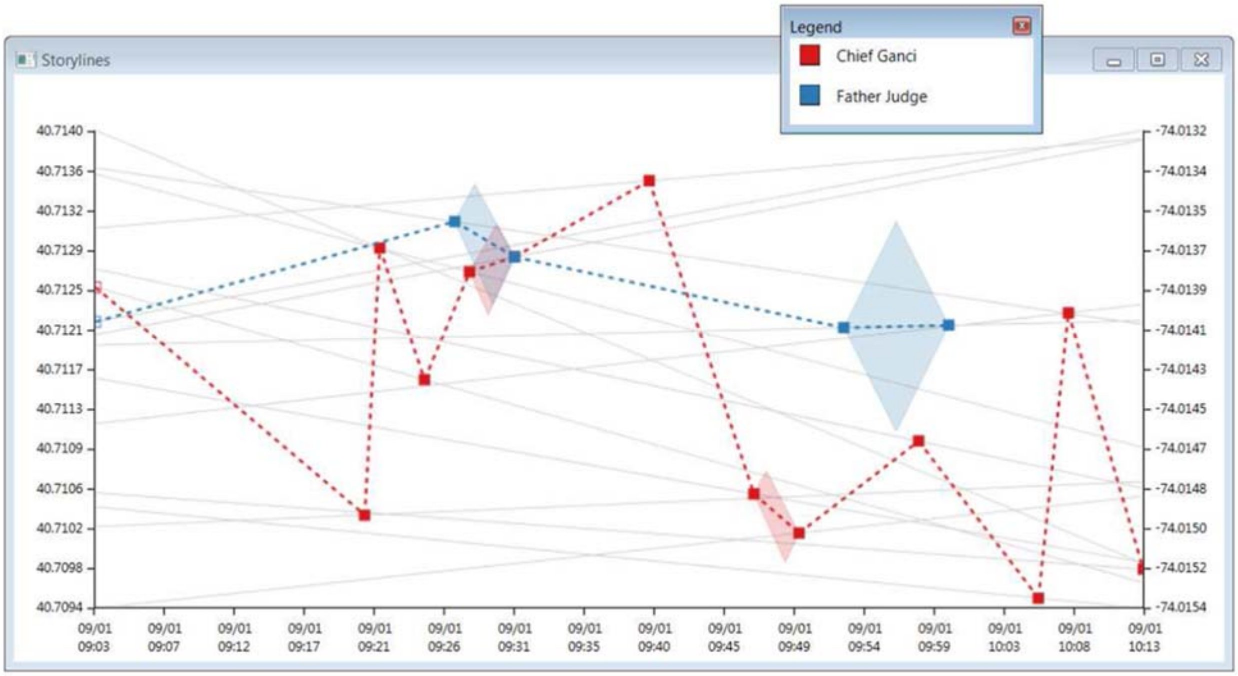

Figure 16: Storylines of WTC victims Chief Ganci and Father Judge. Father Judge was officially identified to be the first victim of the incident. Only a few key points have prisms between them. For other key points, the distance between them cannot be travelled within the given time at a velocity set by the user. This could either mean missing data points, change in velocity, or data reporting error.

10. NARRATIVE PROCESSING TO AID ENTITY IDENTIFICATION

In our research, we found that narratives of survival have two primary granular structures: they predominantly tell of events that either probably or certainly happened from the witness' perspective (see Section 7: Measuring Certainty), and they are composed of small narrative units comprised of person at location at time. Mieke Bal in *Narratology: Introduction to the Theory of Narrative* terms this composition first recognized by Russian Formalists, fabula, which precedes and facilitates the organizational work of narrative, or syuzhet. Language processing tools, when combined, can provide the processes for identifying the various fabulaic elements for extraction and incorporation into other analyses. Although the fabula/syuzhet schema has been well criticized, it serves as a good foundation for testing the possibilities of algorithmically produced transversal narratives and experimenting in methods for cross-document co-reference via narrative alignment. Narrative alignment is the basic recognition that a given cluster of person at location at time in one statement might be same as a cluster from another statement.

The Digging into Human Rights Violations project developed two mechanisms for testing this method for reinforcing cross-document co-reference. The first looked to explore how a person's relations during an event might typify the identity of that person across more than one testimony. The second looked to test if abstracted strings of key words for event descriptions were good markers of comparison for the reappearance of similar events across a set of testimonies. Each of these two exploratory methods has their situational strengths and weaknesses. The first, for example, works well on narratives that feature a relatively small set of entities. The second seems to perform best on set of stories composed of a series of discrete events.

The first method, one this project calls cross-document narrative frame alignment, looked to adapt high-level analyses techniques from computational linguistics reliant on methods like syntactic comparison to the challenge of event alignment. Automated cross-document comparison of narrative facilitates co-reference and event similarity identification in the retellings of stories from different perspectives. The developed method uses graph representations of the entities' co-presence at every event in a narrative as the basis of a similarity assessment. Composed of the entity-entity relations that appear in the events of a narrative, these graphs are represented by adjacency matrices populated with text extracted using various natural language processing tools. Graph similarity analysis techniques are then used to measure the similarity of events and the similarity of character function between stories. Preliminary experiments correctly co-referenced differently named entities from story variations and indicated the relative similarity of events in different iterations of the tale despite their order differences. Though promising, this work in progress also indicated some incorrect correlations between dissimilar entities.

Comparison of events across documents relies on the production of structured representations of events. In the case of this study, that structure is a matrix of entity-entity relations for each event. Events were automatically marked in the narratives using the Events in Text Analyzer (EVITA). EVITA uses statistical and linguistic approaches to identify and classify the language denoting orderable dynamic and stative situations (Llorens et al., 2010). Some language features indicative of events include finite clauses, event-referring nouns, and nominalized noun phrases (Sauri et al., 2005). Comparison via a neighborhood similarity function provided our primary comparison method to highlight event and character similarities. Although EVITA is fantastic in recognizing a high number of events in any given report, it lacks a robust theoretical model for what constitutes an event. A more robust theoretical model for what constitutes an event is being developed for implementation by the NewsReader project (Tonelli et al., 2014).

For every event detected by EVITA, a matrix of people and locations present at that event was produced. These matrices were representations of slices of a larger graph, which when compared, allowed for recognition that a given person from one narrative functions similarly to a person from another story. One secondary finding from this method related to the storytelling style of witness testimonies: they typically lack specific temporal and spatial referents. Rather, witnesses make subjective references to time, "before," "after," and to place, "in front of," "behind." Consequently, most NLP tools fail to identify the places of a testimony as places and the times of a testimony as times. A method like this one that relies on entity-to-entity connections and the sequence of those connections has the potential to align characters from different versions based upon their position within the story. The full methodology for the graph similarity process is documented in "Cross-Document Narrative Frame Alignment" (Miller et al., 2015).

The project's second method focused on abstracting the language used to narrate events and then comparing the sequences of abstracted event descriptions against each other. What this processes yielded was a method for identifying short, coherent narrative units and then finding occurrences of similar narrative units across a corpus. For example, the project tested the method with a set of newspaper articles about an incident that took place in Iraq in 2007. On September 16, 2007, operatives of a private security company known as Blackwater killed 17 civilians and injured 20 more during an operation that went through Baghdad's Nisour Square.

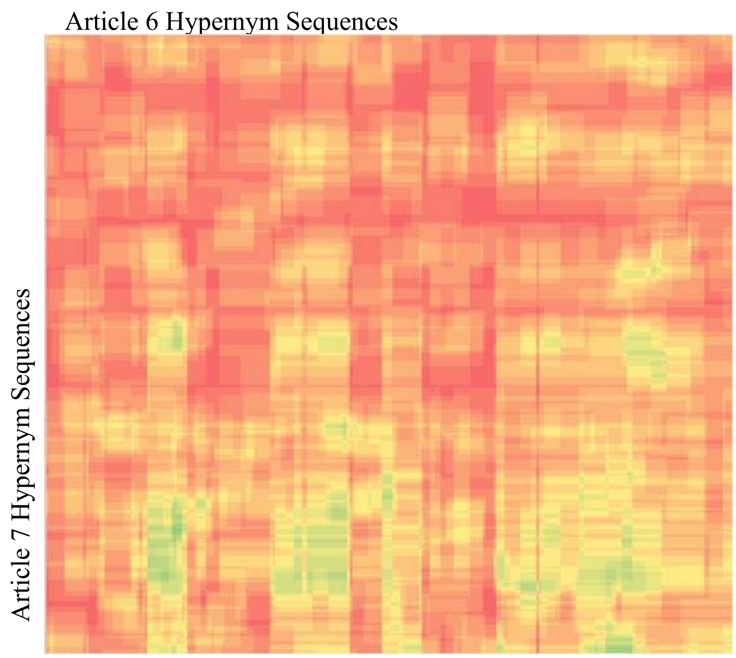

Figure 17: Dissimilarity graph showing the hypergram comparison across articles 6 and 7 using a color gradient scale where red indicates < 50%, yellow indicates50%, green indicates > 50% and up to 100%

Articles on Blackwater approach their story from many angles. Some focus on the appearance of key Blackwater executives before congress. Others look to relate witnesses' perspectives on the massacre and contain translated quotes. Yet others summarize the trial that convicted four of the firm's private security officers. The heterogeneity of the articles' foci on that event prevented the cross-document linking of specific event descriptions based on lexical features or with topic modeling algorithms. That challenge and the articles' description of multiple human rights violations made it a good test case for this method, which is described in detail in "Cross-Document Non-Fiction Narrative Frame Alignment" (Miller et al., 2015b). Briefly, the method used the events detected by the language processing tool EVITA, as did our prior graph-based method. The language describing those events was then abstracted using WordNet's hypernym feature. A hypernym is a word with broader meaning that captures many more specific words. For example, "tree" is the hypernym for many species of tree. In the case of event descriptions, one may find many different kinds of attribution ("said," "told," "explained," etc.) are all captured by one broader word.

Sequences of this abstracted language were constructed and, then, compared. For just 9 of the articles being studied, 2.2 million comparisons were made. One way the comparisons were interpreted was with a technique from machine learning known as a dissimilarity graph (see Figure 17). That graph is the pairwise comparison of each of the 227 sequences from article 6 (columns) against each of the 231 sequences from article 7 (rows). Values are color coded on as red to yellow to green along a 0-to-1 scale. Areas of similarity, such as the one just described that appears in the bottom left corner of that figure, fade in and out of the background dissimilarity as the sequences move into increasing then decreasing alignment.

Remarkably, this method is able to find moments of narrative similarity across a set of documents even when the style, language, and focalization of the storytelling are significantly different. As an example, consider the material presented in Table 4. In that table, one can see the article numbers (1, 6, and 7), the sequence numbers for each of the articles strings of hypernyms, and the sentences from the articles that correspond to those sequences. What this method allows for is a sense of narrative flow that is not bound by sentential or other structural units; a necessary advancement given that storytelling, particularly the storytelling of witnesses to traumatic events, often disobeys the ordered logic of grammar and pagination.

Table 4: Correspondences of sequences from 3 articles describing the 2007 shooting of civilians by Blackwater

Src.	Sq.	Hypernym Sequence	Source Sentence
1	209	talk, blast, disappoint, prevent, veto, surprise, blast, act, injure, veto, label, cease, blast, injure	"The shooting began at 12:08p.m., when at least one contractor began to fire on a car that failed to stop. The driver was killed and the car caught fire, but the contractors continued to shoot, killing the passengers and other Iraqis. At least one contractor reportedly called out to cease fire during the shooting, and another pointed his gun at a colleague" (Facts on File World News Digest, 2007).
1	210	blast, disappoint, prevent, veto, surprise, blast, act, injure, veto, label, cease, blast, injure, inform	
6	184	gunfire, express, perceive, perceive, blast, injure, express, blast, express, cut, affect, inspect, express, act	"All he saw, Sabah said, was that 'the white sedan moved a little bit and they started shooting.' As events unfolded and the Blackwater guards unleashed a storm of gunfire into the crowded square, Mr. Waso and Mr. Ali both said, they could neither hear nor see any return fire. 'It was one-sided shooting from one direction,' Mr. Waso said. 'There wasn't any return fire.' Mr. Waso said that what he saw was not only disturbing, but also in some cases incomprehensible. He said that the guards kept firing long after it was clear that there was no resistance" (Glanz, 2007).
6	185	express, perceive, perceive, blast, injure, express, blast, express, cut, affect, inspect, express, act, blast	
7	43	act, scat, injure, prevent, act, change_state, blast, injure, express, challenge, appear, injure, veto, talk	"The car continued to roll toward the convoy, which responded with an intense barrage of gunfire in several directions, striking Iraqis who were desperately trying to flee. Minutes after that shooting stopped, a Blackwater convoy – possibly the same one – moved north from the square and opened fire on another line of traffic a few hundred yards away, in a previously unreported separate shooting, investigators and several witnesses say. But questions emerge from accounts of the earliest moments of the shooting in Nisour Square. The car in which the first people were killed did not begin to closely approach the Blackwater convoy until the Iraqi driver had been shot in the head and lost control of his vehicle. Not one witness heard or saw any gunfire coming from Iraqis around the square" (Glanz and Rubin, 2007).
7	44	scat, injure, prevent, act, change_state, blast, injure, express, challenge, appear, injure, veto, talk, come	
7	45	injure, prevent, act, change_state, blast, injure, express, challenge, appear, injure, veto, talk, come, injure	
7	46	prevent, act, change_state, blast, injure, express, challenge, appear, injure, veto, talk, come, injure, suffer	
7	47	act, change_state, blast, injure, express, challenge, appear, injure, veto, talk, come, injure, suffer, perceive	
7	48	change_state, blast, injure, express, challenge, appear, injure, veto, talk,come, injure, suffer, perceive, cut	

11. PARTNER OUTCOMES

11.1. HRDAG and BENETECH

- HRDAG Director of Research Megan Price joined other DHRV team members at the 2013 World Social Science Forum to present general information on quantitative approaches to understanding mass violations of the right to life
- HRDAG's quantitative work was supplemented by DHRV qualitative investigations using NLP for tasks like Named Entity Recognition
- Connection to Ayush Shrestha and ideas about using Storygraphs to map perpetrator, victim, and witness movement through sets of documents (archives, testimonies, etc.)
- New uptake of NLP of free text fields to identify subset of records that included homicide victim information
- Project personnel Shrestha and Miller met with Benetech, the developers of human rights bulletining software, *Martus*, about what kinds and how to integrate visualization techniques into their package. Prior to that meeting, NGOs and individuals using that secure software were removing data from the secure environment to visualize it using tools like MS Excel. Integrating data visualization into the platform would help preserve the information security of rights violations documenting organizations working in active violation areas.

11.2. HURIDOCS

- Project personnel Miller met with HURIDOCS (Human Rights Information and Documentation Systems), a Geneva-based organization that develops systems and ontologies for violations reporting. Two efforts emerged from that meeting. 1) recognition of the need for a more balanced, well-structured rights violation ontology. 2) collaboration of a project to identity asymmetric applications of justice in the courts of the Pacific region using NLP techniques.

12. APPENDIX 1: WORKS CITED

Abul, O., Bonchi, F., & Nanni, M. (2008, April). Never walk alone: Uncertainty for anonymity in moving objects databases. *Data Engineering, 2008. ICDE 2008. IEEE 24th International Conference on*. 376-385.1109/ICDE.2008.4497446.

Brodlie, K., Osorio, R. A., & Lopes A. (2012). A review of uncertainty in data visualization. In *Expanding the Frontiers of Visual Analytics and Visualization*. London: Springer.

Broom, A., Cheshire, L., & Emmison, M. (2009). Qualitative researchers' understandings of their practice and the implications for data archiving and sharing. *Sociology*, *43*(6), 1163-1180.

Caruth, C. (1995). Introduction. In *Trauma: Explorations in Memory*. C. Caruth (Ed.). Baltimore, MD: The Johns Hopkins University Press. 3-12.

Caruth, C. (1996). *Unclaimed Experience: Trauma, Narrative, and History*. Baltimore, MD: The Johns Hopkins University Press.

Chang, A. X. & Manning, C. (2012, May) Sutime: A library for recognizing and normalizing time expressions. *LREC*, 3735–3740.

Cliggett, L. (2013). Qualitative data archiving in the digital age: Strategies for data preservation and sharing. *The Qualitative Report*, *18*(24), 1.

Demsar, U. & Virrantaus, K. (2010). Spacetime density of trajectories: exploring spatiotemporal patterns in movement data. *International Journal of Geographical Information Science* 24(10), 1527–1542.

Erman, L. D., Hayes-Roth, F., Lesser, V. R., & Reddy, D. R. (1980). The hearsay-II speech understanding system: Integrating knowledge to resolve uncertainty. *ACM Computing Surveys (CSUR)*, 12(2), 213–253.

Finkel, J. R., Grenager, T., & Manning, C. (2005). Incorporating non-local information into information extraction systems by gibbs sampling. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, 363–370. Association for Computational Linguistics.

Gilbert, L. S., Jackson, K., & di Gregorio, S. (2014). Tools for analyzing qualitative data: The history and relevance of qualitative data analysis software. In *Handbook of research on educational communications and technology* (pp. 221-236). Springer New York.

Hafner-Burton, E. M., & Ron, J. (2009). Seeing double. *World Politics*, *61*(02), 360-401.

Hägerstrand, T. (1989). Reflections on "what about people in regional science?" *Papers of the Regional Science Association*, 66(1), 1–6.

Hyland, K. (2005). *Metadiscourse*. London: Continuum.

Lenntorp, B. & Hort P. (1976). Paths in space-time environments: A time-geographic study of movement possibilities of individuals. *Lund Studies in Geography*, 44, Royal University of Lund.

Lewins, A., & Silver, C. (2007). Using software in qualitative research: A step-by-step guide. Sage.

Kraak, M. J. (2003). The space-time cube revisited from a geovisualization perspective. *Proc. 21st International Cartographic Conference*, 1988–1996.

Llorens, H., Saquete, E., & Navarro-Colorado, B. (2010, August). TimeML events recognition and classification: learning CRF models with semantic roles. In *Proceedings of the 23rd International Conference on Computational Linguistics* (pp. 725-733). Association for Computational Linguistics.

de Marneffe, M. C, Manning, C. D., & Potts, C. (2011). Veridicality and utterance understanding. *Proceedings of the 2011 IEEE Fifth International Conference on Semantic Computing, ser. ICSC '11*, 430–437, IEEE Computer Society. http://dx.doi.org/10.1109/ICSC.2011.10

de Marneffe, M. C, Manning, C. D., & Potts, C. (2012). Did it happen?: The pragmatic complexity of veridicality assessment. *Computational Linguistics* 38(2), 301-333.

Miller, B., Li, F. X., Shrestha, A., & Umapathy, K. (2013). Digging into human rights violations: Phrase mining and trigram visualization. *Digital Humanities*,*13*.

Miller, B., Shrestha, A., Olive, J. & Gopavaram, S. (2015, in press). Automated Comparison of Narrative and Character Function Similarity. In M. Finlayson, A. Lieto, B.Miller, & R. Ronfard (Eds.), *Sixth International Workshop on Computational Models of Narrative*. Atlanta, GA: CMN 2015.

Miller, B. Olive, J. Gopavaram, S. & Shrestha, A. (2015b, in press). Cross-Document Non-Fiction Narrative Alignnment. In *Proceedings of the 53nd annual meeting of the association for computational linguistics*.

Miller, H. J. (2005). A measurement theory for time geography. *Geographical analysis*, 37(1), 17–45.

Moreira, J., Ribeiro, C., & Saglio, J.M. (1999). Representation and manipulation of moving points: an extended data model for location estimation. *Cartography and Geographic Information Science* 26(2), 109–124.

Murray, A., Liu, Y., Rey, S., & Anselin, L. (2012). Exploring movement object patterns. *The Annals of Regional Science*, 1–14. 10.1007/s00168-011-0459-z.

Neutens, T., Weghe, N., Witlox, F. & Maeyer P. (2008). A three-dimensional network-based spacetime prism. *Journal of Geographical Systems*. 10(1), 89–107. http://dx.doi.org/10.1007/s10109-007-0057-x.

Orellana, D., Bregt, A. K., Ligtenberg, A., & Wachowicz, M. (2012). Exploring visitor movement patterns in natural recreational areas. *Tourism Management,* 33(3), 672 – 682.

Peters, V., & Wester, F. (2007). How qualitative data analysis software may support the qualitative analysis process. *Quality & quantity*, *41*(5), 635-659.

Pfoser, D., & Jensen, C. S. (1999). Capturing the uncertainty of moving-object representations. In *Advances in Spatial Databases*. 111–131. Springer: Berlin.

Potter, K., Rosen, P. & Johnson, C. R. (2012). From quantification to visualization: A taxonomy of uncertainty visualization approaches. In *Uncertainty Quantification in Scientific Computing*. 226–249. Springer: Berlin.

Ryan, G. W., & Bernard, H. R. (2003). Techniques to identify themes. *Field methods*, *15*(1), 85-109.

Saurí, R., Knippen, R., Verhagen, M., & Pustejovsky, J. (2005, October). Evita: a robust event recognizer for QA systems. In Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing (pp. 700-707). *Association for Computational Linguistics*.

Sauri, R. & Pustejovsky, J. (2009). Factbank: a corpus annotated with event factuality. *Language Resources and Evaluation*, 43(3), 227–268. http://dx.doi.org/10.1007/s10579-009-9089-9.

Shaw, S. L. & Yu, H. (2009). A GIS-based time-geographic approach of studying individual activities and interactions in a hybrid physical-virtual space. *Journal of Transport Geography,* 17(2), 141 – 149. http://www.sciencedirect.com/science/article/pii/S0966692308001324.

Shrestha, A., Zhu, Y., Miller, B., & Zhao, Y. (2013). Storygraph: Telling stories from spatio-temporal data. In *Advances in Visual Computing*. 693–702. Springer: Berlin.

Thrift, N. An introduction to time-geography. (1977). Geo Abstracts, University of East Anglia.

Tonelli, S., Sprugnoli, R., Speranza, M., & Minard, A. L. (2014). *Newsreader guidelines for annotation at document level*. Technical Report NWR-2014-2, Fondazione Bruno Kessler. http://www. newsreaderproject. eu/files/2013/01/NWR-2014-2. pdf.

Verhagen, M., Mani, I., Sauri, R., Knippen, R., Jang, S. B., Littman, J., Rumshisky, A., Phillips, J., & Pustejovsky, J. (2005). Automating temporal annotation with TARSQI. In *Proceedings of the ACL 2005 on Interactive poster and demonstration sessions*. 81–84. Association for Computational Linguistics.

13. APPENDIX 2: PRESENTATIONS AND PUBLICATIONS

Miller, B., Olive, J., Gopavaram,S., & Ayush, S. (in press). Cross-Document Non-Fiction Narrative Frame Alignment. In C. Wanxiang & G. Zhou (Publication Chairs), M. Kan (Ed.), *The First Workshop on Computing News Storylines (CNewS 2015): ACL Anthology.* Bejing, China: Association for Computational Linguistics.

Miller, B., Olive, J., Gopavaram,S., & Ayush, S. (2015, July). *Cross-Document Non-Fiction Narrative Frame Alignment.* Paper presented at the First Workshop on Computing News Storylines (CNewS 2015). Bejing, China.

Miller, B., Shrestha, A., Olive, J. & Gopavaram, S. (2015, July). *Automated Comparison of Narrative and Character Function Similarity Using Graph Theory*. Paper presented at Digital Humanities 2015. Sydney, Australia.

Miller, B., Shrestha, A., Olive, J. & Gopavaram, S. (in press). Automated Comparison of Narrative and Character Function Similarity. In M. Finlayson, A. Lieto, B.Miller, & R. Ronfard (Eds.), *Sixth International Workshop on Computational Models of Narrative*. Atlanta, GA: CMN 2015.

Miller, B., Shrestha, A., Olive, J. & Gopavaram, S. (2015, May). *Automated Comparison of Narrative and Character Function Similarity.* Paper presented at the Sixth International Workshop on Computational Models of Narrative, Atlanta, GA.

Miller, B., Shrestha, A., & Olive, J. (2014). Visualizing Computational, Transversal Narratives from the World Trade Towers. *Literary and Linguistic Computing*. Manuscript submitted for publication.

Shrestha, A. (2014). *Visualizing spatio-temporal data* (Unpublished doctoral dissertation). Georgia State University, Atlanta, GA.

Shrestha, A. Exploring Events (E2): Spatio-temporal data visualization using Storygraph (Version 1.0). [Software]. Available from https://github.com/sayush/E2.

Shrestha, A., Zhu, Y., & Manandhar, K. (2014). NetTimeView: Applying Spatio-temporal Data Visualization Techniques to DDoS Attack Analysis. In *Advances in Visual Computing* (pp. 357-366). Springer International Publishing.

Xiao, L., Elueze , I., & Kavanaugh, J. (2014, November). Human Rights Researchers' Data Analysis and Management Practices. In Grove, A. (Ed.), *Proceedings of the 77th Annual Meeting of Association for Information Science and Technology (ASIS&T)*. Seattle, WA: Association for Information Science and Technology.

Miller, B., Olive, J., Shrestha, A., Subtirelu, N., Zhao, J., & Zhao, Y. (2014, October). *Graph Comparison of Narrative Texts*. Paper presented at the 2014 Chicago Colloquium on Digital Humanities and Computer Science (DHCS 2014). Evanston, IL.

Miller, B., Olive, J., Shrestha, A., Subtirelu, N., Zhao, J., & Zhao, Y. (2014, October). *Narrative Markup with TEI*. Paper presented at the 2014 Text Encoding Initiative Conference and Members Meeting. Evanston, IL.

Miller, B., Olive, J., Shrestha, A., Subtirelu, N., Zhao, J., & Zhao, Y. (2014, September). *Language Processing Pipeline for Narrative Emergence: Digging into Human Rights Violations*. Paper presented at the 2014 Japanese Association for Digital Humanities Conference (JADH 2014). Tsukuba, Ibaraki, Japan.

Miller, B., Shrestha, A., & Olive, J. (2014, July). *Visualizing Computational, Transversal Narratives from the World Trade Towers*. Paper presented at Digital Humanities 2014. Lausanne, Switzerland.

Miller, B., Shrestha, A., & Subtirelu, N. (2014, March). *NLP Approaches to Rights Violation Classifying*. Paper presented at RightsCon 2014. San Francisco, CA.

Miller, B. & Xiao, L. (2013, October). *Digging into Human Rights Violations: Anaphora Resolution and Emergent Witnesses*. Paper presented at Digging into Data 2013. Montreal, Canada.

Miller, B., Shrestha, A., Derby, J., Olive, J., Umapathy, K., Li, F. & Zhao, Y. (2013, October). Digging into Human Rights Violations: Data Modeling Collective Memory. In Hu, X., Lin, T. Y., Raghavan, V., Wah, B., Baeza-Yates, R., Fox, G. Nambiar, R. (Eds.), *Big Data, 2013 IEEE International Conference on*. 37-45. Silicon Valley, CA: IEEE Xplore.

Miller, B., Shrestha, A., Derby, J., Olive, J., Umapathy, K., Li, F. & Zhao, Y. (2013, October). *Digging into Human Rights Violations: Data Modeling and Collective Memory*. Paper presented at the 2013 IEEE International Conference on Big Data. Big Data and the Humanities Workshop. Santa Clara, California.

Griffin, D., Roughton, B., Watkins, T., Zamost, S., Barthlow, D., & Miller, B. (2013, September). *Text Mining Historical News Sources.* Panel on Big Data and Investigative Journalism. In *Future of Journalism Symposium*. Atlanta, GA.

Shrestha, A., Zhu, Y., & Miller, B. (2013, September). *Visualizing Time and Geography of Open Source Software with Storygraph*. Paper presented at the IEEE First Working Group on Visualization of Software (VisSoft) / International Conference on Software Maintenance. Eindhoven, Netherlands.

Shrestha, A., Miller, B., Zhu, Y., & Zhao, Y. (2013, August). Storygraph: Extracting patterns from spatio-temporal data. In *Proceedings of the ACM SIGKDD Workshop on Interactive Data Exploration and Analytics* (pp. 95-103). ACM.

Shrestha, A., Zhu, Y., Miller, B., & Zhao, Y. (2013). Storygraph: Telling stories from spatio-temporal data. In *Advances in Visual Computing*. 693–702. Springer: Berlin.

Miller, B., Li, F., Shrestha, A., & Umapathy, K. (2013, July). *Digging into Human Rights Violations: Phrase mining and trigram visualization*. Paper presented at Digital Humanities 2013. Lincoln, NE.

Miller, B. (2013, April). *Transversal Reading of Collective Memory*. Paper presented at the 2013 Humanities, Arts, science, and Technology Advances Collaboratory Conference (HASTAC 2013). Toronto, Canada.

Miller, B. (2013, April). *Digging into Human Rights Violations*. [Invited Speaker]. Virginia Polytechnic Institute and State University, Blacksburg, VA.

Miller, B. (2013, April). *Crisis Mapping: Lines, Circles, Surfaces*. Paper presented at the American Comparative Literature Association 2013 Conference (ACLA 2013). Toronto, Canada.

Weber, B. (2013, March 17). HASTAC Digital History Group's Spring Spotlight Series # 3: Digital History's Relationship to Human Rights Archives and Data Analysis. [Blog]. Retrieved from http://www.hastac.org/blogs/bendavidweber/2013/03/17/spring-spotlight-3-digital-history%E2%80%99s-relationship-human-rights-archiv.

Miller, B. (2012, November). *Anaphora and Transversal Reading*. [Invited Speaker]. Information Interfaces Research Group, College of Computing, Georgia Institute of Technology, Atlanta, GA.

Miller, B., Xiao, L., & Umapathy, K. (2012, June). *Digging into Human Rights Violations*. [Invited Speakers]. Western University, London, Canada.

14. ACKNOWLEDGEMENTS

Digging into Human Rights Violations gratefully acknowledges and appreciates the support of the National Science Foundation (Award 1209172), the Social Science and Humanities Research Council, the National Endowment for the Humanities Office of Digital Humanities, and the Digging into Data Challenge. The project would also like to thank the organizations that helped enable the project by carefully sharing this difficult and sensitive data, the institutions that hosted this research, and the researchers and investigators who consulted and advised the project on its goals, models, and outcomes. Lastly, the project would like to thank the many students who contributed to this research; their efforts, diligence, and curiosity were remarkable and inspiring. The generosity of these agencies, organizations, and individuals made the project more well-informed, relevant, and ultimately, meaningful.