January 2015

Evidence of Things Not Seen: A Semi-Automated Descriptive Phrase and Frame Analysis of Texts about the Herbicide Agent Orange

Sarah Beth Hopton

University of South Florida, sbhopton@gmail.com

Follow this and additional works at: http://scholarcommons.usf.edu/etd

Part of the Other Communication Commons, and the Rhetoric Commons

Scholar Commons Citation


This Dissertation is brought to you for free and open access by the Graduate School at Scholar Commons. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact scholarcommons@usf.edu.
Evidence of Things Not Seen: A Semi-Automated Descriptive Phrase and Frame Analysis of
Texts about the Herbicide Agent Orange

by

Sarah Beth Hopton

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy Rhetoric & Composition
Department of English
College of Arts and Science
University of South Florida

Major Professor: Carl Herndl, Ph.D.
Meredith Johnson, Ph.D.
John Skvoretz, Ph.D.
Jurgen Pfeffer, Ph.D.

Date of Approval:
May 4, 2015

Keywords: digital discourse analysis, communicating science, technical communications.

Copyright © 2015, Sarah Beth Hopton
DEDICATION

To those who still suffer.

*Mutato nomine et de te fabula narrator:*

Change only the name and this story is about you.
ACKNOWLEDGMENTS

This work would not have been possible without the help and contributions of many people.

I am particularly indebted to my committee, consisting of Drs. Herndl, Johnson, Skvoretz, and Pfeffer, and to my research collaborators and coding partners, Fred Morstatter and Dana Rine.

I am also indebted to my family, whose unflattering faith kept me buoyed in the roughest of seas.
# TABLE OF CONTENTS

LIST OF TABLES .................................................................................................................. iii

LIST OF FIGURES .............................................................................................................. v

ABSTRACT ............................................................................................................................... viii

CHAPTER 1: HISTORY & CONTEXT .................................................................................... 1
  Overview ............................................................................................................................... 1
  The Birth of Agent Orange ................................................................................................. 4
  America in Vietnam ........................................................................................................... 7
  Agent Orange as Scientific Controversy ........................................................................... 10
  Agent Orange in the US News Media ............................................................................... 16
  Agent Orange as Legal Controversy ............................................................................... 18
  American Skepticism & Vietnamese Acceptance of Agent Orange ......................... 19
  Networked Rhetorics ....................................................................................................... 27

CHAPTER 2: THEORY & DEFINITIONS ............................................................................... 31
  Metaphors, Descriptive Phrases, and Why They Matter ............................................. 31
  From Metaphors to Frames .............................................................................................. 42
  Critical Discourse Analysis as a Means to Interpret Metaphors and Frames ............ 45
  Historical and Modern Approaches to Computer-Aided Discourse Analysis ......... 47

CHAPTER 3: METHODOLOGY ............................................................................................ 52
  A Definition of Big Data .................................................................................................... 52
  Brief Description of Research Site .................................................................................. 57
  Corpora in Detail ................................................................................................................ 59
    Congressional Hearings & Reports .............................................................................. 59
    New York Times Articles ............................................................................................... 61
    New York Times Editorials ............................................................................................ 61
    The Vietnamese News Agency ..................................................................................... 62
  Research Questions .......................................................................................................... 62
  Determining the Unit of Analysis .................................................................................... 63
  Method for Data Collection and Analysis ...................................................................... 66
  Distant Reading, Interpretation, and Analysis ................................................................. 71
  Limitations ......................................................................................................................... 80
    Human Error ................................................................................................................... 81
    Multiple Theoretical Perspectives ............................................................................... 82
    Problems Specific to Big Data ....................................................................................... 82
    Questions about Mixing Methods ............................................................................... 84
Conclusion ..........................................................................................................................85

CHAPTER 4: RESEARCH FINDINGS ..................................................................................86
  The Environment ..............................................................................................................90
  The Uncertainty of Agent Orange ..................................................................................97
  Victims & Victimization .................................................................................................100
  Remuneration ................................................................................................................103
  Ways of Describing Agent Orange ................................................................................108

CHAPTER 5: INTERVENTION, VISUALIZATION, AND THE FUTURE
  PROSPECTS OF BIG DATA AND AGENT ORANGE ..................................................116
  Intervention & Social Change .......................................................................................116
  The Value and Possibilities of Network Analysis & Visualization ...............................120
  Distant Reading through Network Visualization ........................................................121
  Overcoming Limitations Specific to this Project ..........................................................124
  Studying Emerging Sites of Political and Scientific Controversy ...............................126
  Summary of Results ....................................................................................................130
  Conclusion ....................................................................................................................133

REFERENCES ..................................................................................................................134

APPENDIX ......................................................................................................................145
LIST OF TABLES

Table 1. List of Corpora Contents........................................................................................................62

Table 2. Examples of Themes and Expressions......................................................................................69

Table 3. Equations and Examples of Resulting Metaphorical Frames .................................................70

Table 4. The Top 5 Keywords per Stakeholder for Cross-section of Corpus: “Environment.” ...............91

Table 5. Top 5 Descriptive Phrases per Stakeholder for Cross-section of Corpus: “Environment.” ........93

Table 6. Top 5 Co-Occurrences per Stakeholder for Cross-section of Corpus: “Environment.” ...............95

Table 7. Top 5 Keywords per Stakeholder for Cross-section of Corpus: “Uncertainty.” .........................97

Table 8. Descriptive Phrases per Stakeholder for Cross-section of Corpus: “Uncertainty.” ...................98

Table 9. Top 5 Co-Occurrences per Stakeholder for Cross-section of Corpus: “Uncertainty.” .................99

Table 10. Top Keywords per Stakeholder for Cross-section of Corpus: “Victim.” .................................100

Table 11. Top 5 Descriptive Phrases per Stakeholder for Cross-section of Corpus: “Victim.” ................101

Table 12. Top 5 Co-Occurrences per Stakeholder for Cross-section of Corpus: “Victim.” ....................103

Table 13. Top 5 Keywords per Stakeholder for Cross-section of Corpus: “Remuneration.” ...................103

Table 14. Top 5 Descriptive Phrases per Stakeholder for Cross-section of Corpus: “Remuneration.” ..........105
Table 15. Top 5 Co-Occurrences per Stakeholder for Cross-section of Corpus: “Remuneration.” ................................................................. 106

Table 16. Top 5 Keywords per Stakeholder for Cross-section of Corpus: “Agent Orange.” ................................................................. 108

Table 17. Top 5 Descriptive Phrases per Stakeholder for Cross-section of Corpus: “Agent Orange.” ................................................................. 109

Table 18. Top 5 Co-Occurrences per Stakeholder for Cross-section of Corpus: “Agent Orange.” ................................................................. 110
LIST OF FIGURES

Figure 1: Co-location Matrix ...........................................................................................................................................49

Figure 2: Data Organization in Excel ..........................................................................................................................71

Figure 3: New York Times Network Representation ....................................................................................................77

Figure 4: New York Times Editorial Network Representation ..........................................................................................79

Figure 5: Congressional Corpora Network Representation ..........................................................................................112

Figure 6: New York Times Corpora Network Representation ..........................................................................................113

Figure 7: New York Times Editorial Corpora Network Representation ........................................................................114

Figure 8: Vietnamese News Agency Corpora Network Representation ........................................................................115

Figure 9: Sources in Media ..............................................................................................................................................120

Figure 10: Frequency List for Congressional/Environment 1960-1969 ........................................................................145

Figure 11: Descriptive Phrases, Congress/Environment, 1960-1969 ...........................................................................146

Figure 12: Co-Occurrence Congress/Environment, 1960-1969 ..................................................................................147

Figure 13: Frequency List, Congress/Environment ......................................................................................................147

Figure 14: Descriptive Phrases, Congress/Environment .................................................................................................148

Figure 15: Co-Occurrence List, Congress/Environment ................................................................................................148

Figure 16: Frequency List, New York Times/Environment ..............................................................................................149

Figure 17: Descriptive Phrases, New York Times/Environment ......................................................................................149

Figure 18: Co-Occurrence List, New York Times/Environment ......................................................................................150

Figure 19: Frequency List, Vietnamese News Agency/Environment ..............................................................................150
Figure 20: Descriptive Phrases, Vietnamese News Agency/Environment ......................................... 151
Figure 21: Co-Occurrence List, Vietnamese News Agency/Environment .................................. 151
Figure 22: Frequency List, Congress/Uncertainty ............................................................................. 152
Figure 23: Descriptive Phrases, Congress/Uncertainty ................................................................. 152
Figure 24: Co-Occurrence List, Congress/Uncertainty ................................................................. 153
Figure 25: Frequency List, New York Times/Uncertainty ............................................................... 154
Figure 26: Descriptive Phrases, New York Times/Uncertainty ..................................................... 154
Figure 27: Co-Occurrence List, New York Times/Uncertainty ....................................................... 155
Figure 28: Frequency List, Congress/Victimization ..................................................................... 155
Figure 29: Descriptive Phrases, Congress/Victimization ............................................................. 156
Figure 30: Co-Occurrence List, Congress/Victimization ............................................................. 157
Figure 31: Frequency List, New York Times/Victimization ........................................................... 158
Figure 32: Descriptive Phrases, New York Times/Victimization .................................................. 158
Figure 33: Co-Occurrence List, New York Times/Victimization .................................................. 159
Figure 34: Frequency List, New York Times Editorial/Victimization ........................................... 160
Figure 35: Descriptive Phrases, New York Times Editorial/Victimization ................................... 160
Figure 36: Co-Occurrence List, New York Times Editorial/Victimization ................................... 161
Figure 37: Frequency List, Vietnamese News Agency/Victimization ......................................... 161
Figure 38: Descriptive Phrases, Vietnamese News Agency/Victimization ................................. 162
Figure 39: Co-Occurrence List, Vietnamese News Agency/Victimization ................................. 163
Figure 40: Frequency List, Congress/Remuneration .................................................................... 164
Figure 41: Descriptive Phrases, Congress/Remuneration ............................................................ 164
Figure 42: Co-Occurrence List, Congress/Remuneration ............................................................ 165
Figure 43: Frequency List, New York Times/Remuneration ........................................165
Figure 44: Descriptive Phrases, New York Times/Remuneration ..................................166
Figure 45: Co-Occurrence List, New York Times/Remuneration ......................................166
Figure 46: Frequency List, New York Times Editorial/Remuneration ...............................167
Figure 47: Descriptive Phrases, New York Times Editorial/Remuneration ........................167
Figure 48: Co-Occurrence List, New York Times Editorial/Remuneration ........................168
Figure 49: Frequency List, Vietnamese News Agency/Remuneration .............................168
Figure 50: Descriptive Phrases, Vietnamese News Agency/Remuneration ......................169
Figure 51: Co-Occurrence List, Vietnamese News Agency/Remuneration .......................169
Figure 52: Frequency List, Congress/Agent Orange ......................................................170
Figure 53: Descriptive Phrases, Congress/Agent Orange ...............................................170
Figure 54: Co-Occurrence List, Congress/Agent Orange ...............................................171
Figure 55: Frequency List, New York Times/Agent Orange ............................................171
Figure 56: Descriptive Phrases, New York Times/Agent Orange ....................................172
Figure 57: Co-Occurrence List, New York Times/Agent Orange ....................................172
Figure 58: Frequency List, New York Times Editorial/Agent Orange .............................173
Figure 59: Descriptive Phrases, New York Times Editorial/Agent Orange ......................173
Figure 60: Co-Occurrence List, New York Times Editorial/Agent Orange .......................174
Figure 61: Frequency List, Vietnamese News Agency/Agent Orange ............................174
Figure 62: Descriptive Phrases, Vietnamese News Agency/Agent Orange ......................175
Figure 63: Co-Occurrence List, Vietnamese News Agency/Agent Orange .......................175
ABSTRACT

From 1961 to 1971 the United States and the Republic of South Vietnam used chemicals to defoliate the coastal and upload forest areas of Viet Nam. The most notorious of these chemicals was named Agent Orange, a weaponized herbicide made up of two chemicals that, when combined, produced a toxic byproduct called TCDD-dioxin. Studied suggest that TCDD-dioxin causes significant human health problems in exposed American and Vietnamese veterans, and possibly their children (Agency, U.S. Environmental Protection, 2011). In the years since the end of the Vietnam War, volumes of discourse about Agent Orange has been generated, much of which is now digitally archived and machine-readable, providing rich sites of study ideal for “big data” text mining, extraction and computation. This study uses a combination of tools and text mining scripts developed in Python to study the descriptive phrases four discourse communities used across 45 years of discourse to talk about key issues in the debates over Agent Orange. Findings suggests these stakeholders describe and frame in significantly different ways, with Congress focused on taking action, the New York Times article and editorial corpus focused on controversy, and the Vietnamese News Agency focused on victimization. Findings also suggest that while new tools and methods make lighter work of mining large sets of corpora, a mixed-methods approach yields the most reliable insights. Though fully automated text analysis is still a distant reality, this method was designed to study potential effects of rhetoric on public policy and advocacy initiatives across large corpora of texts and spans of time.
CHAPTER 1: HISTORY & CONTEXT

Overview

Agent Orange as a site of study is particularly complicated, as it includes nearly 45 years of scientific, political, historical, environmental, technical, media, and military discourse. This thesis uses a mixed-methods approach to semi-automate the identification, extraction, and comparison of descriptive language from the discourse to illustrate the central debates and frames used by key stakeholders in the rhetorics of Agent Orange. The aim of this project is to analyze and hypothesize about the ways in which describing and framing Agent Orange has helped or hurt the cases made by organizations and individuals advocating for remediation of the habitats Agent Orange destroyed, and remunerating those it disabled.

In Chapter 1, I describe how early framing of weeds as an “enemy”, the widespread adoption and use of herbicides domestically, and a general trust and excitement about technology, primed the country to accept the idea of weaponized herbicides and their use during the Vietnam War. I then explain why and in what ways Agent Orange became a site of significant scientific, political, and legal controversy, and the effects that uncertainty has had on policies that mediate access to care for those who attribute illness to exposure.

The production and use of Agent Orange has been controversial from the start. Military leaders were unconvinced weaponized herbicides would provide military advantage; scientists were concerned about human health consequences; environmentalists were concerned about long-term ecological effects; and activists lodged the use of the chemical violated the Geneva
protocol. The controversies did not end there though. Not long after Operation Ranch Hand—the name of the primary spray program—was canceled, veterans and Vietnamese began complaining of significant health problems, but until veterans mobilized politically in the 1980s, their claims were not taken seriously. Though veteran and Vietnamese health claims were eventually studied, the science of Agent Orange was also embattled, which led to greater skepticism and doubt about Agent Orange’s effects, and greater veteran frustration and bitterness; the U.S. government felt veterans were malingerers, and Vietnam veterans felt their country had betrayed and left them to die.

The controversy surrounding Agent Orange remains ongoing, as veterans, the Vietnamese and now the children of Vietnam veterans petition the U.S. government and the scientific and international humanitarian communities to remediate the soil contaminated by dioxin, and remunerate victims poisoned by the chemical. Meanwhile, the chemical companies who originally produced Agent Orange, like Monsanto, are embroiled in modern controversies related to Agent Orange. Recently, Monsanto tried to cut a deal with the Vietnamese government to import genetically modified food products, but the public relations challenges were too significant to overcome, and the project was scrapped. For many Vietnamese, chemical manufacturers like Monsanto are directly responsible for the ecological damage wrought on the environment in Vietnam and the U.S. (there are several toxic Superfund sites in the United States where former Agent Orange manufacturing plants once operated). Yet, with few exceptions, these companies have never claimed responsibility for their role in the production, distribution or the harm caused those who were exposed.

I conclude Chapter 1 with three suggestions. First, like other scientific controversies, including the Monsanto Protection Act, genomic patenting, bio-nuclear terror protocols and
global warming, Agent Orange underscores the importance of constructing more humane science and technology policies. Second, the ways we describe and frame subjects and issues shape our political practices, driving more or less human public policies. Thus, mapping how we describe key themes or objects in moments of political controversy is an important tool by which to make visible the often-invisible ways that language shapes our political values, attitudes and behavior. Third and finally, technical communicators are increasingly called to do this kind of work—work that is both methodologically rigorous, technical, and ultimately applied.

These arguments and background sets the stage for the work of Chapter 2, where I detail the theoretical underpinnings of this project, which are grounded in metaphor and frame analysis, critical discourse analysis, network analysis and corpus linguistics. Once I outline the value in studying metaphorical frames, I turn, in Chapter 3, to the methodology developed to locate, extract, and calculate descriptive phrases on a specific corpus composed of texts about Agent Orange from the Congressional Record, The New York Times, and the Vietnamese News Agency. Chapter 3 also details the potential problems and pitfalls scholars face when undertaking this kind of “big data”, mixed-method textual analysis, terms that are also explained and problematized. Chapter 4 visualizes and explores the results of my analyses, and attempts to connect findings to shifts in policy or practice in American and Vietnam. Finally, Chapter 5 argues suggests practical applications of this method and argues that mixed-methods that borrow from Sociology, Computer Science, and Linguistics offer an opportunity to extend Rhetoric’s reach and disciplinary validity. This chapter also offers suggestions for how to overcome the central problems of study design and corpus building.
The Birth of Agent Orange

Though this project is not meant to be a history, history is where we must start and history demonstrates that it is war—not peace—that often spurs scientific and technological innovation. Consequently, science and technology have made substantial impact on battlefields.

On the heels of the Second World War, the American Federal Government was particularly interested in how science could improve wartime outcomes. Vannevar Bush, head of the U.S. Office of Scientific Research and Development (OSRD), led many of Bush’s research projects. His influence as both an administrator and inventor was far-reaching. He worked on the development of analog computers, founded the defense and aerospace company Raytheon, and patented numerous inventions like the microfilm viewer and the Memex, a proto-hypertext system of storage and retrieval. In 1945, Bush described his vision for the Memex system writing that, “wholly new forms of encyclopedias” would appear with a “mesh of associative trails running through them”—clearly, an early vision of the Internet and the semantic web (Bush, 1945). It was an exciting time full of technocratic hope, and the country generally felt positive about the use and development of technology to solve both war and peacetime problems (Johnson, 2012).

Billions of public dollars were spent on research and technology during this time (The Administration, 1945), including dollars spent on battling an anticipated exhaustion of food resources that an increasingly prosperous middle class America demanded. Plant biologists, many of whom worked for the U.S. government, found a particularly promising area of focus in the study of hormones (Kaempffert, 1948).

From the 1940s to the 1960s, America’s foreign policy fluctuated between post-World War II isolationism and the interventionist responses to intensifying Cold War pressures to
vanquish a spreading Communism. While American military were waging literal war against Fascism and Communism, the American scientific community was waging a metaphorical and chemical battle against Nature, with much of the language used to describe this battle invoking metaphors of combat. War metaphors were especially popular in the discourse of weeds. In both the popular media and in science journals weeds were metaphorized as killers; they were “pernicious” and “noxious” and had to be wiped out before they “strangled” helpful grasses.

Writing in the *Journal of Royal Horticultural Society* in 1956, William Stearn observed that people in plant science characterized weeds less as a botanical category than as a psychological one. “A weed is simply a plant which in a particular place at a particular time arouses human dislike” (p. 285). Such emotional characterizations pulsed through the popular media too, invoking powerful disdain for weeds from garden-club members in New York to ranchers in Utah. This disdain, coupled with the golden era of advertising, would power the domestic consumption of weed killers, more commonly known as herbicides.

There were many scientists working on the weed problem in the early 40s, but in June of 1941, according to environmental historian David Zierler, the first published account of a chemical named 2,4-Dichlorophenoxyacetic acid (2,4-D) was credited to chemist R. Pokorny. The following year, John Lontz, an American chemist who worked for E.I. du Pont de Nemours Company (called DuPont today) applied for and received a patent for 2,4-D, but it was a scientist named Arthur Galston who would launch 2,4-D into infamy.

While finishing his doctoral degree, which analyzed the properties of soy plants, Galston stumbled upon the incredible herbicidal capacity of a second chemical called 2,3,5-Triiodobenzoic acid (2,4,5-T) (Zeirler, 2011, p. 37). Though growth hormone research was used domestically to govern an ever-larger scope of growth and production issues, Galston’s research
ultimately lead to tailored hormones that could produce or inhibit particular properties pinpointed
in specific plants. The foundational research that scientists like Pokorny, Lontz and Galston
conducted not only provided the appropriate conditions necessary for the development and
release of weaponized herbicides like Agent Orange—a fact that that troubled Galston
particularly—but would also help launch the highly controversial genetically modified food
movement decades later (Dien, 2014).

Domestically released in 1946, 2,4-D launched a revolution in agricultural output and
became a highly successful selective herbicide that could isolate and kill harmful weeds that
destroyed wheat, corn, rice, and cereal crops without hurting helpful grasses that protected them.
Since its creation in the late 40s, 2,4-D has remained one of the top grossing herbicides used
globally because it is cheap and extremely effective (Hawkins, 2010).

After the production of 2,4-D and the discovery of 2,4,5-T, the use of domestic herbicides
exploded. Domestic production of 2,4,D went from 14 to 36 million pounds per year, and
production of 2,4,5-T increased from virtually 0 to 10 million pounds in a single decade.
Predictably, herbicides quickly became a significant profit center for chemical manufacturing
companies, and the preferred weapon in the domestic war against weeds because they were
cheap and readily available. Soon Diamond Alkali (now Dow Chemical), Hercules, Hooker
Chemical (now Occidental Chemical Company), and Monsanto produced more than 75 million
pounds of herbicides annually, generating well over $100 million dollars a year in revenue
(Martini, 2012, p. 23).

As production of herbicides increased, so did the militaristic ways weeds were written
about in technical, scientific, and journalistic discourse. Words like “alien”, “feral,” “invader”,
and “invasive” were common ways of describing weeds, all of which provoke conceptual
metaphors of war, and required an aggressive response in turn. Though “weeds” were described in terms of battle, the term “herbicide” seems to have been perceived neutrally, despite the fact that it was positioned in advertisements of the day as the weapon of choice against invasive weeds. Copywriters and spin-doctors presumably selected the term “herbicide” over “chemical” because of its positive valence.

Not all Americans and scientists embraced the use of herbicides unhesitatingly though. Anxieties associated with their widespread development and adoption began to percolate in the early 1960s and boiled over with the publication of Rachel Carsen’s *Silent Spring*. Carsen’s 1962 book accused the chemical industry of spreading misinformation about the dangers of chemical use and development, and public officials of accepting industry claims unquestioningly. Carsen further contended that industry fudged its science to sell more to its top clients, which, in the late 60s included the U.S. government. This claim would prove regrettably prescient in the case of Agent Orange. Galston, who helped invent Agent Orange, was also among those in the scientific community that protested its weaponization and use during and after the Vietnam War.

**America in Vietnam**

The Vietnam War, sometimes referred to as the Second Indochina War, was a proxy war that took place in Vietnam, Laos, and Cambodia between December 1956 and the fall of Saigon, April 30, 1975. It was fought between Northern Vietnam, backed by the Soviet Union, China and her communist allies, and the government of South Vietnam, supported by the US and other anti-communist allies (*Vietnam: A History, 1997*).¹

In 1961, experiments at the joint American-South Vietnamese Combat Development Test Center generated a request to the US Department of State to use chemicals for a program

¹ Though a gross oversimplification, America’s entry into Vietnam was the result of a series of complex ideological impulses and political miscalculations. Even an abridged history or justification for the U.S. entry into the Vietnam War is beyond the scope of this thesis, but are the topic of great works of scholarship by Stanley Karnow (*Vietnam: A History*) and Phillip Caputo (*A Rumor of War*).
conceived to destroy enemy crops and food sources, and thereby demoralize the Việt Cộng and North Vietnamese Army, but the request was denied because Kennedy’s advisors felt the political and public fallout would be too high a cost to an administration already embroiled in an unpopular war. Less than a year later however, as death tolls clicked along at an astonishing pace, an increasingly desperate President Kennedy, known already for a technocratic management style inflated by idealism (Martini, 2012, p. 19), approved the controversial program. In November of 1962, the first sortie of planes flew low above the triple-thick jungle of South Vietnam, retrofitted to spray a super-concentrated mixture of 2,4-D and 2,4,5-T, later nicknamed “Agent Orange.” Operation Trail Dust was the name of the overall crop destruction program, of which Operation Ranch Hand, the most famous of the campaigns, was a part. Both call signs were macabre metaphors in and of themselves.

From the start, the use of weaponized herbicides was a divisive tactic that divided the Department of State. Opponents argued that Operation Ranch Hand and other crop destruction campaigns violated the Geneva Protocol, was an unproven technology, and might have far-reaching consequences on human health. But, in a war where nearly 60,000 U.S. soldiers and more than 2 million Vietnamese soldiers and civilians would eventually die, arguments against trying new, potentially life-saving technologies were discredited as anti-American (Berman, 2012, p. 2012), and so opposition was overruled. All told, nearly 20,000 crop destruction sorties were flown in the decade between 1962 and 1972.

Typically, “ranch handers”—the nickname given the elite squadron of pilots who flew the first missions—practiced mixing Agent Orange with kerosene or diesel, spraying the amalgam from C123 transport planes and then setting the ground ablaze with firebombs of napalm. This was not the only way Agent Orange was deployed into the environment though. It
was also sprayed from retrofitted helicopters, riverboats, backpacks, trucks, and used in and around U.S. and joint operation bases. By the end of Operation Ranch Hand, some 4.7 million acres of Vietnamese jungle were defoliated and nearly one half million acres of crops destroyed (Lewy, 1978, p. 258). Because of such widespread and indiscriminate use, there are no official estimates of the total number of veterans or Vietnamese exposed to the chemical. Exposure was not straight forward, as wind conditions, quantity of product, and absorption rates were but a few factors that had to be considered and the data making such calculations possible was often unavailable. This is one of the many reasons why the idea of Agent Orange causing significant medical illness is such a contested idea. From the beginning, military and chemical manufacturing companies claimed that—even at concentrated levels—the amount or duration of chemical exposure would not have been high or long enough to cause serious health complications. But such claims—from both parties (veterans and chemical companies) would remain largely that—claims—because military records could not provide the kind of data needed to model exposure rates or predict potential health consequences or environmental contamination levels. The closest thing to an official exposure record available today is the number of claims filed through the Veteran’s Administration for relief of diseases associated with Agent Orange exposure, but this only covers Americans who might have been exposed. That number is roughly 230,000 and the remuneration program enacted in 1991 has cost the U.S. government more than $3.6 billion dollars to date (The U.S. Department of Veterans Affairs, 2015).

This project is a useful response to the political problem of Agent Orange because it attempts to systematize the study of descriptive terms that politicians, advocates, victims and technical communicators use to talk about controversial military and scientific practice. Reality

---

2 Like much of the scientific and military discourse on Agent Orange, sources vary on the exact quantities of defoliant sprayed or the duration soldiers were likely exposed, or even the areas where missions were sprayed. That said, Lewy based his figures on the Military Assistance Command, Vietnam (MACV) history.
is constructed by first describing what we do not yet understand, over time, these descriptions become codified into “reality” or, as Latour would say, “blackboxed.” Identifying and analyzing such linguistic fossils can help us crack them open and overcome the political stalemate that often ensues after reality is created, positions taken, and scientific knowledge hardens into unalterable belief.

**Agent Orange as Scientific Controversy**

Uncertainty, anecdotal evidence and inadequate technical knowledge and scientific practice conspired to make Agent Orange one of the most scientifically and politically embattled issues in modern history. Discussions about Agent Orange are emotionally charged and medically complex, which drives the controversy further.

Agent Orange is the most notorious chemical used during the Vietnam conflict, but it was only one of more than 60 chemical combinations of herbicides, pesticides, fungicides, rodenticides, and riot control agents used to mitigate hostile and unfamiliar environmental conditions, to destroy enemy food supplies, or the protection of jungle cover (Martini, 2012, p. 22). Though some of these chemicals were known entities, many chemicals were used beyond their original intent, or were mixed in experimental combination and were often deployed—just as Carsen accused in *Silent Spring*—without adequate consideration given to the possible human health consequences or the long-term environmental costs.

Part of the confusion characteristic of Agent Orange discourse has to do with the fact that it isn’t Agent Orange that is toxic to humans at all, but the 2378-TCDD-dioxin—a byproduct of its manufacture. Though exposure—how much, how intense, and how often—is one of the key controversies in the Agent Orange debates, it is supposed that veterans were exposed to Agent Orange, and thus dioxin, in multiple ways. Some soldiers handled it on air bases as it was
moved, stored, and used for *Operation Ranch Hand* and other spray missions. Some were exposed to it as they humped through jungles that had been sprayed, or ate local, contaminated food crops, or showered in bomb craters filled with contaminated water. Some carried packs of Agent Orange on their backs to spray as they hiked through dense foliage. Others were exposed to blowback, or residue that saturated cargo planes transporting Agent Orange, or by handling poorly cleaned equipment and nozzles that sprayed it.

By the end of the crop destruction program in 1972, Monsanto was the single largest provider of Agent Orange to the U.S. military, and the levels of TCDD dioxin found in its product were well above the 1 part-per-million EPA-approved levels (Martini, 2012, p. 148). As early as the 1950s and 1960s, it was common scientific knowledge that dioxins were cancer-causing agents, and scientists knew small exposures could have terrible effects on humans (Hites, 2011), though the breadth and depth of those consequences was speculative.

The exact concentrations of 2378-TCDD in the Agent Orange used in Vietnam is not a matter of public record, but the most reliable estimates average about 3 parts-per-million. Given the total volume of herbicide sprayed, it follows that 150kg of 2378-TCDD could have been added to the environment of southern Vietnam (Hites 2). It is also still a mystery when, exactly, chemical companies discovered their products were contaminated, but there is evidence that once they knew, they attempted to prevent U.S. military and government leaders from finding out.³ This fact, along with a general mistrust in big business, as well as the use of a range of chemicals mixed and used indiscriminately, are some of the conditions that swathed Agent Orange in a

---

³ On March 23, 1965 senior leadership from the largest manufacturers of Agent Orange: Diamond Alkali, Hercules, hooker Chemical and Monsanto met at the Midland Country Club in Michigan to discuss the problem of 2,4,5-T toxicity. What exactly was discussed at the meeting is unknown, but an internal memo from Dow’s director of Biochemical Research Lab suggests that they discussed recent experiments that showed animals subjected to heavy amounts of dioxin were significantly and adversely impacted. Though chemical manufacturers clearly knew about the toxic impurities in their products and their consequent biological effects, it is unclear that this information was ever communicated to the Pentagon or the USDA, the two largest clients of herbicides in the 60s (Martini, 2012).
shroud of conspiracy early on.

In the early 1970s, teams of scientists were sent to conduct field tests on Agent Orange’s ecological impacts. These studies and public and scientific outcry led the U.S. government to restrict the use of Agent Orange and other herbicides in 1970. The Department of Defense temporarily suspended the use of Agent Orange late in the year and the last spray run for Operation Ranch Hand was flown in January, 1971. After the discontinuation of Agent Orange in 1972, a large epidemiological study was organized in 1979 to try and associate Agent Orange exposure information with consequent health effects as determined through medical examination. The study focused on veterans in the U.S. Air Force who participated in Operation Ranch Hand and would best serve as a representative sample population to study. Roughly 1000 veterans involved in spray missions and an equal number of those who were not were enrolled in the study and evaluated every five years. Early results found few statistically significant differences in health outcomes from either group, but later it was shown that the study was troubled with methodological problems. The original study found no statistically meaningful differences in diseases among the control group versus the ranch-handers. Later, when the methodological problems surfaced, the botched study led veterans groups and the media to characterize the Veteran’s Administration as incompetent and the Centers for Disease Control (CDC) as being in cahoots with the federal government, intentionally mismanaging the study, and ruining the credibility of science. Though future studies regulated by other agencies attempted to study the effects of Agent Orange with greater scientific rigor, all Agent Orange studies with which the government was involved seemed haunted by these early, bungled efforts.

As the science of Agent Orange matured, calculations of dioxin exposure and health consequence were then based on measured tissue or blood concentrations and at last, significant
health differences began to emerge. Despite the progress, in 2006, the 27-year study, which cost the U.S. government upwards of $140 million dollars, was axed over significant protest from veterans groups and the scientific community. Though the study was scrapped, the most recent assessment of the Ranch Hand data, demonstrated “sufficient evidence of an association” between herbicide exposure and incidence of soft-tissue sarcoma, non-Hodgkin’s lymphoma, Hodgkin’s disease, chronic lymphocytic leukemia, and chloracne” (National Academy Press, 1994, p. 5). As a consequence of this research, veterans won what is called “presumptive exposure,” which affords free treatment for all medically related illnesses associated with exposure to Agent Orange-dioxin.

Part of the scientific controversy of Agent Orange—as opposed to the public or political controversies as represented by remuneration or environmentalism—has to do with how it was classified and talked about. Diseases that came to be associated with exposure are recognizable biomedical classes occurring among populations also not exposed to dioxin. Agent Orange however, as a disease category, remains in a class almost all its own. It is even unlike “Gulf War Syndrome,” a medical classification for an unexplained set of symptoms associated with military service in the Gulf War. Most experts on Agent Orange don’t talk about “Agent Orange disease” or even “Agent Orange syndrome”; they describe it as an “exposure-related illnesses.” But most non-credentialed experts—victims, primarily—will say they suffer from “Agent Orange” or they “got the Agent Orange.” In other words, the association between disease and the chemical that caused it has no bearing on one’s diagnoses or treatment (Uesugi, 2011, p. 8), making Agent Orange confusing and difficult to talk about with scientific, medical, or technical accuracy. This inaccuracy contributes to the scientific uncertainty that infuses discourses about Agent Orange and complicates medical, political, and humanitarian response.
Agent Orange’s chemical complexity is one reason why discourse on Agent Orange is so troubled, but it is equally complicated semantically, which is why treating Agent Orange only as a scientific, medical, or material object tells but half the narrative. Agent Orange works on a linguistic level—specifically a metaphorical level—that is just as important to shaping policy and practice as is the scientific and medical evidence so often entered into Congressional hearings where these policies are written and practices observed. And yet, a systematic study of the rhetorics of Agent Orange is missing from the literature, leaving Agent Orange texts an untapped discourse ready for deep text mining.

In studying the rhetorics of Agent Orange, it is difficult to miss the uncertainty embedded in the discourse. Part of this uncertainty was cultivated by the fact that Agent Orange-dioxin does not affect all bodies in exactly the same way; there is a range of illness attributed to exposure. Much of the testimony delivered to Congress by veterans was anecdotal, and was easily diminished or dismissed in lieu of expert testimony. But even expert evidence delivered during deliberations was variable and didn’t perfectly match the disease categories science said exposure caused. Concurrently, Congress was skeptical of veteran malingering and this suspicion was exacerbated by new psychological disease classifications like post-traumatic-stress-disorder, which often presented symptoms that looked an awful lot like those purportedly caused by Agent Orange.

Some in Congress were worried that Agent Orange was simply another manifestation of the crazed Vietnam veteran, an image that, by the time of the Congressional science wars in the 1980s, had taken root in the American psyche and visualized in movies like Rambo: First Blood and Apocalypse Now. As Martini explained, though scientists learned a significant amount about how dioxin worked on and in the body during the Congressional inquiries of the 80s, and though
they developed better tools with which to measure ever more discrete units of dioxin, these discoveries “did not lead to greater ability to predict with accuracy what the effects of that exposure would be, particularly in human health” (2012, p. 7). Therefore, uncertainty, inadequate diagnostic and lab methods, and public opinion all shaped the discourse on Agent Orange.

When science showed that Agent Orange/TCDD-dioxin indeed had significant human health impacts on U.S. veterans—despite multiple presentations—Vietnamese civilians and former soldiers began what would become a 45-year petition for recognition, apology, and recompense. Not long after Operation Ranch Hand launched, Vietnamese women began complaining that their children were born with terrible deformities, but, like veteran accounts of exposure, these narratives were quashed, chalked up to NVA propaganda. Once U.S. soldiers returned from war and had children of their own though, they too told similar tales. Four decades on, Vietnamese families continue to claim children are born with significant birth defects attributed to dioxin exposure, and many U.S. veteran children, now grown, have organized claim they too suffer from exposure-related illnesses passed on through their genetic code. Such claims remain largely unstudied, though dioxin has been shown to be a teratogen and mutagen in animal studies of the chemical.

Concurrent to the “Ranch Hand” study, several additional studies were conducted that sought to identify possible patterns in data that might prove causation or even correlative links. The Center for Disease Control (CDC), the Australian government, and the Vietnamese government all conducted independent studies—all which yielded mixed results—which further compounded the controversy already embattling Agent Orange science. While some studies found parents exposed to Agent Orange had higher than normal rates of Neural Tube Birth defects like Spina Bifida and Anencephaly, the relatively low incidence of birth defects per the
large sample size resulted in weak correlations. To complicate matters, the publication of significant but biased findings by Vietnamese researchers fueled the belief that the Vietnamese, like the U.S. veteran, was simply malingering. Questions of exposure, risk, and the limits of science would be debated in courtrooms and in Congress through the end of the 20th century. Veterans, scientists, lawyers, corporations and politicians were “each pitted against the other in a complicated and compelling confluence of anecdote, metaphor, scientific knowledge and state authority” (Martini, 2012, p. 149).

Agent Orange in the US News Media

Agent Orange was not always known by this name. For a significant period of its early history, it was known by the somewhat innocuous name, “Herbicide Orange.” The defoliant itself gave off a reddish-brown hue, but the name “Agent Orange” derived from the color of the giant 55-gallon drums in which it was stored. Though war correspondents were embedded with several units in Vietnam, little was reported on the largely secret Ranch Hand crop destruction missions until the late 60s, and consequently, in an analysis of newspaper reports of the time, Agent Orange shows up under different names like “Super Orange,” “Herbicide Orange,” and “Super Herbicide,” or merely one of the “rainbow chemicals,” another of the war’s many macabre metaphors and allusions.

There was very little mention or discussion of Agent Orange in the popular American press before 1968 when the first occurrence of the term appears in the New York Times in an article titled, “Pentagon Backs Use of Chemicals.” The term is found even later in the more conservative Wall Street Journal, appearing first in a 1979 article, “US Plans to Assess the Possible Ill Effects of Dioxin on Humans.” Despite its late entry into the discourse on Agent

---

4 For a complete description of studies on Agent Orange and birth defects, see the meta-analysis by Anh Duc Ngo and colleagues, “Association between Agent Orange and birth defects: systematic review and meta-analysis” (2006).
Orange, the media had and continues to have tremendous impact on the ways Agent Orange and its attendant issues are thought about, framed and enacted. In fact, it is highly unlikely that veterans would have won “presumptive exposure” (explained in detail later in this work) without media interest, and the public boost that Paul Ruetershan lent the cause in 1979 when he addressed the nation on The Today Show. In a stunningly persuasive and impromptu sound bite, Ruetershan said he, “died in Vietnam and didn’t even know it” (Wilcox, 1989), alluding to exposure to Agent Orange to which he attributed his rare form of cancer. This cancer killed him not long after the Today Show appearance.

The 28-year-old Ruetershan leveraged the power of the mass media to raise awareness of his issue after his claim for medical relief to the VA was denied. His only legal recourse thereafter was to file a personal injury suit against the main chemical manufacturers who produced Agent Orange, and the feature on the Today Show, which detailed the main points of that case, launched what would become the largest class action lawsuit in American history (Jovanovitch, 1988, p. WC25).

Reutershan was a sympathetic character with demonstrable credibility. He was young, good looking, smart, articulate, and had volunteered to fight in the war. Bureau chiefs, many of whom served in the war and perhaps recognized themselves in the young Ruetershan, began filing editorials about Agent Orange, or accepted editorials that gave voice to a public who grew increasingly outraged over what was perceived to be the systematic exploitation of the nation’s veterans and a mass poisoning.

In March of the same year, a CBS affiliate in Chicago aired the documentary “Agent Orange: Vietnam’s Deadly Fog.” It is difficult to understate the impact this documentary, written and anchored by Bill Kurtis, had on the trajectory of Agent Orange policy debates (Wilcox,
1989), but its release reframed Agent Orange from a minor, easily dismissed local veteran’s issue, to an issue of national importance and global consequence. Within a few weeks of a private screening, Congress requested that the Government Accountability Office (GAO) explore the documentary’s main claims about dioxin exposure and disease. Additionally, in response to the growing national media attention focused on the issue of Agent Orange, the Veteran’s Administration created the Agent Orange Working Group, which would eventually help to develop the “presumptive exposure” policy that remunerated exposed veterans (The Institute of Medicine, 1994, p. 46).

**Agent Orange as Legal Controversy**

Paul Reutershan died shortly after his interview aired, but before his death, he had organized and inspired other veterans to participate in the Agent Orange Product Liability Litigation class action lawsuit. The suit was eventually brought before the Second Circuit Court of Appeals in New York with Judge Jack Weinstein presiding.

Weinstein felt the veterans’ case was legally weak, but also knew that a disabled or dying veteran or his crippled child would be extremely persuasive with jurors. Calculating the cost of losing the case, chemical manufacturers agreed to a settlement of $180 million—the largest settlement in U.S. history at the time. Though the settlement was technically a “win” for veterans, it was the chemical manufactures that proved victorious in the end. Not only did chemical companies make tens of millions more off the production of Agent Orange than they had to pay out, but companies negotiated into the settlement a clause that precluded admitting responsibility and prohibiting any future liability brought against them from U.S. or foreign claimants. None of the chemical companies that manufactured the product would ever have to admit their culpability, and the settlement effectively inoculated them against all future liability
for their role in what has been called by some one of the largest ecocides and mass poisonings in United States history (Blumenthal A1).  

Vietnamese nationals have since attempted to sue chemical manufacturers for a range of health issues and for the lasting ecological damage wrought by the products they manufactured, but in 2005, Weinstein dismissed all associated lawsuits, finding that Operation Ranch Hand did not violate international law, and therefore foreign nationals did not have the standing to sue. In March 2009, the U.S. Supreme Court denied Vietnamese plaintiffs’ applications to hear an appeal, permanently ending this or any future lawsuits brought about by foreign nationals in the case of Agent Orange. As David Zeigler woefully notes, it is “perhaps the only aspect of the complex legacy of Agent Orange that has ended with some degree of decisiveness” (Zeigler, 2011, p. 13).

American Skepticism & Vietnamese Acceptance of Agent Orange

Agent Orange is an ongoing global, scientific controversy. Not unlike the controversy surrounding climate change, debates about Agent Orange have “deniers” who believe Agent Orange is not the cause of veteran and civilian maladies so often attributed to it. The most famous of these deniers is probably Alvin Young.

Alvin Young spent his entire career with the U.S. Air Force experimenting and documenting how Agent Orange worked in and on the environment, and despite the contempt his name inspires among veterans groups (he is characterized by many as a turncoat and lackey) few can question the breadth, depth or value of the data he collected during his military career as an Air Force scientist.

---

5 For a fairly thorough explication of the court records and final settlement amounts, see Peter Sill’s “Toxic War: the Story of Agent Orange.” The strength of the book is his access to trial documents, transcripts and files, many of which have been lost to history and never digitized or archived. However, his selected bibliography only includes 24 works, with the most recent works published in 1993 before much of the recent and best scholarship on Agent Orange was written.

Young argued in the 1980s that there was no reliable epidemiological data to suggest Agent Orange was associated with any long-term health effects in humans other than chloracne (Agent Orange: At the Crossroads of Science and Social Concern). While this was mostly true at the time, it was also a fallacious argument. Several attempts by leading research and scientific organizations had been made to collect epidemiological data that might result in valid incidence and mortality results, but with little success. Military records charting troop movements were difficult to come by, samples of soil were contaminated, and population samples could not account for those who had already died or were missing. Young was right to say there was a dearth of epidemiological data, but that didn’t mean the anecdotal evidence of disease was unfounded; it simply meant that by the standards of epidemiological science, a valid study could not be produced because of the limitations in the amount, quality and accuracy of available data.

Young was not hated for this. He was reviled because early and later works suggested a radical alternative hypothesis to medical explanations of veteran claims Young hypothesized that Agent Orange was not a physical disease, but a psychological one, a disease of the psyche that could be “caught” through narratives circulated and redistributed in the media. Though a wildly unpopular position among veterans and other scientists, Young was not alone in his skepticism of veteran and civilian claims. As Edward Martini wrote, Agent Orange might be reasonably considered as an experience “forged at the intersections of memory, media, trauma and legitimate disease concerns” (Martini, 2012, p. 157). At the time Young was writing, there wasn’t a name for what has since been called in network theory circles “social contagion.” Theories of social contagion measure the influence of one’s peer/social network on disease classes that were once conceived of as entirely biochemical—depression, obesity, and alcoholism, in example (Christakis & Fowler, 2012).
There are many theories of collective behavior and social contagion that attempt to explain how networks of people, behaviors, or ideas influence and “infect” others’ attitudes, beliefs, values, and actions. Many of these theories come from structural theory, symbolic interaction theory, and network analysis, all of which emphasize a different aspect of the social construction process as explanation for how meaning and belief are transferred.

Theories of contagion from social network analysis are simple enough. Communication networks serve as the mechanism through which people are exposed to attitudinal messages, information or the behavior of others (Burt, 1980; Contractor & Eisenberg, 1990). Exposure increases the likelihood that network members will develop similar beliefs, assumptions and attitudes to those within their networks (Carley, 1991; Carley & Kaufer, 1993). There are many factors that affect the influence of a group or network: multiplexity, strength of a connection or tie, and frequency of contact, in example, all of which influence the degree to which group member beliefs are influenced (Erickson, 1988).

The applications of contagion theory are broad, applicable to organizations, governments and interest groups, and though there is strong evidence to suggest that contagion theories are legitimate social phenomenon and were very likely at work in the exchange, transmission, and hardening of beliefs around Agent Orange, there has never been a network study conducted of Agent Orange veteran networks and belief. Partly this is because network analysis requires a specific type of data gleaned best from surveys, not unstructured texts. This data does not exist or has been, thus far, inaccessible to most researchers outside the military. The project detailed in the final chapter of this thesis is an important first step in collecting the right kind of data to conduct network analysis and test hypotheses linked to theories of social contagion.
On the other side of the scientific controversy are those who have worked tirelessly to overcome the doubt embedded in Agent Orange discourse by continuing to push for or conduct new medical and soil studies that show links between exposure and disease or exposure and contamination. These are the people who carry on the early awareness projects started by Paul Reutershan and Rachel Carsen, lending to narrative and anecdote the gravitas of calculation and quantification.

As recently as July, 2014, *Time* magazine featured a new study from the University of California, Davis, which strengthen links between neurodevelopmental disorders—particularly autism—with gestational pesticide exposure, particularly organophosphates (Shelton, et al., 2014), and the work of the Aspen Institute’s Agent Orange in Vietnam program continues to address the health and environmental impacts of herbicides sprayed during the war, but the most recent movement of political significance was the December 2014 Consolidated and Further Continuing Appropriations Act signed by President Barak Obama (The Aspen Institute, 2011).

Whether it’s dealing with the social-psychological aspects of Agent Orange, its environmental consequences, or its medical conundrums, Agent Orange is *not* an issue that ended with the war in Vietnam. As U.S. Specialist in Asian Affairs, Michael F. Martin said in a report to members of Congress, “The one significant legacy that remains unsolved is the damage that Agent Orange and dioxin have done to the people and environment of Vietnam” (2012, p. 2), itself a separate controversy. Agent Orange is embattled by controversy, not just between the U.S. government and its veterans, but also between the U.S. and Vietnamese governments. In an effort to address that legacy, President Obama signed into law the Consolidated and Further Continuing Appropriations Act, 2015 (H.R. 83) in December 2014. The act appropriates funding
to address Agent Orange/dioxin remediation and remuneration issues, but more importantly, it guides the State Department and USAID in *how* to spend the funds.

From 2007 until present day, Congress had approved roughly $100 million to remediate dioxin “hot spots” at and around former Air Force bases and $31 million for disability assistance, but defining what counted as an Agent Orange-related disability proved difficult and contentious. Unlike the United States Congress, who spent years funding studies they hoped would prove the causal relationships between exposure and disease beyond doubt, the Vietnamese government sought no such proof. Most Vietnamese civilian and medical communities working with Agent Orange patients/victims have never questioned the relationship between Agent Orange and the diseases suffered; they’ve simply presumed the link Congress felt obliged to prove. The combination of those presumptions and a lack of credible studies on Vietnamese populations however, led to over-reporting Agent Orange-related cases and environmental ruin. In turn, this strained the credulity of the Vietnamese government and the patience of the U.S. Government—a key source of aid dollars. The United States government is not blameless, as it has a long history of denying non-Western scientific studies, claiming standards of science were incommensurable. For years, this tension colored bilateral trade negotiations and stalled significant progress on the ground in Vietnam toward remediating “hot spots” or offering aid to the worst and most obviously affected civilians.

The language of the 2014 Appropriations Act however, offers a way around this 40-year-old impasse by sharply focusing on the use of funding for people with *specific* kinds of disabilities who live in or near areas where Agent Orange was sprayed or stored. The language specifically stipulates that the funds should be prioritized to assist those with “severe upper or lower body mobility impairment or disability and/or cognitive or developmental disabilities”
(H.R. 83). This word—disability—is particularly important because it may be one point of evidence in this study that shows how consistent framing over a period time can influence and change policy and practice. In the results chapter of this work, the term disability and the disabled are identified as significant words and frames in the Vietnamese discourse and have clearly had some rhetorical effect. These words were not always used in the discourse though, and emerged and were responded to in the last 10 years or so, replacing, the more negatively charged term “victim.”

The bill’s stipulations would send aid to the most needy, and to those most likely exposed to Agent Orange-dioxin. The continued scientific uncertainty of Agent Orange has forced this kind of “work around” solution, which, in some ways mirrors the political and legal decision Congress made in the 1980s to presume U.S. veterans were exposed and to cover medical costs related to that exposure. As time passed, science has increasingly impressed upon the American legal, political, and veteran communities an unavoidable responsibility, but a responsibility that will likely always be hedged by political caution.

For decades, the Vietnamese have positioned themselves as “victims”—as this study demonstrates—and have, at times, conflated all disability with exposure-related disabilities. Using more precise language to specify how, where, and to whom aid dollars will go, offers the Vietnamese government an opportunity to save face, while securing much-needed aid, but it also affords the U.S. a continued opportunity to resist accepting blame.

One of the reasons Agent Orange/dioxin is such a confounding scientific issue is because dioxin affects bodies differently. For U.S. veterans, who are now approaching their seventies (an age where the body naturally begins to deteriorate), dioxin poisoning has often been linked to lymphoma, specific types of cancer, or porphyria (The United States Veterans Administration).
Comparatively, the Vietnamese, who have been living near polluted sites and consuming dioxin-contaminated food for 45 years, suffer medical issues that look quite different (Science Daily, 2003). New generations of children are born with significant deformities and developmental delays that look nothing like the diseases that plague U.S. veterans, suggesting that dioxin not only manifests differently in bodies, but manifests in particularly different ways in young and adult populations. To date, no valid studies have been conducted on the child populations in Vietnam who, presumably, suffer dioxin-related illness, disease, and deformity.

What has been studied however, is the soil. Recent soil studies conducted by Hatfield Consultants left little room for question that “hot spots” near former bases were contaminating food supplies and would continue to threaten human health if not remediated (Hatfield Consultants, 2009, p. 2009). Interestingly, far more work has been done on remediating the environment than addressing the human health impacts, and partly because it’s easier to quantify the effects of dioxin on the environment. Most industrialized countries recommend dioxin levels at or below 1,000 ppt (parts per trillion) in soil and 100 ppt in sediment. Most soil from industrialized nations holds less than 12 ppt, but in Vietnam, scientists with Hatfield Consultants found dioxin levels of up to 365,000 ppt at the former Da Nang Air Force base.

In 2012, USAID and the Government of Vietnam launched one of the largest cooperative environmental remediation projects between two former wartime enemies. The Environmental Remediation of Dioxin Contamination at the Da Nang Airport Project will treat approximately 45,000 cubic meters of dioxin-contaminated soil and sediment through traditional means of remediation, similar to that used at the Johnson Atoll to get rid of the Agent Orange stockpile left over after the war. Eventually, the soil will be heated to 335 degrees Celsius (635 F), which is the maximum temperature it takes to incinerate dioxin. USAID contractors completed Phase 2
excavation activities ahead of schedule in early 2015 and the entire project is expected to be complete by 2016, with several more to follow in other “hot spots” throughout the country.

While there is progress on environmental remediation in Vietnam, back in America, the list of illnesses presumably related to Agent Orange exposure, and therefore covered under the 1991 Agent Orange Act, continues to grow, as do the rhetorical challenges faced by veterans groups who seek protection under the law for themselves and their families (The U.S. Department of Veterans Affairs, 2015).

It took the U.S. Congress almost 20 years to acknowledge and accept that science could only answer part of the Agent Orange question. The other half of the Agent Orange controversy was not a scientific question at all, but a political and ethical one: what is a government’s moral obligation to the men and women it sends off to war who come back sick and diseased? Debating these questions about the greater good may be philosophical, but as a republic we arrive at our answers through dialectic, and thus to study how we talk about, describe, and frame the controversies embedded in Agent Orange discourse is an important project of public rhetoric.

Modern discourse about Agent Orange is increasingly generated or archived in digital public spaces that have also gradually gained acceptance as legitimate research sites (The Aspen Institute, 2012). Twitter, investigative journalism websites, Facebook Groups, and blogs, in example, offer new sites in which to study old problems of interest to rhetoricians like identity creation, ethos building, argument, and persuasion. Getting at, organizing, and doing something with the volumes of unstructured text contained across the web presents vexing practical and methodological challenges for discourse analysts scrutinizing discourse in the age of “big data.” As I suggests later in this thesis, it is critical to first systematize ways of identifying and mapping key frames across large corpora of texts before we can do the constructive rhetorical work of
suggesting new frames and ways of describing. To demonstrate the power of frames, I later discuss how, once Agent Orange remediation and remuneration rhetoric was reframed in humanitarian terms instead of scientific uncertainty, far more positive rhetorical possibilities and potentials for social change emerged. Finding, tracing, and comparing how different stakeholders have framed the central arguments in the discourse automatically, is the heart of this work. It is also its greatest challenge and is the key feature of the methods chapter.

First, I must situate this work in the field of Rhetoric, and offer a theoretical lens, which requires borrowing from Computer Science, Cognitive Science, and Sociology. I argue that when a social issue is particularly complicated, humans will seek to reduce that complexity by activating frames, or ways of communicating about and understanding reality. Framing involves the construction, interpretation and judgment of social phenomenon. The media, political leaders, and other actors often frame issues, objects and ideas in ways that shape our values, attitudes, and beliefs about them. Frames are constructed, relayed, and reinforced visually and verbally (Goffman, 1974). Here, I am particularly interested in comparing how different discourse communities describe important keywords related to Agent Orange, words like chemical, environment, and remuneration. I was curious to see if the ways we described, talked about and framed these key terms were positive, neutral or negative, and whether what we talked about followed patterns of statistical significance.

**Networked Rhetorics**

Quantifying what was once only described qualitatively affords our discipline a greater level of authority and legitimacy, as well as the excitement of pushing out in new directions and claiming disciplinary territory. We not only need quantitative methods to help manage and analyze large data sets, but we need to understand quantitative methods because we can then
appeal to dominant topoi, values, and warrants in larger decision-making and policy communities. Quantitative methods need not replace more traditional qualitative analysis, but it can make our analyses more compelling to audiences with whom we have not always been persuasive, and with potential collaborators to whom we have not always had access.

As our rhetorical practices change, they create new communication challenges, which in turn will demand a recasting of old rhetorical theories and a generation of new theories capable of explaining rhetoric’s revised range, responsibility, and role. These changes have prompted modern scholars like Hawk (2004) and Ridolfo & DeVoss (2009) to adapt language and theory from other fields and scholars like Barabasi (2014), Freeman (1977), Borgatti (1998), and Carley (Van Holt, Johnson, Carley, Brinkley, & Diesner, 2013), in order to shape a new rhetoric capable of guiding public advocacy and deliberation in the modern, digital agora.

Mixed methods approaches like those explored in this dissertation are increasingly used to study a range of issues traditionally important to rhetoricians and technical communicators. The U.S. military, the Ford Foundation, academia and industry have used mixed-methods text analysis to study everything from the psychological grip terrorist ideologies have on mainstream (McCullough, 2009); to the influence of social documentaries on behavior (Graduate School of Library and Information Science at the University of Chicago, 2013); competitive advantage in the marketplace (Grimes, Text Content Analytics 2011: User Perspectives on Solutions and Providers, 2011) and decision-making practices in technical communications (Thayer, Evans, McBride, Queen, & Spyridakis, 2007).

As rhetoricians, we can and should, as Collin Brooks suggests, begin to think about how rhetorical elements are embedded in broader ecologies like networks. We should begin to think about network structures and those measures important to networks and embedded in big data
like tipping points and rhetorical velocity (Ridolfo & DeVoss, 2009). We should also study and master new and better ways to visualize and render our findings so that our stories are both more quantitatively accurate and visually persuasive (Borkin, et al., 2013).

Mixed methods approaches and tools allow the technical communicator to trace, plot, compare, analyze and visualize components of discourse, compare actors, positions, frames, and topoi. Rhetoric is always at work in the network of relationships, news, texts, and the ideas that form and reform culture. Networks are simply a new visual means by which to mark and study the structure of those relationships. From Latour’s actor network theory to the distant reading of Franco Moretti’s Lit Lab, which uses both quantitative and qualitative methods to tackle literary problems, mixed methods and improvements in the power of tools available provide social scientists and rhetoricians new methodological possibilities. By operationalizing knowledge, researchers in Rhetoric and Technical Communications can rework our social-scientific practices in ways that transform insight into social and political action. Therefore, we should invite critique, replication, and verification as a means to enhance the scholarship conducted in our field.

Ultimately this is a thesis about methods, metaphor, and frames—about the ways in which different stakeholders have described and framed keywords relevant to the Agent Orange debates, and new ways of collecting and analyzing those units of analyses from within discourse that is increasingly archived digitally, or available in real-time through multi- and social media sites. Collecting, counting, and comparing descriptive terms only gets one so far, which is why I advocate a mixed methods approach that leverages the strengths of quantitative analysis to find and count with the power of critical discourse analysis to interpret and explain social behavior. The computational tools used to extract and tabulate phrases embedded in texts are new and
experimental and are thus of great relevance and value to practicing technical communicators. Like others (Williford & Henry, 2012), I propose those in the Digital Humanities should also master these new tools because they allow us to do our necessary critical work faster, across wider sets of text, and with greater validity. Like other current scientific controversies, including the Monsanto Protection Act, genomic patenting, and global warming, the legacy of Agent Orange serves to underscore the importance and relevance of enacting humane science and technology policies that do not privilege corporate interests over human health or dominion over cooperation.
CHAPTER 2: THEORY & DEFINITIONS

Metaphors, Descriptive Phrases, and Why They Matter

In Tim O’Brien’s now famous book about the Vietnam War, *The Things They Carried*, he wrote, “They carried the sky. The whole atmosphere, they carried it, the humidity, the monsoons, the stink of fungus and decay, all of it, they carried gravity” (1990, p. 39). O’Brien is using allusion and metaphor to explain what it was like to be an American soldier during the Vietnam War.

The term metaphor is Greek, deriving from the word *meta* meaning *with* or *after* and *pherein*, meaning to *bear* or *carry*. Yoking these terms together we understand that metaphors carry meaning from one word, concept, or idea, to another. Though metaphoric language works through its powers of transference, the translation is never exact, meaning that metaphors illuminate some aspects of similarity while diminishing others.

Metaphors are powerful for many reasons, not least of which because they carry an emotional charge, which is what has made them poet and writers’ favorite instrument for centuries. Who among us is untried by Donne’s famous metaphor describing his lover—“She is all states, and all princes, I”—or the disgust at seeing 14 year-old Emmett Till in his open casket in 1955? Donne’s verbal metaphor and the visual metaphor that Till’s open casket became use different means to arrive at a similar rhetorical end: to provoke the receiver to see a situation, object, or idea in a particularly constructed way, to advocate.

The “hoodie” not only has literal, evidentiary value in that the young black man Treyvon Martin was wearing one when he was shot to death by a neighbor patrolling the neighborhood,
but it also holds iconic power, and has since become a visual metaphor symbolizing resistance to the criminalization of black, American youth. Similarly, the most famous photograph of Agent Orange—the image of a C123 plane spraying white trails of herbicide over pristine jungle canopy—has become its own symbol and is co-opted by environmental groups and veteran organizations as an emblem representing the dangers of untested, corporate-funded science put to military ends. The Monsanto typeface, colored in orange, is also used to similar effect on protest posters that decry the use of genetically modified foods today. The awful photographs of children with sunken heads—presumed victims of dioxin exposure—are visual metaphors, and their propagation and distribution across the internet have done more to raise the public profile of Agent Orange remediation and health advocacy efforts than any other single textual message alone.

Metaphors—both visual and verbal—are ubiquitous because we think in metaphors. In fact, cognitive scientist George Lakoff argues that humans can only abstractly reason by way of metaphor, comparing how abstract concept A is like abstract concept B, or how the spiritual weight of war is as heavy as gravity (O’Brien, 1990). “Metaphorical thinking half discovers and half invents the likenesses it describes” (Geary, 2011, p. 9). But, it is descriptive language’s very inventiveness that can become problematic. As studies have shown, how we describe or metaphorized objects and subjects subtly shape our values, attitudes, and beliefs about them and the kinds of treatment they deserve (Rose & Baumgartner, 2013); (Marshall & Gerstl-Pepin, 2004); (Maibach, Nisbet, Baldwin, Akerlof, & Diao, 2010).

In the Poetics, Aristotle defined metaphor as “giving the thing a name that belongs to something else,” a fair but incomplete definition that only reflects metaphor’s linguistic aspects, not its cognitive or rhetorical characteristics (p. 1457b). I prefer Jonathan Charteris-Black’s
definition of metaphor as, “a linguistic representation that results from the shift in the use of a word or phrase from the context or domain in which it is expected to occur, thereby causing semantic tension” (Charteris-Black, 2004, p. 21). Take the word “grunt”—a word with many denotative meanings, and a word that carries a particularly electric charge when it is applied in an unexpected way outside its usual context or denotative meaning of guttural sound-making.

In Java scripting language, a grunt is a task runner; in biology it’s the process of calling worms from the ground; but in military parlance it’s an infantryman. In popular culture, grunts are often portrayed as dim-witted, even expendable. To characterize a soldier as a grunt doesn’t conform to the classic metaphorical equation of X is like Y or X is Y, but it serves a similar function in that it describes an abstract idea (e.g. the conditions and expectations for infantry) in terms of more common experiences and understanding (e.g. the life of a worm). What seems to be an innocent linguistic choice, when applied creatively and outside literal meaning, suddenly becomes saturated with judgment, and as history has shown, how societies judge subjects often affects their treatment of them.

Metaphoric or descriptive language does not merely affect how those with power wield that power, but our choice of language seems to affect our very experience of life. A study conducted by the ESRC Corpus Approaches to Social Science Center investigated the use of metaphor in the experience of end of life care delivered in the United Kingdom (Semino, Andrew, Koller, & Rayson, 2013). The study involved the systematic analysis of metaphoric language found in a corpus of 1.5 million words aggregated from interview and online forum texts composed by different stakeholder groups of patients, unpaid family caregivers, and healthcare professionals. Using a semantic annotation and calculation tool called WMatrix, the
team of investigators found that how stakeholders described key concepts and treatment options significantly impacted their end of life experiences.

A “good death” was often conceptualized and talked about as a journey as opposed to a battle, thus signaling the patient’s receptivity and openness to their impending transition. This subtle linguistic choice had very real material consequences, as patients perceived greater quality of care and remarked that they were glad to have had more time with family and friends. Conversely, a “bad death” was one conceptualized as a battle, or one that enacted other war metaphors. These metaphors held implicit directives for patients to “go out swinging” or “fight to the bitter end.” In the qualitative interviews conducted after a patient’s death, caregivers and family members remarked that they felt robbed of time and touch, impinged by a constant revolution of tests, wires, charts, and scans—all the necessary armor one needs to not go gently into that good night, as Dylan Thomas wrote. The overall goal of the end of life metaphor project was not only to improve communication between patients and caregivers, but to also improve end of life experiences for the dying. As a result, the British government has now stripped its marketing and information material of all metaphoric references to war.

I am arguing, as others have before me (Geary, 2011); (Lakoff & Johnson, 2003); (Cameron & Maslen, 2010) that metaphors are more than linguistic flourishes, and studying ways of describing keywords and concepts embedded in a discourse is important because people use unconscious, emotion and frame-driven forms of reasoning when considering, enacting, or responding to public policies.

As with most any unit of language, there are multiple kinds of metaphors. Conventional metaphors are words or phrases so frequently used in discourse they are almost unrecognized as a metaphor at all. The body of an essay, the hands of the clock, even, the horrors of war, are all
conventional metaphors. Conventional metaphors are so commonly used they have lost much of their semantic tension and emotional charge. Conversely, novel metaphors are either new or haven’t been widely used in discourse and thus retain their semantic tension and rhetorical power. If semantic tension is a bell curve, novel metaphors are at the top end where few others are in emotive and persuasive power, so they tend to be sticky, meaning they are memorable; they seem to reverberate in the mind. An example of a novel metaphor might be Thomas Friedman’s “flat world” metaphor, which he used in his book by the same title to describe the economic impacts of globalization (2005) or, in the Agent Orange discourse, describing the legacy of the chemical as a “haunting ghost.” The stickier the metaphor, the quicker it is often adopted and circulated within a discourse community and the harder it is to get rid of or replace.

Novel metaphors carry a certain quality of pleasure, attributing to their salience in a discourse. As Aristotle noticed some 2,500 years ago in On Rhetoric:

Those words are most pleasant which give us new knowledge. Strange words have no meaning for us; common terms we know already. It is metaphor, which gives us most of this pleasure (p. Book III).

Using metaphoric language that characterizes a specific discourse or community’s language is part of what it means to belong to that community—to share its values, vision, and understanding. To that extent, adopting a particular metaphor is not a conscious choice. The “ghosts of war” is often used to describe Agent Orange in diplomatic discussions. This was, perhaps, at one time, a novel and perhaps even effective rhetorical strategy, since distinctive discourse always has rhetorical affects on readers and hearers. But over time, this way of describing Agent Orange has hardened into convention and has lost much of its affectivity.
The cultural, cognitive and semantic process that results in a rhetor selecting one metaphor or way of describing over another is somewhat mysterious and unpredictable. Cognitive semantics asserts that metaphorical expressions are systemically motivated by the conceptual metaphor, a kind of master metaphorical phrase that is, more or less, universally recognized across cultures (Lakoff & Johnson, 2003). Examples of conceptual metaphors include “down is bad,” which then produces metaphorical phrases like “stock market crash” or “war is hell.”

According to Lakoff & Johnson, conceptual metaphors work because we believe a single idea can illustrate multiple other ideas. There may be several metaphorical expressions in which one domain—“fight”—is conceptualized in terms of another—“justice”—but the conceptual metaphor represents the basic idea/image that undergirds the metaphor and can trigger new and novel metaphors in inexhaustible combinations. The basic conceptual metaphor “bad is down” is what makes the metaphor “war is hell” make sense and feel right—at least to a Western, Christian audience.

Agent Orange works on many semantic and metaphorical levels, which is one of the most difficult parts about automating the detection, extraction and comparison of its individual parts. Linguistically, Agent Orange is a proper noun, describing its materiality, but it is often described in terms of the underworld, complete with “creepy moonscapes,” “phantom jets” and “significant ghosts.” There are at least two conceptual metaphors that might give rise to these descriptions of Agent Orange if Lakoff’s conceptual metaphor theory is correct. The first is “existence is having a form” and the second is “treating illness is fighting a war” (Treating Illness is Fighting a War, 1994). The entire discourse on Agent Orange is replete with descriptive and elusory language.
The jungle canopy is a “curtain of weeds” that “sheltered” the enemy and soldiers were “the walking dead.”

Over time, Agent Orange would come to carry the weight of multiple meanings and connotations. To a Vietnamese family who cares for a catastrophically disabled child, Agent Orange is sometimes metaphorized as a karmic return visited on families for ancestral sins (Fox, 2007), but most often it is described as a human misery that must somehow be endured. To American politicians, Agent Orange has been described as a weapon that created military and later political advantage. To politicians, 40 years of Agent Orange remediation negotiations offered the United States significant leverage and made “valuable contributions” to bilateral relations (Secretary, Office of the Press, 2006). To Vietnam veterans, Agent Orange has come to symbolize scientific uncertainty, betrayal, and even conspiracy (Finney, 1969, p. 2); (Wilcox, 1989). I should note here that metaphors and symbols are not entirely interchangeable. A metaphor is a rhetorical device wherein the traits of one thing are attributed to something else. Conversely, symbols are not specific or definitive in their interpretation, and can carry a wide range of ideas. They are also entirely context-dependent. The difficulty, as we shall see in the Chapter on methodology, is that each of these examples of metaphorical language is that they neither conform to the traditional equation for a metaphor, and they require contextual reading to make sense of their connotative rather than denotative meaning. Context depending meaning and polysemy (the capacity for a sign to have multiple meanings) makes machine approaches to collection and analysis very difficult.

There are multiple theories and philosophical investigations into the poetic tradition, and multiple linguistic approaches to metaphor (Underhill, 2013). For this thesis, however, I return to Lakoff:
‘Metaphor’…has come to mean ‘a cross-domain mapping in the conceptual system.’ The term ‘metaphorical expression’ refers to a linguistic expression (a word, phrase or sentence) that is the surface realization of such a cross-domain mapping. (1993, p. 203)

Though this is the definition of metaphor I use for this work, it became evident early in this project that the definition did not always map in ways that were traceable, but it was worth overlooking these limitations because the definition is sound. In other words, the limitations of this work were not in how one defines a metaphor, but in how one converts that definition to a machine-readable format; changing the definition of a metaphor wouldn’t make finding them in texts any easier. I did not actually find any traditionally phrased metaphors, as in A is like B or A is B in the corpus, yet there were dozens of phrases that described and ultimately framed key concepts in the Agent Orange debates. The problems of searching for metaphoric language are attended to in later chapters.

Metaphor and descriptive phrases work by projecting structures from source domains of cultural experience onto abstract target domains. Hence, we understand “human life” (the target domain) in terms of our personal experience of “journeying” (Lakoff & Johnson, Metaphors We Live By, 2003). This function of transfer is how we know Emily Dickinson’s poem ‘Because I could not stop for Death’ is not a story about a day riding through a New England town, even though that is what happens literally. Similarly, our ability to overlay source domains of experience onto abstract target domains of experience is how we know that “Family of Fallen Leaves” is not as a collection of stories about nature, but is a collection of stories about the hardships endured by families with members who suffer dioxin poisoning. Given the obtuse nature of language, one can begin to see the many difficulties of hunting metaphors in dense corpora.
Though descriptive language and metaphor harden over time and thus lose some of their potency, they remain persuasive nonetheless because they continue to circulate in discourse, operating in and on society unconsciously. Conventional metaphors become a kind of linguistic habit, remaining largely unexamined for continued efficacy or accuracy, and this is why they can become problematic (Deignan, 2005, p. 34). It is only through the hard work of identifying, comparing and then reading against dominant metaphor or descriptive language common to a discourse that we can conceptualize and perhaps rearticulate more novel, accurate, and humane ways of describing subjects. Determining a metaphor’s mutability then is important to understanding the political or economic decisions that might be activated by metaphor’s use and reuse (Lakoff, The Contemporary Theory of Metaphor, 1993).

The power of metaphoric language to influence decisions and actions has been demonstrated in several disciplines. In rhetorical studies of science, the power of metaphor to frame or reframe debates is reflected in the work of scholars like Evelyn Fox Keller (1995), who studied the effect of metaphors on genetics, and Amy Koerber (2006) who examined the way changing metaphors altered the theory of pediatric immunology. In political science, scholars have studied metaphor’s effects on political communication and public opinion (Ottatti, Renstrom, & Price, 2014); in social psychology metaphor has been used to investigate the body (Casasanto, 2014); and metaphor has been used extensively in cognitive psychology and marketing to study buying behavior and consumptive motivations (Zaltman & Zaltman, 2008).

In a class experiment, a group of students were told a small democratic country had been invaded and had asked the US for help. The students, as representatives of the US had to make a decision: appeal to the UN, intervene, or do nothing. Each group was given a different textual description of this hypothetical crisis, which used specific language designed to trigger a
different historical analogy in their mind’s eye—WWII, Vietnam, and the third was historically neutral. The group exposed to the WWII language, made more interventionist recommendations than the others. Just as we can’t ignore the literal meaning of words, we can’t ignore the analogies and ideologies triggered by them, which can have clear, and sometimes significant, policy implications, even if only in a fictionalized project for a college class (Geary, Metaphorically Speaking, 2009).

This dissertation project seeks to become more aware of the ways we describe and thus frame the discourse on Agent Orange. Choosing to conceptualize Agent Orange as a “battle” instead of a “humanitarian crisis”, for example, reinforces an ideology suggesting “might is right.” A battle requires winners and losers; there are casualties; battles necessitate destruction and protectionism, not cooperation and conciliation. How might the lived experiences of veterans and the Vietnamese change if key stakeholders described Agent Orange as scientific opportunity instead of scientific uncertainty?

Metaphors never operate outside of context. Cognitive structures—or how a rhetor chooses one metaphor or descriptive phrase over another—can’t be isolated or considered apart from the context of the rhetorical act or its persuasive function in discourse. Thus, as Burke’s definition of rhetoric suggests, the rhetor intends and the audience interprets, attributing intention to the rhetor, though not always the same intentions that the rhetor had (A Rhetoric of Motives, 1969). Intentions can be inferred, but they are never entirely known, and so far, computer programs used to conduct corpus linguistic work cannot account for rhetorical intention; most cannot even account for context, making human analysis a critical component necessary to valid studies of language and metaphor.
The cognitive view of metaphor suggests it is not only a mode of language, but also a mode of thought (Fauconnier, 2003). However, there is still debate in contemporary metaphor studies about whether metaphoric language reflects enduring patterns of metaphoric thought, or whether metaphors are ubiquitous because of their power to persuade. In other words, if one assumes that people come to think about an abstract subject in metaphoric terms (justice as a battle, for example), the presence of metaphoric discourse is motivated by thought. But if rhetors use metaphor only as an available means of persuasion to affect certain social ends, then metaphor is less a cognitive phenomenon and more a matter of communication (Gibbs, 2014, p. 18).

The pragmatic view of metaphor assumes the rhetor employs the powers of metaphor to persuade, and the linguistic view sees metaphor as a semantic choice. Neither a purely cognitive, semantic, or pragmatic view of metaphor however, can explain metaphor’s function or power completely. In combination though, they each add to our understanding of metaphor, and ultimately strengthen the argument that metaphorical language is more than mere stylistic flourish. Here Schon (Schon, 1993) is helpful:

There is a very different tradition associated with the notion of metaphor, however—one which treats metaphor as central to the task of accounting for our perspectives on the world: how we think about things, make sense of reality and set the problems we later try to solve. In this second sense, ‘metaphor’ refers both to a certain kind of product—a perspective or frame, or a way of looking at things—and to a certain kind of process—a process by which new perspectives on the world come into existence” (p. 137).

Thinking more broadly and examining the ways people think metaphorically helps to dismantle the arbitrary divides between thought, language, and communication because
employing metaphoric language is as much a way of *thinking* as it is *persuading* or *semantic wrangling* (Charteris-Black, 2004, p. 22). Metaphor is a context-sensitive tool that can meet multiple challenges of social and political life. Such a view of metaphor opens space for new ways of understanding its importance in the development of theories about language and behavior across important domains including science, politics, and social change.

**From Metaphors to Frames**

In the social sciences, framing is a set of concepts, perspectives, or practices between individuals, groups or societies and how they communicate about reality. Framing requires subscription to social construction theories of reality; a theory that posits the world is jointly constructed through social phenomenon by the mass media, political and social movements and organizations, or other individuals and actors (Berger & Luckmann, 1966). Framing is a rhetorical act because it involves selecting, highlighting, and diminishing aspects of an object, issue or example in an effort to influence one’s perception of the thing being framed. There are frames of thought—mental representations, interpretations, and stereotypes, in example—and communication frames, or the ways actors attempt to persuade one another through word and phrase selection, delivery, and tone (Leeds-Hurwitz, 2009). Lakoff says of framing that it is “one of the most ordinary things” humans do, and yet framing is extraordinary in that the ways we describe and frame subjects *can* result in significant attitudinal and behavioral shifts toward the subjects being framed (Lakoff, The Contemporary Theory of Metaphor, 1993).

Though framing and reframing has been used to subjugate, it can also be a means by which to liberate. Consider the results of a study published in the *Policy Studies Journal* by Max Rose and Frank Baumgartner, which looks at data on media framing of poverty and government generosity between 1960-2008. How the poor were characterized—as “cheats” or “laggards”—
directly affected government generosity towards the poor. An astonishing 82 percent of the changes in government generosity to the poor can be explained by the tone of media coverage in the preceding ten years. Rose and Baumgartner identified five frames found in poverty stories—three positive and two negative. Since the 1960s, there has been significant growth in the number of poverty stories that invoke the “Lazy” frame and a significant decline in stories that frame poverty in terms of “Misery and Neglect” with an overall decrease in positive framing matched by a correlating decrease in government generosity. As they found, media tone over a period of time is an excellent predictor of the current level of government generosity.

One way to understand the intuitive results of this study is to focus on how they summarize and put into context what several qualitative and quantitative studies have shown over the past 30 years. “After the War on Poverty, the discussion turned toward a more negative view of the poor and the policies that supported the poor, making them easy targets when looking for spending cuts” (Rose & Baumgartner, 2013, p. 42). As media discussion of poverty shifted from arguments that focused on structural cause of poverty to portrayals of the poor as “cheaters” public policies and funding to aid poverty-related programs decreased. Using a simple statistical model, Rose and Baumgartner showed poverty spending by the severity of the problem, GDP, and media coverage and used their “generosity” measure to show that government policy toward the poor was highly related to the descriptive content found in news stories of the poor. In this example, policy followed the frame, which means that, ideally reframing can have significant, positive social change effects.

In 1974, Irving Goffman’s seminal work Frame Analysis popularized the linguistic work of Charles Fillmore, who, among many other contributions to the field, is credited with the discovery that virtually every word in the English language conforms to a frame, which is why I
can collect descriptive language that does not conform to the classic formula of metaphor and still find descriptive language that frames. To summarize, Fillmore describes framing as the process by which humans try to make sense of texts and talk within a given context. Frames are devices through which we interpret an object or term given that particular context. Frames can either be created by or reflected in language (Fillmore, 1982, p. 227), which is why studying metaphorical and descriptive phrases is also to study frames.

Frames are of value analytically because by studying them and their component parts scholars can further theorize social mobilization and political action. Frames often denote position and it is important to any act of negotiation or persuasion to begin by first recognizing each stakeholder’s position (Toulmin, 2003). There are many conditions that affect the efficacy of a framing campaign including the robustness of the effort; the relationship between the frame and larger ideological systems of belief; the relevance of the frame; as well as cycles of protest (Benford & Snow, 2000). This is why I do not offer frame-change prescriptions in this work. That said, changing the way a social issue or subject is described can often spur alternative solutions or changes that were before impossible to see or conceptualize simply by virtue of the subtle and often unconscious ways that descriptive language and frames work in and on society (Henderson, 1982, p. 151). The conflicting and often inhumane social policies surrounding the remediation and remuneration of Agent Orange are not only found in the literal language of Congressional testimony or New York Times editorials, but in the competing metaphors embedded in those texts, and the overall ways those texts are framed, which do more to suggest the true positions and policies of actors than overt or literal law ever could (Charteris-Black, 2004, p. 23).
Critical Discourse Analysis as a Means to Interpret Metaphors and Frames

In the social sciences and humanities, “critical” is often a term used to refer to theoretical perspectives and methodologies that aim to alter the existing social and political order by making transparent the hidden ideologies that are often habitualized or embedded in discourse (Fowler, 1991, p. 89) Critical Discourse Analysis (CDA) then, is a method by which the researcher raises awareness of the social relations forged, maintained, and reinforced by language with an aim to change them. Primarily, critical discourse analysts seek to correct a “widespread underestimation” of the significance of language in the “production, maintenance, and change” of social relations of power (Fowler, 1991, p. 90) Secondarily, CDA practitioners strive to increase the ways in which language use contributes to the domination of some people by others, believing that consciousness is the first step to emancipation (Fairclough, 2001).

Critical discourse analysis is interdisciplinary, combining linguistic, political, historical, sociological, and psychological approaches to language in an effort to make visible the invisible networked webs of power operating on and in societies. Critical discourse analysis places texts within a social content in which relations of power or hegemony, to use Gramsci’s term (Selections from the Prison Notebooks, 1971), become the focus of textual analysis that aims to demonstrate how discursive practice reflects power structures and how modification of those power structures can benefit those they may otherwise disadvantage. CDA’s history can be traced to the Frankfurt school of critical theory and its overarching aim is to make explicit the implicit motivations of those in and with power. As Stubbs (1996) argues:

The world could be represented in all kinds of ways, but certain ways of talking about events and people become frequent. Ideas circulate, not by some mystical process, but by
a material one. Some ideas are formulated over and over again, such that, although they are conventional, they come to seem natural. (p. 149).

Not all metaphor is consciously constructed and employed, and so not all descriptive or metaphoric language reflects differences in social position or power, or rather, not all language is ideological. Nor is all metaphor perfectly coded and decoded. Even so, as a unit of analysis, metaphor and descriptive language represents a semantic, cognitive, and pragmatic choice that conceals underlying social processes and values, and thus collecting and critically analyzing metaphors and their conceptual keys can aid in making visible implicit ideologies embedded in discourse.

Though I am arguing that metaphoric language and frames are powerful tools that represent, construct and can balance disturbed power structures, we must guard against the temptation to see in any of these methods a panacea. Agent Orange is a network of material and nonmaterial, human and nonhuman actors than cannot be “fixed” by one means alone (Latour, 2004); (Herndl & Graham, 2013). I want to be clear that while I see great potential in studying descriptive terms that frame, I’m not suggesting that merely reframing key arguments or constructing more imaginative ways of describing key elements in the Agent Orange controversy will solve such a complicated and longstanding socio-political-technical dispute. For all the evidence to suggest reframing campaigns work, there are some issues that seem particularly resistant to reframing. Thus, even brilliant framing has limits (Druckman, 2001). I’m not arguing that framing alone is what matters, but I am arguing that framing matters; that frames serve a critical rhetorical function in social change initiatives (Benford & Snow, 2000) and are therefore of interest to technical communicators who are increasingly working to change frames so they
are more just or humane (Thayer, Evans, McBride, Queen, & Spyridakis, 2007); (McKenzie, 2002); (Tutt, 2009).

**Historical and Modern Approaches to Computer-Aided Discourse Analysis**

Computer-assisted text mining approaches first surfaced in the mid-1980s, but advances in the past decade, including access to digital archives of “big data” have radically changed the sites and possibilities of study. Text mining is interdisciplinary, drawing on methods of information retrieval, data mining, machine learning, statistics, and computational linguistics, and is used in both industry and academic settings (Grimes, 2014).

As an analog method, text analysis has been around much longer than the 1980s. In fact, text analysis is the preferred method of hermeneutics, and many modern digital tools used to analyze texts were developed from hermeneutical theories and practices. The concordance features in commercial text analysis software, for example, were modeled after analog concordance practices used by rhetoricians to investigate multiple works of literature from the *Old and New Testaments* to the works of Euripides (Allen & Gabriel, 1954) and *Homer* (Prendergast, Homer, & Marzullo, 1971).

There are multiple, individual techniques for the systematic collection and analysis of words and phrases: free lists, paired comparisons, triad tests, and semantic network analysis, to name but a few, and specific routines used to analyze units of texts including key-words-in-context, word counts, structural analysis and cognitive maps. A detailed explanation of these methods is beyond the scope of this dissertation, but for a review, definition and case featuring them see Borgatti (1998).

At the heart of analyzing unstructured texts quantitatively is counting. Though sometimes dismissed as rudimentary, counting words is a way to discover hidden patterns in a corpus,
which in turn can be used to develop theory-building categories that reflect or predict human behavior. Political science scholars have used word counts to trace the ebb and flow of support for political ideas across time. Danielson and Lasorsa mapped 100 years of front-page news and history in The New York Times and the Los Angeles Times, revealing trends that might have otherwise gone unnoted over shorter periods of time, including the rise and fall of communism as a social threat; the ascendancy of “experts” and “professionals” in American society, and the rise of “quantification” in the social perception of reality (Perceptions of Social Change: 100 Years of Front Page Content in The New York Times and The Los Angeles Times, 1994).

Though this project does not conform to traditional network analyses (explained in detail in later chapters), the research design and much of digital corpus linguistics, in general, is informed by network theory. Network analysis is both theory and method. As the authors note in the introduction to Studying Social Networks, “network paradigms are on the rise,” (Henning, Brandes, Pfeffer, & Mergel, 2013) and derive their theoretical foundations from sociology, anthropology, and social psychology, among other disciplines.

A social network is a structure composed of individuals, organizations, words, or actors called “nodes” which are connected by specific interdependencies like interest, proximity, or knowledge, forming “ties,” resulting in a graph-based structure that serves as the object of analysis. Social network analysis is no longer merely a suggestive metaphor, but an analytic approach and, as Pfeffer and others have argued, a paradigm with specific methods, theoretical statements, and software to support it. Network analysts reason from whole to part, from structure, to relation, to individual, and from behavior to attitude (Scott & Carrington, 2011, p. 40)
The earliest work in semantic network analysis was conducted before computers were commercially used to process data, when, in 1959 psychologist Charles Osgood calculated a “semantic differential.” Osgood built co-occurrence matrices and applied factor analysis (a statistical method used to describe variability among observed, correlated variables in terms of lower numbers of unobserved variables called factors) and dimensional plotting (a set of points plotted in space) to explain the relationships between words in a text, as the figure below demonstrates.

![Co-location Matrix](image)

**Figure 1.** Co-location Matrix
*Uses WordStat to demonstrate a co-location matrix that shows the total frequency count of each word with all other words, displayed as a count total and sorted by keyword.*

Co-occurrence matrices show how often every pair of words co-occurs in a text and when analyzed can describe the set of relationships of major words, phrases, or themes to one another. The subfield of semantic network analysis is the grandchild of these early analog experiments (Doerfel & Barnett, 1999, p. 1992).

Krippendorff calls semantic network analysis “relational content analysis” (Krippendorff, 2004); (Roberts, 1989); (Popping, 2000) and it has been used extensively to study the structure
of communication associations (Doerfel & Barnett, 1999), communication in terrorism cells (Diesner & Carley, 2005); knowledge discovery (Michael, Pfeffer, & Carley, 2013), and news media discourse among other areas (Van Atteveldt, 2008).

This dissertation does not qualify as a network analysis, though I used network representations of various discourses to quickly read texts, generate key words and themes to study, and to confirm quantitative findings. It is not a network analysis primarily because I am not investigating the structure of the semantic networks each discourse represents, though that is one potential way to expand this research in the future. Whereas network analysts reason from whole to part, I am reasoning from part (unit of language) to whole (public policy/culture), and whereas they look at behavior to find attitude, I am interested in attitudes that shape behaviors.

In this chapter, I defined what metaphor is and its importance in the formation of social practices. I detailed several studies that suggest the ways framing and describing impact the lived realities of others, sometimes positively as in the case of end of life care in Britain, and sometimes negatively, as in the case of framing poverty. As noted earlier, metaphors and metaphorical frames shape the ways those with power think and act and thus are powerful influences on public policies and scientific practices. Because of the ways that metaphorical and descriptive thinking works—largely on an unconscious level—using machines to automatically scan, find, retrieve, display, and compare counts and co-occurrences of words and phrases can often reveal the hidden ideologies embedded in descriptive discourse, and thus is a critical project for the technical communicator.

In the following chapter, I extend the theoretical conversation, arguing that a mixed-methods approach to studying unstructured texts is ideal, and that technical communicators are well suited to carry out this work. I then describe the particular methodology used to study the
corpora of Agent Orange texts assembled because of their representations of key stakeholders and competing positions. Finally, I conclude the chapter by highlighting some of the limitations and constraints such methodological work poses. In Chapter 4 I detail findings, and in Chapter 5 discuss future research goals.
CHAPTER 3: METHODOLOGY

A Definition of Big Data

Though most humanities scholars are perhaps more familiar with qualitative methods—surveys, interviews, and close reading of texts—new quantitative methods and tools allow humanities scholars to conduct different forms of investigation that may yield new and richer insights. Much of the call in humanities research to embrace quantitative methods stems from the availability of “big data,” which requires tools capable of processing and sense making at scale. But what is big data? And how can big data and data mining techniques expand the intellectual range and reach of the humanities?

In its simplest form, big data is a term that describes the growth and availability of video, voice, biological, mechanical, textual, and financial data. Some have argued that big data will turn out to be as revolutionary as the invention of the Internet with futurists envisioning a “global digital nervous system,” as Bill Gates called it. Though big data has been a critical component in business analytics for decades, it is increasingly featured in humanities research projects, and is the coin of the realm in social science research presently. Not unsurprisingly, even the White House is interested in big data. In 2012, the White House’s Office of Science and Technology Policy department announced the “Big Data is a Big Deal” program, a program worth $200 million in funding managed by six federal departments and agencies (The Administration, 2012).

Big data’s promise is simple: more data found and mined through algorithmic discovery leads to insights that traditional research cannot. For technical communicators working in industry, big data suggests obvious advantages in operational efficiencies and cost reductions.
and is used in risk management and in managing social media communities, but its applications are virtually limitless (Thayer, Evans, McBride, Queen, & Spyridakis, 2007). In the Humanities and Social Sciences, big data is used in increasingly interdisciplinary and interesting ways like the 2013 project run by the Humanities, Arts, Science, and Technology Alliance and Collaboratory at Duke University, better known as HASTAC, called “Making Data Matter: Big Data Tool, Ethics and Social Change.” This project looked at new tools, methods and collaborations capable of mining historical records for social engagement and ethical urban planning processes (Grant, 2013). The purpose of the project was to engage students in a meta-analysis of the ways in which interdisciplinary teams of students, faculty, and community members could think about the ethics of big data, as well as what big data might reveal or obscure about urban planning. Much like the aim of this project, students developed new kinds of open source tools with which to analyze records, artifacts, maps, images, texts and oral narratives. They also learned to work on cross-disciplinary teams composed of computer scientists, engineers, mathematicians and urban and environmental planners. And they explored the challenges and opportunities inherent in big data work. In return, the results of collaborative projects that mine big data often paint broader and better—or at least more complete—pictures of the interworkings of culture than one method, from one discipline, or a small sample of data ever could.

Given the volume of textual data alone stored online (by some estimates 80% of all human knowledge presently), humanists interested in quantifying aspects of digital texts are limited only by their ability to collect and process it. Concordances, collocation tables, and word frequencies collected by hand are effective for many types of questions, but these methods cannot scale without machine assistance. Moreover, analog methods do not lend themselves to
discussions of completed works (like the complete works of Shakespeare), or works from other periods (comparing Shakespeare’s use of metaphor to Milton, in example), or issues related to gendered or ethnic language, or other units of measurement that are small and difficult to find and count at scale. Leveraging machine capacity to quantify and identify patterns across massive corpora offers our discipline a new, productive and powerful capability.

Big data can be thought of not so much as a collection of objects as it is a set of sites with certain properties. In 2001, industry analyst Doug Laney first suggested that a set of data counted as “big” if the data set abided the three Vs: Volume, Velocity, and Variety. Volume was defined as unstructured data streaming from social media, or sensor-to-machine data. Velocity is the literal speed or pace at which data streams into repositories for collection and analysis, and variety is determined by the kinds of data that can be analyzed.

This project doesn’t abide all the traditional terms in order to qualify as a big data project. But, I do draw from a volume of texts that far exceeds a human’s capacity to closely read and analyze, and this text is indeed varied. The Congressional corpus analyzed in this thesis (and explained in detail later) alone is composed of more than 2 million words. Velocity was less of an issue in this project simply because I chose to analyze static, archived texts that were easily accessible, though I deal with problems of velocity in a related project that attempts to archive streaming posts and metadata collected from a Facebook group called “Children of Agent Orange.” There was ample variety in the corpus analyzed, as the unstructured texts that make up the corpora studied came from congressional, military, news, and scientific documents, but variety poses its own limitations, as detailed later in this chapter.

Digital texts are a favored source of data because they are cheaply or freely available or easily accessed through subscription databases like Proquest or LexisNexis, circumventing
problems of velocity without giving up variety or volume. Whereas traditional qualitative tools that result in richer texts or “thicker” description—surveying, observation or close reading, in example—may deliver deeper psychological insights than news articles or policy briefings, these instruments are costly in terms of time and money. Such was the case for this project. As it was designed, I had to consider what forms of text I could access quickly, consistently, and inexpensively, which necessarily constrained my overall vision, but provided an access point—a way in.

So much data streaming from so many points however, changes traditional ideas of a “field site” suitable for humanistic or rhetorical research, and our field has not yet come to terms with the implications of this. Composition conducted “teacher research” from the 1970s on, and Rhetoric conducted non-academic writing research from the 1980s forward, but, as Stephen Gilbert Brown wrote, “in the aftermath of the postmodern assault” the concept of a “field site” was redefined in broader terms to include sites of language and discourse found in our communities and classrooms (2004, p. 303), which now must certainly include digital space.

Though scientists, researchers, and writers used texts for hundreds of years as a means to understand a culture’s values, attitudes, and beliefs, the sites of collection have changed. Magazines, newspapers and books are now digitized; survey responses and interviews are accessible 24/7 through web archives, and more archival data is increasingly crowdsourced and curated, raising issues of authenticity and ethicality (Alistair, 2012). Because our sites of research have changed, and because textual data has changed, our methods of organizing, accessing, and processing textual data must change in turn. In the same way that digital communications have changed notions of intellectual property, authorship, textuality, and what counts as “writing”, digital communications have also changed the research sites, questions and
methods we use to study humanistic concerns. Margaret Baker Graham suggests professional communications would be richer were it to extend the boundaries of its research (Graham, 1999, p. 183). Though she does not mention the sweeping digitization of texts specifically, I would argue that studying digital texts with quantitative tools is simply the logical next step in contemporary rhetorical research. When venturing into new methodological territory however, we must pack the right tools and methods, and know our ethical limits.

As recently as 2000, ethnography did not include digital, virtual communities and methods, though now many ethnographic studies of digital communities like Facebook or Twitter are commonplace. In such studies technology can serve as both the subject and the tool with which the subject is studied. Though I consider the various corpora to represent stakeholder communities, this particular project is not a digital ethnography.

But offshoots of this project, specifically the Facebook analysis, which studies how ethos is created and maintained in digital lay communities, highlights that the concept and nature of “fieldwork” in the digital age has indeed changed, the concept of “text” has been expanded, and the ethics guiding “big data” textual research is still being vetted.

Conducting an ethnographic enquiry through the use of CMC [computer-mediated communication] opens up the possibility of gaining a reflexive understanding of what it is to be a part of the Internet. This provides a symmetry to the ethnography, as the ethnographer learns through using the same media as informants.... An ethnography of, in, and through the Internet can be conceived of as an adaptive and wholeheartedly partial approach which draws on connection rather than location in defining its object (Hine, 2000).

As we enlarge our disciplinary understanding of “data,” incorporate new tools that extend our ability to analyze and criticize new sites of text, talk, and testimony, and as we codify what
counts as legitimate methods of collection, analysis, interpretation, and visualization, we will understand the affordances and limits of digital rhetorics, and of conducting mixed methods big data research.

**Brief Description of Research Site**

For this project, I created a research site composed of texts representing four stakeholder views: the U.S. Congress, the U.S. news media, the U.S. media-editorial view, and the Vietnamese state-owned media. Part of the reason for choosing these texts and not others had to do with access—all were either available through subscription, or were available online and could be collected using a web scraping tool (a technique of extracting information from public websites). It was particularly important to me to include the Vietnamese perspective and voice in this research, as that is a voice often diminished in narratives of Agent Orange. I was limited however, by the availability of data that represented the Vietnamese perspective. The Vietnamese News Agency (VNA) was selected for inclusion because it was available, digitized and published in English. As automatic translation tools are still fairly rudimentary, I was worried I would miss important nuances if I attempted to use machine-translated texts. Ideally, and perhaps in a future study, narratives would be gathered from Vietnamese victims, translated, digitized and then analyzed.

I created the corpora by keyword Boolean searching for “Agent Orange” and “Vietnam” and not “band” (because there is a rather prolific punk band by the same name, which initially slanted results) in ProQuest. All except the Vietnamese News Agency texts were collected from either ProQuest or ProQuest Historical Newspaper databases. I attempted various techniques for searching the database with the aim of collecting all available texts that were predominantly about Agent Orange without collecting duplicates and without littering the corpus with articles
that only mentioned Agent Orange but were principally about other subjects. Ultimately, this would not be an issue because of the way I further processed the texts in Python, (an extensible programming language), collected and calculated texts that met certain criteria for inclusion and were within a certain distance from a keyword.

I should note here that I recognize the problems in identifying only a selection of possible documents as material evidence of a discourse community. One of the greatest values of big data text mining is that N can equal all. As this was a preliminary investigation into method however, and as accessibility issues were a constant problem, N is only a fractional representation of the potential documents one might include in a more robust study where time, money, and research assistance was unlimited.

Clearly, the congressional members who sat on committees and heard testimony contained in the documents that make up that corpora do not represent the opinions, values, attitudes, or beliefs of all U.S. politicians any more than the three or four Vietnamese who testified before Congress represent all of Vietnamese victims. For this study, I define a discourse community as a group of people who share a set of discourses and use communication and rhetoric to achieve communicative goals within the discourse (Porter, 1992) and so, while the genres, conventions, and purposes of the texts vary, the collections represent a community—of politicians, scientists, reporters, and writers—with shared communication goals.

In the hard sciences, variables are controlled, but texts are, by their nature, variable and multivocal. Different texts are bound by different genre conventions, and shaped by immeasurable editorial and political pressures. In future iterations of this project, I would reduce the scale of texts considered, count more variables like gender, time, context, speaker, appeal or
political party and ensure the data I collected to process included all variables across all texts, thus limiting the impacts of a text’s variability.

Though there was variability in the texts in number, size, and time, this was not and insightful or helpful kind of variability, I found. The variance in time, for example, made it impossible to correlate shifts in time with the introduction of certain phrases, and the lack of variables like turn-taking or even political party resulted in fewer opportunities to correlate linguistic shifts to wider cultural markers.

**Corpora in Detail**

**Congressional Hearings & Reports**

In the 1960s, Congress was certainly aware of Agent Orange, but the public controversy had not yet begun, and so texts added to this decade of the corpus include technical memorandums like that composed by Richard A. Hensen in August of 1965, describing the “Physical Properties of Normal Butyl Esters of 2,4-D and 2,4,5-T and “Orange.” Also included are science and policy Legislative Assessment Reports. Such reports were delivered to the Subcommittee on Science, Research, and Development of the Committee on Science and Astronautics in the U.S. House of Representatives, 91st Congress, 1969, but are quite different in purpose, audience, tone and genre than the congressional hearing documents that follow in later decades.

In later decades, the Congressional corpus includes reports to the Comptroller General on the costs of covering negative health effects related to Agent Orange. By the 1980s, and at the height of scientific controversy, the corpus includes special reports delivered to the Committee on Veterans Affairs, Hearings, and the testimony of Admiral Zumwalt. Zumwalt was an important character in the narrative of Agent Orange, as he and his son, both high-ranking
officials in the US military, died of exposure-related disease. There are also health and science briefings by Dr. Alvin Young. By the 1990s and 2000s, most of the texts reflect testimony gathered in hearings that resulted in bills and appropriation reports, which include the scientific and technical data that make up the earlier sections of the corpus. The congressional corpus is the most diverse of the four.

One of the key challenges in big data projects is to find the most meaningful ways to delimit data and reduce complexity. This is particularly challenging when the data is variable in form, length, or time. For example, it was easy to delimit the congressional corpora by decade but not the Times editorials or the VNA because there simply wasn’t enough data in those corpora. To compensate, at least for the Times editorial corpora, I might have included editorials from other papers, but I didn’t have access to other historical newspaper databases.

I chose to delimit the corpora by decade because at first, I wanted to not only compare the ways in which each discourse community talked about Agent Orange and its issues, but I also hoped to track shifts within individual stakeholder communities. This proved impossible to do with the existing corpora; it simply wasn’t large enough to trace shifts at this level of detail.

Though there wasn’t enough text in some of the corpora, there were thousands of pages and millions of words to keep track of. At the risk of stating the obvious, when dealing with hundreds of documents and millions of words, sound document management practices are critical. If a text was already machine-readable, it was placed in a folder for processing and given a unique file number. If a text was not machine-readable, it was processed using Abby Fine Reader, a Macintosh-compatible optical character recognition (OCR) software tool that turns flattened PDFs (portable document format) into machine-readable documents.
New York Times Articles

All of the articles from 1985-2014 from *The New York Times* were digitized, but the historical articles, those from 1968-1985 were not and, like the older Congressional records, had to be converted from scanned PDFs to machine-readable texts. As I have argued elsewhere, the media played a significant role in constructing the narrative of Agent Orange and without media intervention and dedicated coverage of Agent Orange, it is highly unlikely veterans would have won “presumptive exposure” from the U.S. government, or that the former Air Force bases in Vietnam would now receive the kind of aid dollars necessary for the large-scale remediation efforts presently underway. But it bares repeating that *The New York Times* was not selected as a media outlet because it was representative of all U.S. news media, or even because it was presumed to be objective\(^7\), but because the *New York Times* has invested considerable time and infrastructure in digitizing its historical and modern articles collection, because I had unfettered access to content from a very specific time frame (1968-2014), and because the news media was an important source of historical evidence given the scope and aim of this project.

New York Times Editorials

Like the Congressional and *Times* article corpora, editorials from *The New York Times* were selected and treated as an individual discourse community because empirical analysis suggested they were written by veterans or friends and family of veterans and thus were as close to the thicker description surveys or interviews would have provided than any other textual data available to me. Editorials were also specifically of interest because they often employ the use of metaphoric language more frequently than articles or scientific reports, which are bound by genre conventions and guided by an ideal of objectivism. Further, editorials are more often

\(^7\) The presumed liberal bias of *The New York Times* should be noted here: Some studies suggest the *New York Times* is quite liberal and biased (Groseclose 2004), while other studies suggest what bias exists is focused on social, not political issues (Media Matters).
examples of deliberative discourse, which propose policy and action that often depend on expressing explicit values in ways that news stories do not.

**The Vietnamese News Agency**

The Vietnamese News Agency corpora was compiled with the assistance of Dr. Jürgen Pfeffer, who used a web scraping tool to pull the texts from the digital archive using a keyword search function to delimit articles of interest from all available articles.

The statistics of each corpus constructed and analyzed are detailed in the following table.

<table>
<thead>
<tr>
<th>Name of Publication</th>
<th>No. of Cases</th>
<th>Number of Words*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Congressional Hearings/Reports (1960-2000)</td>
<td>210</td>
<td>2.8 million</td>
</tr>
<tr>
<td><em>The Vietnamese News Agency</em> (2006-2014)</td>
<td>994</td>
<td>304,700</td>
</tr>
</tbody>
</table>

*Note: This table details the number of cases and number of words analyzed in the study. Word counts are rounded up to the nearest 100.*

**Research Questions**

This study was designed to answer the following research questions:

1) What combination of tools and procedures best identifies, extracts, and calculates metaphorical frames from large corpuses of unstructured texts?
2) What were the most common metaphors or descriptive phrases used to talk about Agent Orange across discourses?
3) What were the significant similarities or differences between the ways different discourse communities talk about and describe key themes related to Agent Orange?
4) What affects might these differences have on public policies about Agent Orange?
5) In what ways does semi-automatic text extraction methods advance and limit our research capabilities?
Determining the Unit of Analysis

The term “metaphorical frame” comes from Otatti and Renstrom (2014). A communication cue is activated by a “root metaphor” in the recipient’s mind and this root metaphor contains images, a central theme or narrative associated with the subject being described. Therefore, a metaphorical frame both describes and frames concurrently (p. 2), but a metaphorical frame is different from a metaphor and different from a frame in that it only reflects aspects of metaphor and frame. The term “environmental modification,” in example would have been considered a metaphorical frame in this data set because it is descriptive instead of literal, and it was a term used repeatedly to help frame a controversial action as helpful, necessary, and largely positive. Remember that framing, like metaphor, highlights some aspects of the object, subject, or issue being framed, while diminishing or excluding others. In this case, environmental modification was a far less emotionally charged way of describing Operation Ranch Hand than a chemical weapons program. Finding metaphorical frames however, is very difficult for machines because their identification often relies on context. So the first challenge in automating this process was to attempt to convert context-dependent phrases and words into equations a machine could read.

Working in tandem with a research assistant, we separately analyzed 10% of the corpora, randomly selected. The assistant was given instructions on what a metaphor and metaphorical frames were, and how they functioned linguistically. She was also given several examples. Separately then, we analyzed the same sample, searching for and highlighting any instances we felt abided the definition of a metaphor, metaphorical frame, or way of describing. Sometimes, a metaphorical frame was presented as a single lexical item, as in the term victim; other times the ways of describing were phrases like unjustified suffering. As one can see, neither of these words
or terms conforms to the definition of a classic metaphor, and so I use the term metaphorical and
descriptive frame interchangeably to represent the lexical items and phrases identified that
describe and frame.

I did not attempt to count the frames of each text. Very little research has been conducted
to validate automatic frame identification methods (David, Atun, Fille, & Monterola, 2011),
partly because the process is so context-dependent. As Entman writes, frames require a “tacit
understanding” of text, context, and subtext (Entman, 1993), which machines find difficult.
Though machines are not yet capable of finding, reading, and organizing frames, they can find
particular words that occur in particular patterns, and some of these patterns, we found, were
more or less indicative of a frame. Here is an excerpt from an article written by the Vietnamese
News Agency from 2005 with what both coders counted as metaphorical or descriptive frames
(underlined).

Three Agent Orange victims filed lawsuits against 36 US chemical firms, asking them to
compensate for damages and suffering they have endured as a result of American War
injuries.

After we identified metaphorical frames or ways of describing in the sample texts, we
then went back and converted each identified phrase into a parts-of-speech equation and began to
tally the equations, searching for the most common. The most common parts of speech patterns
that most often signaled a descriptive frame included:

• adjective>noun
• adverb/adj/noun
• verb>preposition
• adverb>verb
• noun>preposition>noun.

Of these, the most common signal of a descriptive frame was an adjective followed by a noun as in the term “toxic chemical.” Once these parts of speech equations were identified, we had a means through which to parse volumes of text, extracting only the sentences that related to relevant keywords and conformed to the linguistic equations most likely to yield a metaphoric frame. Once this routine was written in Python, the extraction, calculation and comparison between the different corpora was automated and what would have taken a team of researchers months to read and analyze, took a matter of days.

Inter-coder reliability is a critical component of any text analysis projects. By itself, coder agreement does not ensure validity, but without coder agreement, data and findings cannot be considered valid at all. As Neuendorf argues, "given that a goal of content analysis is to identify and record relatively objective (or at least intersubjective) characteristics of messages, reliability is paramount. Without the establishment of reliability, content analysis measures are useless" (Neuendorf, 2002, p. 141).

There are multiple indices to measure intercoder reliability. Popping (2000) identified 39 different “agreement indices,” but in practice, two are adopted most frequently for their ease and flexibility: Cohen’s kappa (k) and Krippendorff’s alpha (a). While these calculations were once done by hand, most coding software suits like NVivo and QDAMiner offer built-in inter-rater coding calculations. For this project, I calculated the inter-rater agreement measures using Cohen’s kappa (k) because it eliminated unnecessarily complex weighting systems. Cohen’s kappa is the measure of agreement between two raters that determines which subjects belong to which categories once chance is factored out. So, raters either agree or disagree in their rating (this is or is not a metaphorical frame). No weightings are used. Whether using Cohen’s kappa or
other reliability measures, reliability is typically described as a percentage of agreement. If reliability is the extent to which any test yields the same results when repeated, then inter-rater agreement is evidence that a metaphorical frame, keyword or category of language has some validity apart from mere whim. Standards of intercoder reliability are somewhat arbitrary, but Krippendorff promotes 70 percent agreement. For work in text analysis of which intercoder reliability is a significant piece see (Holsti, 1969); (Pool, 1951); (Krippendorff, 2004). For this project, the aim was 70% inter-rater agreement and we achieved 73% across our sample.

**Method for Data Collection and Analysis**

I next collaborated with a computer science doctoral student at the University of Arizona and together, we wrote a custom Python 2.7 script that used the Natural Language Processing Toolkit (NLPT) to parse the corpora accordingly (see appendix for code). NLTK (Steven, Loper, & Klein, 2009) is a leading platform for building Python programs to work with language sets (De Smedt & Daelemans, 2012). Its interface integrates the most common corpora and lexical resources like WordNet, which we used as a basis for dictionary creation, both of which are explained in more detail below. The NLTK also afforded us traditional processes employed in text analysis research like classification, tokenization, stemming, tagging, and parsing.

Some might wonder why we chose to write our own script instead of using any number of commercially available qualitative/quantitative software tools like WordStat or NVivo. Though these programs indeed offer more robust features and visualization functions, using a commercial program weds the researcher to the program’s logic and visualization interfaces. For a significant portion of the post-processing visualization work, I indeed used WordStat, but found its system of rules to be inflexible and inadequate when it came to parsing text at the level.

---

8 For more information, visit: [https://www.python.org/](https://www.python.org/)
9 Documentation found here: [http://www.nltk.org/](http://www.nltk.org/)
I needed it parsed to find metaphorical frames. WordStat uses a cumbersome system of rules and folders and after significant experimentation with this rule system, it became obvious that I needed the data processed (cleaned) and parsed before I could use WordStat to tabulate and visualize findings.

One of the reasons text analyses can be difficult is because there is no single, systematic way to process texts. Each inclusion or exclusion in a pre-processing routine carries its own consequences. Having said that, there are procedures that are more or less commonplace like stemming (reducing inflected words to their stem), lemmatization (grouping different inflections into one form), and using stop lists (lists of words too common to be of linguistic value).

After significant experimentation with various cleaning and pre-processing routines, I settled on the following process:

1) **Data input and preprocessing**
   (included lemmatization and the use of stop lists)

2) **Theme matching**
   (parsing a sentence from a paragraph if it contained a relevant key word like “Agent Orange” and associating the parsed sentence with its “bucket”)

3) **Equation matching**
   (further parsing the keyword-identified sentence by the linguistic equation contained in it in an effort to find descriptive terms and metaphorical frames)

4) **Result filtering**
   (filtering results, concatenating files, exporting and processing in WordStat and visualizing findings)
In the first step, texts were read from the raw data files provided, which were converted into machine-readable text files. We sanitized all text files by replacing extraneous characters like “smart quotes,” junk metadata that scrambled in the conversion, and numbers with standard characters that would not be counted in our results. We then split each document into a list of sentences using NLTK.

The second step included finding the themes present in each sentence by first identifying a list of keywords. These are found using the “pattern” library for Python. The regular expressions for each theme are shown in Table 2 on the next page.

Determining keywords was a somewhat arbitrary decision made after closely reading several historical texts about Agent Orange, consulting with subject matter experts, and identifying keywords relevant to each discourse using network analysis tools to distantly read the texts (more description on distant reading follows later in this chapter). On the first pass, there were 20 themes identified and the corpora were parsed accordingly, but after digging into the results, it became obvious that there were too many overlaps, which reduced the validity of the findings, and so the 20 keywords were reduced to the most significant 5. These five keywords were selected not only because they were central to the discourse and debates about Agent Orange, but also because the keywords did not overlap.

In the third step of this method, we applied “pattern” to find “part-of-speech” regular expressions in the text. Remember, with the help of my research assistant, we had already analyzed, gathered, and tabulated the most common linguistic patterns suggesting metaphorical frames. In this step, we associated the patterns (adj before noun) with a keyword. These matches represented probable metaphorical frames within the text, specific to the keywords of interest.

---

11 For this project I worked with Edward Martini, but also interviewed Fred Wilcox, author of Waiting for an Army to Die, and David Zeirler, the author of The Invention of Ecocide.
Table 2. Examples of Themes and Expressions

Themes defined in this work and the regular expression used to find them within the text.

<table>
<thead>
<tr>
<th>Keyword</th>
<th>Pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dioxin</td>
<td>dioxin</td>
</tr>
<tr>
<td>Exposure</td>
<td>exposure</td>
</tr>
<tr>
<td>War</td>
<td>war</td>
</tr>
<tr>
<td>Veteran</td>
<td>veteran</td>
</tr>
<tr>
<td>Vietnamese</td>
<td>vietnamese</td>
</tr>
<tr>
<td>Justice</td>
<td>justice</td>
</tr>
<tr>
<td>Environment</td>
<td>environment</td>
</tr>
<tr>
<td>Uncertainty</td>
<td>uncertain*</td>
</tr>
<tr>
<td>Quantification</td>
<td>economics</td>
</tr>
<tr>
<td>Remediation</td>
<td>remediation</td>
</tr>
<tr>
<td>Remuneration</td>
<td>remunerat*</td>
</tr>
<tr>
<td>Time</td>
<td>*time</td>
</tr>
<tr>
<td>Health</td>
<td>health</td>
</tr>
<tr>
<td></td>
<td>effects</td>
</tr>
<tr>
<td>Victim</td>
<td>victims</td>
</tr>
<tr>
<td>Ranch Hand</td>
<td>ranch hand</td>
</tr>
<tr>
<td>Legal</td>
<td>legal</td>
</tr>
<tr>
<td>Veteran’s Administration</td>
<td>The VA</td>
</tr>
<tr>
<td>Proof</td>
<td>Proof</td>
</tr>
<tr>
<td>Chemical Manufacturers</td>
<td>chemical manufacturers</td>
</tr>
</tbody>
</table>

Note: “|” indicates a logical OR relationship, where any match will be categorized under the theme, and “*” indicates that any characters following the word will match. For example “spray*” will match “spray,” “sprayed,” “spraying,” etc. Originally, this work attempted to trace these 20 keywords across the four corpora, but ultimately, this list was reduced to 5 distinctive key words that could be easily traced.

The Part-of-Speech regular expressions are found in the following Table 3. We used WordNet, a large (35,000 words) lexical database of English words including nouns, verbs, adjectives, and adverbs to generate synonyms related to keywords so we found not only ways of describing “veteran” but also common synonyms for veteran like “soldier.” Below is a table with a few representations of keywords, the linguistic equations found and the metaphorical frames pulled and tallied.
Table 3. Equations and Examples of Resulting Metaphorical Frames

<table>
<thead>
<tr>
<th>Keyword</th>
<th>Linguistic Equation</th>
<th>Metaphorical Frame</th>
</tr>
</thead>
<tbody>
<tr>
<td>Veteran</td>
<td>JJ* NN*</td>
<td>Veteran Affairs</td>
</tr>
<tr>
<td>Veteran</td>
<td>JJ*</td>
<td>Vietnam Veterans</td>
</tr>
<tr>
<td>Veteran</td>
<td>JJ* CC</td>
<td>IN JJ* NN*</td>
</tr>
<tr>
<td>Veteran</td>
<td>VB* NN*</td>
<td>Said Scott</td>
</tr>
<tr>
<td>Veteran</td>
<td>NN* IN NN*</td>
<td>Fact that Tom</td>
</tr>
<tr>
<td>Veteran</td>
<td>VB* IN</td>
<td>Fighting with</td>
</tr>
<tr>
<td>Veteran</td>
<td>VB*</td>
<td>Measure</td>
</tr>
<tr>
<td>Veteran</td>
<td>VB* PRP* IN</td>
<td>NA</td>
</tr>
<tr>
<td>Veteran</td>
<td>is like a</td>
<td>NA</td>
</tr>
<tr>
<td>Veteran</td>
<td>is like an</td>
<td>NA</td>
</tr>
<tr>
<td>Veteran</td>
<td>is like</td>
<td>NA</td>
</tr>
</tbody>
</table>

In step 4, we ensured all of the relevant linguistic equations were within three words of a keyword, removing any that were not. Though at first we tried a window of seven and five keywords, we found that a window of three keywords was a better measure of a metaphorical frame, or was more likely to signal descriptive language relevant to the keyword or theme of interest.

Next, and finally, I combined the text files from all corpora into one master text file representing each stakeholder community and keyword, which resulted in all sentences identified as having a potential metaphorical frame in a given discourse on a given topic. I then used WordStat to calculate the occurrences of phrases, keywords, and co-occurrences and finally, I went through each of the phrases that occurred 20 or more times and resolved any conflicts (words or phrases that were counted but did not conform to a metaphorical or descriptive frame) manually.

---

12 In AutoMap (a tool that Deisner and Carley created at Carnegie Mellon to process text and export networks that can be read in the companion software ORA) statement formation choices are implemented as various operationalizations of the distance between concepts (Carley 1993). These operationalizations reflect different approaches of the windowing method (Danowski, 1993). By settling on a window size the user defines how distant concepts can be from each other and still have a relationship. The window size then is applied to the specific text unit. Windowing defines the size of a window as a length of a string of adjacent concepts that get linked into statements if they match the statement formation criteria. See also Weber (1990) and Lasswell & Pool (1951).
We managed our data exports using Microsoft excel, specifically for their sorting formulas, which resulted in files that look like Figure 2. These files visualized the parsed textual data by individual file name, sentence in which the keyword occurred, related themes, related equation identified, frequency, the TF*IDF of each phrase or word, and where the word/phrase occurs in other parts of the corpora. The ultimate counts and bar charts however were developed using WordStat.

Figure 2. Data Organization in Excel
A screenshot of the exported results from Python into Excel demonstrating how the data was organized by file name/sentence/next theme/equation/words/count and TF*IDF (term frequency*inverse document frequency).

Distant Reading, Interpretation, and Analysis

What are we to make of so much textual data, so many ways of delimiting, counting, and describing the terms and concepts important in a narrative constructed by multiple stakeholders? One of the advantages of machine-assisted text analysis is that it enables the researcher to approach texts relatively free of existing notions and bias. That said, processing texts so a machine can find patterns of language embedded within it only gets one so far; analysis, theming, and interpretation are the final steps toward application. That is the work of the following chapter—to detail the findings and make sense of those findings. The payoff of this kind of work is that it can lead to greater awareness of key issues and actors, better fundraising opportunities or collateral development, and perhaps even more humane and inclusive public
policies and scientific practices. Before reviewing the results, I need to define the key statistical measures used to value my findings: *frequency*, TF*IDF* (the metrics on which the most frequently occurring phrases were based), and *co-occurrence*.

Scott defines *frequency* as those words or phrases that occur in a corpus with higher than expected or statistically significant counts (Scott M., 1999); (Manning, Schutze, & Raghavan, 2008). The frequency of a word or cluster of words is partly what defines nodes central to a network, which is why the network visualizations included in the appendix closely reflect the frequency counts in the tables beneath them. Frequency is an important statistical measure because it helps to define both *salient* words (words of significance), as well as word *valence* (a word’s emotional charge). Valence is an important measure because it often denotes a frame. I do not include valence dimension measures here because that would be a very different project, but taking an aspect of the corpus and measuring valence would be an interesting way to extend this work further. Both salience and valence are only accurately measurable once stop words have been removed. As noted earlier, *stop words* are those words that occur too frequently to be rhetorically valuable, words like “a” “and” or “the.” Removing stop words is a common text preprocessing routine that reduces the complexity and volume of a text.

Once raw text files were delimited by corpora, decade, and keyword, all sentences that contained a linguistic equation of value were concatenated (combined) and analyzed using WordStat. From within WordStat, I produced the bar charts featured in the following chapter, which measured frequency counts for keywords, key metaphoric frames, and co-occurrences.

Frequent phrases were determined important both because they scored high in frequency but also had a high TF*IDF* measurement. TF*IDF* stands for term frequency-inverse document frequency, and this weight or measurement is commonly used in information retrieval and text
mining because it evaluates how important a word or phrase is to a document in a collection. The importance of the word or phrase increases proportionally to the number of times the word or phrase appears in the document, but is tempered by the frequency of the word/phrase in the corpus overall. For the calculus of TF*IDF, see Chris Manning’s *Introduction to Information Retrieval* (Manning, Schutze, & Raghavan, 2008). The NLTK in Python and WordStat program both had built-in TF*IDF calculators, which were used in this project.

Finally, I chose to look at co-occurrence counts because co-occurrence words sometimes offer additional insight into how themes in a narrative are framed. A word or phrase may occur frequently and thus be valuable in demonstrating the content of a text, but what it occurs alongside frequently adds additional, less obvious insight, sometimes suggesting a frame. In example, the co-occurrence of the terms “dioxin” and “victims” in one corpora but not in another may suggest a specific rhetorical move made to associate the two words and frame discussion in a rhetorically compelling way. This example comes from the corpus actually, where dioxin co-occurs with words like “environment” and “victim” and is used in far greater frequency in the Vietnamese News Media text than it is in the congressional or *Times* corpora, suggesting more than just the content of the corpus, but also the way the content was framed.

Frequency is the same whether applied to keywords or phrases, but co-occurrence is a slightly different measure altogether. Co-occurrence is the above-chance frequency of co-occurring words within a determined span of words (Sinclair, 1991). The decision to include words on either side within a certain “window” is somewhat of an arbitrary decision that can only be decided once a researcher is “in” the corpus and experiments with how widening or narrowing the widow affects results. I found that a word or phrase within a seven-word window produced meaningful results. The calculation of co-occurrence is based on the combinatory
results from frequency of the term, frequency of the words co-located to the term, and the
number of times the co-location occurs above chance, making the co-location meaningful.

Co-occurrences are an important metric because they contribute to a term’s contextual
meaning (Nattinger & DeCarrico, 1992) but they can also illuminate unseen associations
(Hunston, 2002). Here, two additional terms are helpful because they identify two important
types of textual interpretations that can be drawn: semantic preference and semantic prosody.
Semantic preference is the relationship “between a lemma or word form and a set of semantically
related words” (Stubbs, Words and Phrases: Corpus Studies of Lexical Semantics, 2001, p. 65).
The term “herbicide in” naturally co-occurs with certain words like “Vietnam” or the “the
environment,” which demonstrates a kind of linguistic preference but it might not indicate
anything beyond this. That would make finding this term used in this way less meaningful in one
sense—because it was commonly used—but perhaps meaningful in another, as one looks at the
words with which the phrase co-occurs. If the phrase “herbicide in” co-occurred multiple times
with “the devastated environment” then that would be significant finding, a phrase that is
pragmatic, linguistic, and rhetorical, while signaling a frame. It should be said here that not all or
even most of the top ten phrases identified for each keyword and stakeholder offer this much
depth and insight. That’s one of this project’s surprising disappointments. I have graphed only
the top ten keywords, phrases, and co-occurrences, but that doesn’t mean there aren’t high-
frequency counts or phrases that are more linguistically or rhetorically valuable or interesting yet
occur less frequently. It is to suggest that text-mining projects demand simplification of the
corpus and strict data reduction procedures so as to separate the signal from the noise.
Identifying and abiding statistical measures (windows of words, counts, and cut offs for
frequency counts, in example) are an important means by which to achieve this separation.
Having defined semantic preference, I also need to define semantic prosody. *Semantic prosody* is evaluative in that it can reveal the rhetor’s frame of mind or reference. Prosody “extends over more than one unit in a linear string” (Stubbs, Words and Phrases: Corpus Studies of Lexical Semantics, 2001, p. 65). Stubbs uses the word “cause” to explain this idea. The word “cause” occurs overwhelmingly often alongside or in close connection with words that also signal unpleasant events (Stubbs, Words and Phrases: Corpus Studies of Lexical Semantics, 2001). Co-occurrence then is an important tool that may help identify keywords in context, but also might help to identify hidden valence. It is also important to mention here that co-occurrence is not the same thing as collocation. Collocation means physical proximity to. Co-occurrence is the presence of two terms within a certain window of words.

As mentioned earlier, the critical aspects of this project demand not only linguistic accounting, but also an attempt to understand why or under what circumstances the rhetor made specific linguistic choices among available other options, and whether those choices map onto policy or practice in meaningful ways. This second interpretive and analytic step requires additional variables beyond text to correlate. For example, to map to public policy changes, I might take *only* the Congressional hearings related to Agent Orange debate, pull out each “turn” in the testimony, and map the turn taker or rhetor to his or her political party and the dominant rhetorical appeal used (logos, ethos, pathos) to advocate for his or her position. That would allow me to show which parties used were more likely to use what kind of appeals when advocating for or against certain policy decisions. If I further delimited the corpora by decade (assuming there was enough testimony to do so), I could also show how rhetorical appeals or ways of describing Agent Orange among the major American political parties have changed across time and within each party and I might be able to make arguments about which elected officials, under what
circumstances, employ which rhetorical appeals to “tow the party line.” Though this project wasn’t set up to offer that kind of insight, that’s not to say the findings here are without merit. It is to say one of the greatest lessons learned in corpus analysis is that it is an iterative process with results dependent as much upon research design as data analyzed.

In some research situations, text mining would reveal its results, which would then be confirmed, validated, or finessed by a subject matter expert, and then refined, re-run and analyzed again until no new results are generated. For this project, I investigated aspects of the wider historical and contextual settings relevant to the debates about Agent Orange in the United States and Vietnam as a means to verify and sense-make my findings. I examined what and how key terms were relevant to the discourse, how various keywords were privileged and how phrases constructed frames in official and unofficial sources. Additionally, I used network analysis tools to empirically validate the statistical findings. In early research design experiments, I found that studying network measures of the corpora didn’t yield insightful or different results but that network visualizations were an excellent way to quickly and visually validate word counts found using WordStat. In other words, though the algorithms used to produce a network versus those used to produce a simple bar chart of counts are not the same, network visualizations were an excellent indicator of key themes and topics and sometimes even identified outliers, or topics, phrases, or words one would expect to be there in a frequency chart, but were not.

There are dozens of network analysis tools available on the open source market to use to visualize texts as networks. For this project, I used Texteture, an open, free, nonlinear distant reading and text network visualization tool developed by Nodus Labs. This tool can quickly
“read” and visualize texts as networks of words and situate them in associated contexts for further and deeper exploration.

![Network Representation of Agent Orange](image)

**Figure 3.** New York Times Network Representation

*A network representation of a single article from the New York Times corpus on Agent Orange using the software program Textexture.*

In the network representation above, the size of the node denotes its relative importance in the network and within its cluster of contextual terms. By definition, a graph is a collection of nodes (vertices) along with identified pairs of nodes, often called edges or links. It is important to note that nodes are not linked because they are next to each other in a sentence, but because they are central to a window of context-dependent words. This node-edge structure is encoded and visually represented as a graph using the open source network visualization tool Gephi.

These and similar network visualizations depend on Gephi metrics and community detection algorithms, a deeper explanation of which can be found on the Gephi Consortium web page.13

Table 3 above visualizes the network representation of a single article from *The New York Times*

---

13For more information on Gephi, visit: https://consortium.gephi.org/
about Agent Orange. Based on the size, position, and color of the nodes in the network, one can see the central issues concern tracking veteran health through the Agent Orange registry. The terms “veteran”, “registry” and “data” are central to the network with each offering a cluster or community of subtextual content. For example, “veteran” is connected to a context that talks about health, questions, and the public.

The size of the nodes in the network graphs depends on a network calculation called betweenness centrality, a standard network measure that accounts for how often the shortest path or distance between two randomly chosen nodes appears in a network (Freeman, 1977). In semantic network analysis, betweenness centrality is an important measure because those words that measure high on this scale often appear at the junctures of meaning. A juncture of meaning is an area of the graph that visualizes a significant relationship or, in rhetorical terms, an exigency. In other words, the nodes at a juncture of meaning do not just occur frequently, but are influential. Nodes can have high betweenness centrality within clusters of concepts too, making them influential terms within a specific context or subsystem of ideas. In the network above, the most influential keywords in this text include: “registry,” “veteran,” “data,” and “health.” As detailed in Figure 4 below, the keyword “registry” is central to both the document, and a particular context that, when drilled into, maps to the central discursive ideas that one would find if reading closely. This particular text was from a 1983 editorial from The New York Times, “V.A.’s Unhelpful Data on Agent Orange.”
Networks are visually stunning, persuasive, and useful as both a way into and through the content of corpora. Beautiful network visualizations were utilized in projects like the Republic of Letters, which traces, maps, and visualizes the influence of ideas from Erasmus to Franklin (Stanford University, 2013). Network analysis and visualization has also been used to compose the Writing Studies Tree, an interactive, crowdsourced archive that maps the relationships of influence in the field of Composition and Rhetoric (City University New York, 2011).

Issues of validity are a significant difficulty and a significant opportunity in text analysis projects. Most of the outputs of unsupervised text analysis procedures are difficult to assess. External validation methods that involve external data and allow for classical statistical tests do not apply here because the text is too specific to the subject. In other words, there is no “control” in text mining. Exciting work however, is ongoing on internal validation methods based on resampling techniques that use Monte Carlo and other methods, which may be able to tackle the
problem of the plurality of words, lemmas, and sentences (Lebart, 2004). One of the advantages of mixing methods and applying more quantitative methods to qualitative studies, is that the quantitative approach can evaluate the explicit features of a text (Huckin, 2004) but the statistical analyses can extend beyond counting and offer insight into how variables may relate (Boettger & Palmer, Quantitative Content Analysis: Its Use in Technical Communication, 2010, p. 3). Using statistical measures like co-occurrence and TF*IDF suggest that relationships between words or phrases are not due to chance. Purely qualitative content analyses might make similar claims, but lack the statistical rigor to defend them. This kind of internal validation requires expert collaboration with a statistician, which was unavailable to me at the time.

**Limitations**

This work was wrought with frustration, challenge, and difficulty, which seem to mark most large-scale text analysis projects (Cermak, 2003). One of the most frustrating limitations in this method had to do with how language in a discourse works. It is rare, in example, for writers to describe issues, events, or objects in *exactly* the same way. Computers work best with specificity, not ambiguity, and so the lack of result for some keywords I expected to find was disappointing. In other words, some things are rather important to qualitative research—like unique ways of describing—but are not common and therefore, do not register as statistically significant or important quantitatively.

Another limitation of this method is that readers often bring meanings to a text, and as yet, there’s no way to code for latent meaning. Similarly, computers are poor readers of context, polysemy and irony, and, though our method rightly identified a majority of metaphorical frames, it was still necessary to make hundreds of individual judgments about metaphoric and
frame value, which made this project painstaking and calls into question one of the presumed affordances of computer-assisted text analysis: speed.

Ultimately, in a mixed method project like this, the primary investigator is responsible for analysis, which requires significant understanding of the historical context in which the texts were composed. When subject matter expertise is required to make sense of what was quantified, it raises other questions of value and validity: does big data text analysis reveal new insights or merely confirm what we could already know or learn from history or close reading? And, can we trust these insights?

**Human Error**

While the collection and processing of texts is systematic, a researcher has to give sufficient time to developing analytic criteria, which then must be strictly applied. When dealing with such large volumes of textual data, much of which must be converted and cleaned before it can be analyzed, data distortion is a constant worry. Though data management (file sharing, naming conventions, etc.) can be mitigated through the use of shared technologies, researcher judgment cannot, which is particularly true when the goal is high-level interpretation. Because of the unique viewpoint and expertise of the researcher, the analysis and its conclusions will also be unique. Another researcher analyzing these corpora may arrive at the same “counts” but radically different conclusions. While qualitative analysis allows for insightful evaluation of texts, it also leaves open the possibility of idiosyncratic interpretations and the “pull” of prior experiences and expectations, even among scholars dedicated to objectivity. Though thick description is appealing and often more insightful than the thin description of words reduced to numbers, counting is often perceived as more credible because it lacks the bias inherent in interpretive work.
Though interpretation and prescription may feel more satisfying, I think we have to resist the temptation to attempt a single, “correct” interpretation or to suggest that merely reframing will resolve the complicated socio-technical-political problem. We also have to continue to design our mixed methods studies to account for and correct for bias (Parnell, 2014).

Multiple Theoretical Perspectives

The dizzying array of theoretical approaches to applied text mining can confound the novice researcher at the start of her project. Perspectives from semiotics (Manning & Cullum-Swan, 1992), psychoanalysis (Biesecker, 1998), film genre theory (Dickinson, 2007), critical and cultural theory, (Tarlau, 2014), feminist theory (Thompson L. , 2013), literary theory (Fischer, 2003), and critical discourse analysis (van Dijk, 1991) have all served as theoretical frames for text mining and corpus linguistic work. Add to this the time it takes to learn the tools needed to conduct the analysis, often borrowed from semantic and social network analysis, computer science, graph theory, and visualization, and the multiplicity of available frames and methods becomes paralyzing. It is important to narrow the focus of any research project, simplify the theoretical paradigm within which one will work, and reduce complexity.

Problems Specific to Big Data

Earlier in this dissertation, I mentioned that “big data” projects were often defined by volume, velocity, and variety. Each of these qualifiers of big data projects poses unique challenges specific to working with large corpora of texts. There are many sources of text: transactional data, streaming data from social channels, and transcribed testimony to name a few. In the past, storing such volumes and varieties of data was problematic. Today, storage capacity has increased considerably so storage by itself is not so much a problem, but can still be prohibitively expensive for a researcher with limited research funding. Additional considerations
related to volume have to do with perpetual storage and curation. Though there are multiple storage options available, both cloud-based and physical, another possibility includes free, open-sourced, hosted platforms like Omeka\textsuperscript{14}. Omeka was designed for cultural institutions, academics and armchair historians who want to publish and host online collections and exhibitions of various sizes. Additional options include sending the data to digitally managed archives at other universities, like the Vietnam Center and Archive at Texas Tech University. Regardless of the solution chosen, managing data volume and curation is a difficulty specific to text mining projects of this kind.

Velocity is and continues to be a problem for researchers interested in working with digital texts, especially texts generated in real-time from social media sites. With RFID tags, sensors, smart metering and more than 340 million Tweets tweeted per day, torrents of data are available, but not always easily accessible, even to professors with research credentials. Access to the “fire hose” of data streaming from social sites like Twitter has, until recently, been severely limited and is often financially prohibitive, requiring third party vendors like Gnip, recently acquired by Twitter. The Library of Congress archival project for all Tweets and the Twitter Data Grants project are alternative routes to collecting data from this particular streaming social site, but reacting quickly enough or having access to resources to deal with the torrent of available data continues to be an ongoing problem for many researchers.

The variety of data is perhaps the most pressing challenge specific to big data. While structured, numeric data found in traditional databases is relatively easy to analyze, numeric data is the smallest of all available data sets. Most of the world’s data is stored as unstructured text. Texts are increasingly composed as emails, embedded in videos, and transcribed from audio formats. Accessing, managing, and merging varieties of data—textual and otherwise—from

\textsuperscript{14}http://omeka.org/
multiple sites of collection are serious problems in big data text research. Data reduction techniques are particularly important to text mining projects and sometimes the power of new tools overwhelms the importance of this practice. A tightly conceived research plan, a single question, and a well-toned corpus make it easier to observe the effects of methodological modifications on texts.

Data variety is also a challenge because even unstructured texts need to be matched, cleaned, and transformed into a consistent, machine-readable format before analysis. Some programs, like AutoMap and WordStat can read multiple file formats and have cleaning routines embedded in them, but even with these affordances, standardizing textual data can be particularly labor-intensive. Once these problems are addressed, data pulling, sorting, cleaning, and indexing routines can be automated, which makes lighter work of the data variety challenge.

Questions about Mixing Methods

Compounding problems of the multiplicity of theories, human error, and the technical challenges inherent in managing big data text projects are questions about mixing methods generally. The interest in mixed methods research is evident by the number and kinds of special workshops offered at summer institutes and disciplinary conferences in and across our discipline, but one of the questions these workshops and panel discussions inevitably raise is whether paradigms can be mixed at all, and if so, how well and to what new ends? The Digital Humanities continues to struggle to answer questions about definition, boundaries, and what actually “counts” as mixed methods research. Whether mixed methods research privileges postpositivism or which of the variable design possibilities are best suited to a particular research project, are also necessary questions to consider for researchers engaged in this kind of work. For
an excellent overview of the current controversies in mixed methods research, see Creswell’s
*Designing and Conducting Mixed Methods Research*.

As a discipline, Rhetoric and Composition—though less so Technical Communications
(Boettger & Palmer, *Quantitative Content Analysis: Its Use in Technical Communication*,
2010)—has been criticized for its weak or “mushy” methodologies (North, 1987). Though
qualitative analysis like discourse analysis can be methodical and painstaking work, even when
done well, all qualitative research is subject to claims of bias. Of course, quantitative methods
have also been dismissed on occasion as mere counting and a method of limited humanistic
value, and so I argue that a mixed-methods approach offers the richest possibilities for
methodological validity and insight.

Our lives are constructed and represented by the digital traces we leave behind in
cyberspace. These records—Internet keyword searches, online store purchases, Snapchats—are
increasingly archived in accessible and searchable databases. The shift toward Web 3.0 or the
semantic web provides a stream of new and untapped sources of discourse to analyze, but
requires both quantitative and qualitative research paradigms and skillsets.

**Conclusion**

In this chapter, I stated the questions this project attempts to answer and the steps taken to
try and get to those answers. I explained why using a custom Python script was necessary (to
parse text), and the features and benefits of WordStat for visualization and tabulation. I
concluded by explaining how I used semantic network analysis theory and network visualization
tools like Textexture to quickly “read” and thus validate quantitative findings. In the following
two final chapters, I explore the findings of this study and its limitations and the future iterations
necessary to overcome those limitations.
CHAPTER 4: RESEARCH FINDINGS

The following findings offer insight and answers to all three of the key research questions, but admittedly answer some questions better than others. Q1, which sought to identify the most common ways of describing key terms central to Agent Orange discourse, was satisfactorily answered. As hypothesized, each stakeholder told a different story about Agent Orange with some stories told in markedly different ways using unique descriptive and metaphorical phrases. It is in the moments of difference that it is easiest to tease out stakeholder frames, positions, and rhetorical aims and perhaps to see how alternative ways of framing might promote justice or more equitable political solutions. For example, that the Vietnamese News Agency persistently describes Agent Orange as a “poisonous chemical”, whereas the congressional corpora describe Agent Orange as an “herbicide” suggests an intentionally rhetorical framing of the issue is more or less negative terms.

The actors and issues at the center of each version of the Agent Orange story, as well as the ways of describing aspects of the narrative, suggest what stakeholder’s privilege in discussion and debate, as well as the ways they describe these topics and frame the details of their stories. For example, the Congressional testimony is strongly focused on claims, payments, and proving the human health consequences of exposure. It uses the most uncertain, probabilistic language of all the corpora with veterans and the Veteran’s Administration poised as key actors. Phrases that describe and reinforce scientific uncertainty are present throughout the text. Comparatively, the Vietnamese New Agency positions the “Vietnamese people” as its central actor, talks far more widely about the effects of dioxin on existing generations of children (the congressional testimony rarely talks about children and AO at all), and frames remuneration and remediation as an issue of justice, which is further metaphoricized in terms of war (as in the...
“fight” for justice). The language in the VNA uses the least amount of probabilistic language, also a rhetorical strategy that seems to suggest the Vietnamese have attempted to reframe the conversation away from scientific uncertainty, to the very certain and visible suffering of current generations of the disabled.

These and similar findings helped to answer Q2, which attempted to map the similarities and differences in the discourse on Agent Orange. This question could be further answered more reliably with better data, but the existing data demonstrates that the congressional and Times data is the most similar, and the Vietnamese data is, in some ways like the Times data but in most every way unlike any other corpora.

The third and final question was perhaps the most difficult to answer because the variables tracked in the corpora did not support a reliable answer. Question 3 asked how the differences or similarities in the various corpora explained or mapped onto shifts in public policy or practice. As mentioned elsewhere, to attribute shifts in public policy or practice with language, the study would have to change fundamentally. Rose & Baumgartner’s poverty study provides a workable model to follow. To clearly attribute language directly to policy shifts requires a third, calculable variable like speaker, party, or dollars. I would need to rely more on statistical models that could correlate remediation and remunerative spending, say, to the severity of the problem and the way the problem was characterized over time in the press. Rose & Baumgartner used gross domestic product and media coverage, but this same factor would not work for Agent Orange remediation/remuneration efforts. What might work is the amount of spending on Agent Orange from USAID or the VA across time, though these figures would be difficult to come by since USAID has found it difficult to directly fund Agent Orange efforts,
and instead channels dollars to AO victims by contributing to organizations that also those with other kinds of disabilities as well.

If I took this project in this direction, I could collect the frequency of articles published in US sources on Agent Orange and to calculate the coverage of the remediation effort, as one example, and then correlate that with policy/practice shifts and ways of describing the problem. Though I could not successfully answer Q3, I learned that constructing research projects with text as the only variable will yield thinner results—description of content, in example—than if the corpora are designed to track multiple correlative variables. Text mining projects are helpful because they are summative, but evaluative projects that might recommend a more immediate, prescriptive direction requires more than an analysis of content.

Originally, I anticipated analyzing the findings of 20 keywords, each of which represented larger themes within the corpora. Many of these terms were far too similar and I noticed early on that results were duplicative. To overcome this challenge, I simply whittled the 20 keywords/themes down to what I felt from close reading and network visualization were the five most important keywords and themes in the rhetorics of Agent Orange. These keywords/themes included the “environment”, “uncertainty”, “victim”, “remuneration, “ and “Agent Orange.” These keywords were selected because they were central to all stakeholder discourse and because they were suggestive of larger themes.

Theming is the bridge between the individual unit of text gathered and calculated and the researcher’s reading of it. Theming is an important step in this methodology because it allows the researcher to create a record of what is otherwise transient phenomena (Krippendorff K., 2004), or as Miles and Huberman said, “coding is analysis” (1994, p. 56). Assigning themes or codes is the process by which a researcher prepares to construct models that explain relationships
found among counts, testing models against empirical data, and generalizing findings. Themes are often vague constructs that researchers identify before, during and after data collection and are most often induced from the texts studied.

Though themes and frames are closely related, they are not the same thing; one informs the other or rather, one can frame a theme but one cannot theme a frame. As Quinn noted in her study that documents beliefs about American marriage, theming helps humans organize current knowledge and provides a framework for future understanding. It “exploit[s] clues in ordinary discourse for what they tell us about shared cognition—to glean what people must have in mind in order to say the things they do” (Quinn, 1982, p. 140). Themes are important to corpora because they help resolve the problem of intentionality mentioned earlier.

Foundational studies exploring theming specifically include Holland’s study of interpersonal problems in America (1985), Kempton’s (1987) study of American home heating control and Claudia Strauss’s (1991) study of Ciba-Geigy chemical plant workers’ ideas about capitalism. Strauss’s study is particularly illuminating. Her work focused on a disabled former Ciba-Geigy welder who now lives near the factory. She extracted three “discrepant cognitive schemas” during multiple interviews of him and described them thematically as (a) can’t fight the system; (b) achieve anything you want; and (c) responsibility to others. Each of these schemas is expressed through emotional and descriptive, even metaphoric, language and the voices are associated with specific behaviors or roles: the blue-color worker, potential capitalism, and patriot. Strauss discovered these schemas are not exact replicas of dominant discourses, but reworked versions of them. More on developing themes can be found in the two works by (Willms, Best, Gilbert, Wilson, & Singer, 1990) and (Miles & Huberman, Qualitative Data
Analysis, 1994). The themes identified and explored across the corpora are detailed in the following chapter on findings.

In addition to mapping four stakeholder communities and five keywords, I also traced three statistical measures. At 20 keywords, the volume of data became overwhelming, not only to map and organize, but to make sense of. The selected keywords yield important quantitative and qualitative insight, demonstrating not only the topics privileged in each discourse, but also how they were described and framed, and how they differed from, responded to, or reflected competing discourses and rhetorical demonstrations of power and resistance. The completed charts, which detail the top 10 keywords, phrases, and collocated terms are in the appendix. For ease of analysis, I’ve included only the top five words in each category. Further, I have interpreted each table, underscoring what I took to be the most interesting or insightful aspects of the final counts, considering my theoretical frame, and overall project goals.

The Environment

Agent Orange was designed to change the ecologies into which it was sprayed and so “the environment” is a central theme in all discourse about the chemical. In the “environment” cross-section of the corpora, “environment” is, predictably, the most frequently occurring keyword (for actual frequency counts, see appendix of charts). In and of itself this is not an interesting finding. In the above table the second most frequently occurring words across each of the stakeholder communities are health, dioxin, and Vietnam, respectively. These words are interesting because they signal different emphases in each stakeholder discourse about Agent Orange as it relates to the environment.
Table 4. The Top 5 Keywords per Stakeholder for Cross-section of Corpus: “Environment.”

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Environment</td>
<td>Environment</td>
<td>Not enough data</td>
<td>Environment</td>
</tr>
<tr>
<td>Health</td>
<td>Dioxin</td>
<td>X</td>
<td>Vietnam</td>
</tr>
<tr>
<td>Human</td>
<td>Toxic</td>
<td>X</td>
<td>Agent Orange</td>
</tr>
<tr>
<td>Public</td>
<td>Vietnam</td>
<td>X</td>
<td>Hanoi</td>
</tr>
<tr>
<td>Protect</td>
<td>Health</td>
<td>X</td>
<td>Dioxin</td>
</tr>
</tbody>
</table>

Note: in this and some following tables there were not enough texts, or there were not enough words/phrases for a particular category to count as statistically salient. That these counts are ‘0’ doesn’t mean the related words or phrases never occurred, only that they didn’t occur frequently enough to meet the minimum threshold set for inclusion, which was 10. These findings are significant because they are another way to point to find, suggest, or validate that certain corpora contained, framed, or privileged specific aspects of the Agent Orange debates while others did not. As pointed out elsewhere, the editorials almost never mention the environmental aspects of Agent Orange, focusing instead on veteran’s health issues and the toxicity of the chemical.

The Congressional rhetoric concerning the environment is concerned with the interplay between the environment and its effects on human and public health. It is also compelling to note that despite longstanding allegations of mismanagement, conspiracy and even corruption among the ranks of the CDC, U.S. Military, Veterans Administration, and chemical companies, the fifth most frequently occurring word in this cross-section of the corpus is “protect.” Because this section of the corpora focuses on the environment expressly, the word “protect” might lose a bit of its salience because of rhetorical prosody and preference, meaning when it comes to Congress talking about the environment, the terms “protect” and “human health and the environment” often co-occur. Whether one agrees that Congress has done all it could or should do to protect the environment from AO contamination, or to protect veterans affected by exposure, can be debated, but the frequency with which this term occurs is a clear indication of mission and position.

Studying the keyword frequency counts, one can see that the cross-section of New York Times articles about Agent Orange and the environment has an entirely different focus and frame than the Congressional corpus. This is interesting because it could suggest the “story” was
somewhere other than public and environmental health, or it could simply mean the *Times* wasn’t interested in publishing news set by Congressional agenda. Either way, the *Times* corpora privileges dioxin, toxicity, human health, and what’s going on in Vietnam, both then and now, evidenced by drilling into the various ways Vietnam is used with the keyword-in-context feature in WordStat. The *keyword-in-context* feature suggests what it, in fact, does: it highlights the key term as used in every instance of the text so the researcher can read how the term was used in context.

A quick look through the top ten keywords in the Congressional cross-section of the environmental discourse shows its main focus is protection of humans and clean up of the environment, whereas the *Times* is focused more on the inflammatory concepts and aspects of debate. Note the term “dioxin”—the toxic byproduct of 2,4,5-D and 2,4,T “Agent Orange”—does not appear in the Congressional discourse table, but is frequently identified as a top occurring keyword in the *Times* and *VNA* discourses. In many ways, in fact, the discourse on the environment and Agent Orange between the *Times* and the *VNA* is far more similar than either is to the Congressional discourse, which may be nothing more than an indication of newsworthiness and the fulfillment of editorial mission versus legislative mandate, or it may be a signal of rhetorical intentionality.

Table 5 details the top five descriptive phrases used in each discourse about Agent Orange and the environment. The story that the Congressional corpus tells is one of modification, danger, remediation, and protection. The *Times* and *VNA* cross-sections tell slightly different stories with the *Times* focused again on the news value inherent in protecting the environment and disagreement about how to do that, as well as the ongoing news stories generated by the genetically engineered crop movement led by Monsanto, one of the largest
Agent Orange manufacturers. I also want to note the presence of the descriptive phrase “entered the environment,” that speaks to the materiality of Agent Orange, agency, and responsibility.

**Table 5. Top 5 Descriptive Phrases per Stakeholder for Cross-section of Corpus:**
“Environment.”

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Protect Human Health &amp; the Environment</td>
<td>Agent Orange</td>
<td>Not enough data</td>
<td>Natural Resource &amp; Environment</td>
</tr>
<tr>
<td>Remedial Action</td>
<td>Protecting the Environment</td>
<td>X</td>
<td>Protect the Environment</td>
</tr>
<tr>
<td>Modify the Environment</td>
<td>Appears to be no agreement</td>
<td>X</td>
<td>Toxic Chemical</td>
</tr>
<tr>
<td>Biological &amp; Chemical Agents</td>
<td>Genetically engineered crops</td>
<td>X</td>
<td>Clean Up the Environment</td>
</tr>
<tr>
<td>Imminent &amp; substantial</td>
<td>Entered the Environment</td>
<td>X</td>
<td>Vietnamese People</td>
</tr>
</tbody>
</table>

Diving deeply into the Congressional corpus for a moment and looking at the ways stakeholders talked about Agent Orange, it becomes clear that ways of describing like “sprayed,” “introduced into,” “arise in,” and “unstable materials” are all ways of describing Agent Orange in the environment. These ways of describing do not occur frequently enough to make the frequency chart, but when combined and counted, they make up a significant portion of how Agent Orange is described and metaphorized—as though it were a self-defining, almost alien chemical entity. Agent Orange didn’t just “appear” in the jungle, and while its long-term human and public health effects may have been unknown, unstudied, or unintended, the chemical’s production, weaponization and distribution was not. The problem in describing dioxin as a mysterious, agentive chemical is that it understates the role and responsibility of manufacturers, distributors, and buyers.

This is one of the significant points of difference in this cross-section of corpora and points to both a stakeholder position and frame. The table above shows that the VNA discourse
includes the term “protecting the environment” but also includes the phrase “cleaning up the environment.” Cleaning up the environment suggests that a mess has been made and that someone is responsible for that mess. While the Congressional and even Times discourse rarely address responsibility, the VNA discourse frames Agent Orange and the environment as an issue of responsibility, morality, and justice, suggesting position, frame, and rhetorical aim.

It is also interesting to note that Agent Orange is not described most frequently as the more neutral “herbicide”, but negatively, even aggressively, as in the phrase “toxic chemical.” Another feature of the VNA corpus that is unique is the frequent occurrence of the term “Vietnamese people.” Perhaps this term simply reflects the country’s form of government (socialism), or perhaps it represents a strong sense of national identity. It certainly signals some kind of in-group identification, and is significant because it occurs in none of the other corpora. I think the key takeaway from this term’s frequency is that Agent Orange is experienced differently in Vietnam than it is in the United States. In America, Agent Orange affects a relatively small class of people. With the passage of the Agent Orange Act of 1991, American veterans had access to health and medical benefits; many fully disabled veterans receive compensation in addition to benefits. In Vietnam however, the government is too poor to support its citizens in the same way. There may be only one dioxin victim in a village or family, but that one victim can capitalize all financial and medical resources, and can break the poorest of rural families. Whether because of a lack of economic opportunity, or because of the cultural structures imposed in traditional village life, Agent Orange doesn’t just affect the victim or the victim’s immediate family, but the whole village or family unit.
Table 6. Top 5 Co-Occurrences per Stakeholder for Cross-section of Corpus: “Environment.”

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Effects</td>
<td>Environment</td>
<td>Not enough data</td>
<td>Environment</td>
</tr>
<tr>
<td>Long</td>
<td>Dioxin</td>
<td>X</td>
<td>Dioxin</td>
</tr>
<tr>
<td>Action</td>
<td>Toxic</td>
<td>X</td>
<td>People</td>
</tr>
<tr>
<td>Sites</td>
<td>Public</td>
<td>X</td>
<td>Chemical</td>
</tr>
<tr>
<td>Protect</td>
<td>Environmental</td>
<td>X</td>
<td>Vietnamese</td>
</tr>
</tbody>
</table>

Finally, the co-occurrence table above depicts the top five words that co-occur in each cross-section of discourse related to the “environment.” As mentioned earlier in this chapter, co-occurrence metrics are important because they can—though do not always—suggest a relationship between words that is not always obvious. In example, as the table above shows, there is a relationship in the Congressional text between the words “environment” and “effects” unseen in the previous table of frequencies alone.

Here is an additional point of evidence that suggests the U.S. government and military were deeply concerned over the possible long-term effects of Agent Orange in the environment. It is also interesting to note that “action”, “toxic”, and “people” are the third most frequently co-occurring words with “environment”—each suggesting a way of framing the discourse that we will see consistently throughout the corpora in different cross-sections. Generally and importantly, Congressional testimony tends to frame the topic in terms of what to do; the *Times* tends to frame issues in terms of controversy; and the *VNA* tends to frame issues in terms of who is affected.

There is a Vietnamese proverb that says, “Human life is like a flower, here and open in the morning, closed and gone by night.” This correspondence between human life and the natural world is not uncommon in Vietnamese literature (Waugh, 2013, p. 243). Much of the folk wisdom in Vietnamese literary tradition arises from observations about the natural world, as in
“people have ancestors, like trees have roots, like rivers have springs” (Waugh, 2013, p. 244).

Likening ancestors to tree roots suggests a belief in a larger cosmic whole in which natural things have spiritual equivalents. This isn’t to suggest the Vietnamese traditionally saw themselves as some idealized “one” with nature, but that they viewed humanity as part of a larger whole that works best when in balance. There is “correspondence” between the human and natural worlds (Waugh, 2013, p. 243) and so it is not, perhaps, surprising to see “the Vietnamese people” frequently co-occurring with “environment.”

The glaring absence in this cross-section of the corpora, as well as the next, is the lack of textual information from the New York Times editorial base. As mentioned elsewhere, in some cross-sections of the corpora, the data was too thin, but the absence of textual data in some instances (as in this instance) is a finding in itself. The New York Times editorial corpus is composed of a little less than 100 articles and the content of those editorials is consistent in its themes. The content focuses almost exclusively on the plight of U.S. veterans and their struggle to win reparation through the American legal system for their exposure to Agent Orange. The lack of attention to Agent Orange’s environmental effects does not necessarily suggest editorial writers were unaware or unconcerned with AO’s environmental effects in Vietnam—though that might be true. More likely however, the focus on veterans and the contentious and complicated political and legal struggle for recognition and recompense was the result of editorial policy. Either way, the editorials never speak to environmental issues related to Agent Orange at home or abroad in the years surveyed.
The Uncertainty of Agent Orange

Table 7. Top 5 Keywords per Stakeholder for Cross-section of Corpus: “Uncertainty.”

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Uncertainty/Uncertain</td>
<td>Uncertain/Uncertainty</td>
<td>Not enough data</td>
<td>Not enough data</td>
</tr>
<tr>
<td>Exposure</td>
<td>Dioxin</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Probability</td>
<td>Chemical</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Report</td>
<td>Case</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Risk</td>
<td>Effect</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

If one were to use a single word to characterize the debates about Agent Orange, at least in the United States, it might be uncertainty. The discourse suggests rhetors were uncertain about many things—levels and duration of exposure, the longevity of dioxin persisting in the soil, short and long-term human health consequences, and potential birth defects. The most uncertainty is expressed in the Congressional corpus and it is not just the volume of words or phrases that signal uncertainty, but what rhetors were uncertain about. For example, a close reading of the articles identified as having uncertain language notes that the VNA discourse of uncertainty revolves around how families will afford care, whether claims will be heard or dismissed in US courts, and when and to what degree US NGOs will provide relief.

Alternatively, probabilistic language related to exposure interleaves the Congressional corpus. Words like “probability”, “report” and “risk” suggest not only the depth of the uncertainty that beleaguered Congressional debate, but also the kinds of proof Congressional and military leaders sought and valued as they tried to mitigate that uncertainty.

The Times corpus reports on Congressional uncertainty, or, more specifically the uncertain outcome for veterans pursuing legal action against chemical companies. Whereas the
Congressional corpus leverages the language of uncertainty to sometimes hedge politically, the *Times* seem to report on the uncertainty of milestones, events, and happenings. It is worth noting too that the term “dioxin” is frequently found in discussions of uncertainty, further constructing the myth of AO-dioxin as an unpredictable, Frankenstein chemical agent.

Again, rhetorical silences are results. Notice the *Times* editorials and the *Vietnamese News Agency* have no data in the table above. This supports the findings of a sentiment analysis that characterized these corpora based on the frequency counts of words of positive, negative, and “uncertain” valences (Hopton, 2014). Though there could be many explanations for this, the most likely explanation is two fold. First, to focus on the uncertainty of Agent Orange/dioxin science is to shift the attention away from the rhetorical goals of advocating for financial support and remediation. Second, the absence of hedging or uncertain language is in itself a way of framing the issue. To the Vietnamese, the links between Agent Orange/dioxin and their physical maladies is not in question. Spending time debating the science of exposure defeats their rhetorical aim: to secure medical and financial relief and environmental remediation from those they hold responsible for their illnesses and ecological imbalances.

**Table 8. Descriptive Phrases per Stakeholder for Cross-section of Corpus: “Uncertainty.”**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Scientific Uncertainty</td>
<td>Remains Uncertain</td>
<td>Not enough data</td>
<td>Not enough data</td>
</tr>
<tr>
<td>Health Effects</td>
<td>Relieve Symptoms</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Exposure to Herbicide</td>
<td>Health Effects</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Risk Assessment</td>
<td>Contends that troops did not enter</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Exposure Estimate</td>
<td>GAO report released</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

The above table supports many of the observations and extrapolations made from the keyword table about uncertainty. As noted, the uncertainty for Congress is a different kind of
uncertainty than it is for the *Times*. For Congress it is the very specific “scientific uncertainty” that threatens to topple political and legislative negotiations. Though all AO “effects” are concerning, when it comes to uncertainty the *most* uncertain effects—and thus the hardest to deal with—are the human “health” effects. This observation is perhaps supported by the fact that environmental remediation in Vietnam is underway while collaborative medical support and research between Vietnam and the US government is minimal.

“Exposure”, “exposure estimates”, and “risk assessments” speaks to the language of risk management, but it is the inherent risk in such undertakings that seems to be what the *Times* talks about when it talks about uncertainty. These ways of describing—“remains uncertain”, “contends that”, and even the “GAO report released”—amplifies the uncertainty and skepticism of the Congressional discourse.

Table 9.5 Co-Occurrences per Stakeholder for Cross-section of Corpus: “Uncertainty.”

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Uncertainty</td>
<td>Case</td>
<td>Not enough data</td>
<td>Not enough data</td>
</tr>
<tr>
<td>Exposure</td>
<td>Chemical</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Probability</td>
<td>Company</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Report</td>
<td>Court</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Risk</td>
<td>Dioxin</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

Finally, the co-occurrence table above confirms what the keywords suggested: the Congressional discourse on Agent Orange is full of hedging, uncertain language. The term “uncertainty” co-occurs with “exposure,” “probability,” “report”, and “risk” most frequently in the Congressional discourse, but in the *Times* discourse the most frequently co-occurring word is “case.” This suggests the outcome of the veteran’s case against chemical manufacturers, but also its case for reparation from the U.S. government was described with great uncertainty.
Victims & Victimization

In the Congressional and *Times* corpora, the term “suffer” occurs frequently and seems to be a case of semantic preference, as it is most often used to describe a state of victimhood, as in “AO victims suffer…. As noted elsewhere, the Congressional corpora, when analyzed as a whole, are focused on what can be done: providing better service and improving human health outcomes. Interestingly, sufferers are more likely described as “veterans” primarily, and “victims” secondarily. Conversely, the *Times* remain focused on exposures suffered by the Vietnam Veteran. The editorial section of the corpus continues to focus on suffering as it is associated with the legal battle for justice, a battle many veterans felt they had to fight multiple times across multiple fronts—first with Congress for recognition and protection, then with chemical manufacturers for compensation, and finally, with their bodies, as they battled symptoms and debilitation.

Table 10. Top Keywords per Stakeholder for Cross-section of Corpus: “Victim.”

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Suffer</td>
<td>Suffer</td>
<td>Victim</td>
<td>Victim</td>
</tr>
<tr>
<td>Veteran</td>
<td>Victim</td>
<td>Suffer</td>
<td>AO</td>
</tr>
<tr>
<td>Victim</td>
<td>Veteran</td>
<td>Case</td>
<td>Vietnamese</td>
</tr>
<tr>
<td>Service</td>
<td>Exposure</td>
<td>Lawyer</td>
<td>Dioxin</td>
</tr>
<tr>
<td>Health</td>
<td>Vietnam</td>
<td>Veteran</td>
<td>Hanoi</td>
</tr>
</tbody>
</table>

Though the term “victim” and “suffer” appears in all four of the keyword lists, the prevalence of the term is unique among them. Suffer is the most frequent keyword in the Congressional and *Times* corpora, but “victim” is the top occurring keyword in the two corpora that most closely represent those that were victimized. Another way of looking at this, which supports an earlier observation about how each of the corpora work rhetorically, is to say that the
Congressional corpora is focused on what can be done; the Times on what is being done; and the editorial and VNA corpora on to whom the action is being done.

The Vietnamese News Agency is, as in other cross-sections of the corpus, predominantly focused on “the Vietnamese people” and their access to service and support through government channels, most of which are based in the capital city of Hanoi. What’s particularly interesting in this table is that the terms “AO” and “dioxin” appear consistently in the Vietnamese discourse, but not in the other discourses. The shorthand AO is almost never used in any discourse except the VNA discourse. Though the purpose of this is uncertain, generally, shorthand or abbreviations and acronyms are another way to rhetorically include or exclude others and thus may be another expression of identity or solidarity.

Table 11. *Top 5 Descriptive Phrases per Stakeholder for Cross-section of Corpus: “Victim.”*

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Veterans Suffer</td>
<td>Vietnam War</td>
<td>Victims have Agent Orange</td>
<td>Victims have Agent Orange</td>
</tr>
<tr>
<td>Health Care</td>
<td></td>
<td>Ranch Handers</td>
<td>AO victim</td>
</tr>
<tr>
<td>Mental Health</td>
<td>Exposure to Agent Orange</td>
<td>Document the financial harm they suffer</td>
<td>Vietnamese victim</td>
</tr>
<tr>
<td>Suffer From PTSD</td>
<td>Birth Defects</td>
<td>Suffer from chloracne</td>
<td>Vietnam Association</td>
</tr>
<tr>
<td>Service Connection</td>
<td>Compensate Victims</td>
<td>Lose their job</td>
<td>Chemical Company</td>
</tr>
</tbody>
</table>

Here, the ways various stakeholders describe victims and kinds of victimization further shapes our understanding of stakeholder conceptions of the term. For Congress, understandably, it is the physical and mental suffering of veterans that requires action. We see in this graph the introduction of a new term—PTSD, which may seem like a curious occurrence except that PTSD and Agent Orange not only share many symptoms, but also were two maladies expressly
associated or attributed to a veteran’s experience in Vietnam. Over time, these Agent Orang and PTSD have become symbolic of the Vietnam veteran and the tragedies of war.

As mentioned earlier in this dissertation, Agent Orange is not a recognized disease class. It is even unlike its nearest neighbor PTSD in that it is neither a psychological nor a biomedical condition, and yet, “Agent Orange” is talked about here as something one “has.” “Agent Orange” is what victimizes. Of course, medically this is not true. It’s the dioxin in Agent Orange that is the human health hazard and so it’s worth noting again the infrequency of the use of the term “dioxin” in all but the VNA corpus. Here too, it should be pointed out that for the first time the term “birth defects” is talked about frequently enough to be of statistical significance. Notice too that birth defects do not appear with any frequency in the Congressional corpus, and are thus not associated with victimization in that corpora, but do appear in the Times corpora. In this table it is clear that victims are those who should be “compensated”, those who “suffer chloracne” and those who are victimized by “chemical companies.”

Victimization is an important term to discuss in the rhetorics of Agent Orange especially for veterans and the Vietnamese. Here, in talk about victimization, the Times editorial corpus has much to say. In addition to Agent Orange being something one suffers, it is also curious to note the primary concerns of the Times editorials: the financial burden of AO-related medical expenses, and the difficulty on one’s family after losing a job due to illness. The VNA corpus—more than in any other—attempts to attribute responsibility for victimization in as much as it attempts to remind readers dioxin poisoning is real and suffering is ongoing, hence the presence and frequency of terms like “chemical companies.” Victims are in need of support, offered by many organizations but predominantly VAVA, which is why that term (Vietnam Association of Victims of Agent Orange) features in the VNA discourse, but not in others.
Finally, the words that most frequently co-occur are similar but not exactly the same as those in the keyword table. In example, the term “exposure” more frequently co-occurs in the Congressional text with the word victim, as in “a victim of exposure to Agent Orange”, but exposure does not frequently co-occur with the term victim in the Times editorials or VNA corpora. In the Congressional and VNA corpora, victims “suffer”, but the emphasis is on support in the VNA corpus and “service” in the Congressional. Finally, the terms “dioxin” and “AO” co-occur frequently with the term victim, but neither term appear in the Congressional or Times corpora at all. There is a specificity and conviction in the way that the VNA calls victims specifically AO or dioxin victims that is most likely another rhetorical move that redirects the debate away from scientific uncertainty (which stalls progress) and toward support and action.

**Remuneration**

Table 13. *Top 5 Keywords per Stakeholder for Cross-section of Corpus: “Remuneration.”*

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Claim</td>
<td>Claim</td>
<td>Claim</td>
<td>Disabled</td>
</tr>
<tr>
<td>Veteran</td>
<td>Veteran</td>
<td>Benefit</td>
<td>Child</td>
</tr>
<tr>
<td>Benefit</td>
<td>Benefit</td>
<td>Compensation</td>
<td>People</td>
</tr>
<tr>
<td>Record</td>
<td>Compensation</td>
<td>Veteran</td>
<td>Disability</td>
</tr>
<tr>
<td>VA</td>
<td>Record</td>
<td>Feinberg</td>
<td>Victim</td>
</tr>
</tbody>
</table>
As explained elsewhere in this work, remuneration is a central theme in the discourse on Agent Orange. The debates about remuneration are long standing and generally focus on issues of who gets remunerated, for what, how long, who pays, and whether benefits of remuneration extend to the families of veterans and victims. Remunerative strategies have a long history in the discourse on Agent Orange, beginning in the 1960s when certain farmers, villagers and rubber plant owners were compensated for losses attributed to Operation Ranch Hand and other spray campaigns. The term “claim” in this context is not something one does, but an action one takes as in to “file a claim.” As we have seen previously, the Congressional corpus seems to emphasize action, and so it is of little surprising to find key words that describe the claims process or the military records needed to prove claims.

Similarly, the cross-sections of Times discourse on remuneration maps almost perfectly to the Congressional discourse with one minor exception—the use of the term “compensation.” So too does the editorial corpora map closely onto the other two corpora with one exception—the inclusion of the name Feinberg. Kenneth Feinberg appears as an important actor because he was the court-appointed Special Settlement Master in several cases, but specifically, in cases related to Agent Orange product liability. Since I have already established that the editorial text frames Agent Orange generally in terms of legal battles, it makes sense that Feinberg’s name would be included.

What is perhaps most interesting about this table though comes from the VNA corpora and its connection of the term “disabled” and “disability” with remuneration, and its inclusion of the term “child.” With data that reached back further than 2006 one might be able to trace the development and use of the term “disabled.” My suspicion is that this term was used later in the discourse after U.S. veterans won their compensation claims with both the government and
against chemical companies. All I can definitively say about the term “disabled” here is that it occurs more frequently and sooner in the VNA corpora than in the others and thus it is an important term and shifts the emphasis or frame away from victimization, which perhaps demands apology as recompense, to disability, which more often demands compensation.

As we’ve seen elsewhere, “birth defects” are rarely mentioned in the Congressional corpus, and only sometimes mentioned in the Times corpora. This lack of emphasis or consideration on possible birth defects related to exposure may be one reason there is such a dearth of government-funded medical data on dioxin’s teratogenic effects. Of course, the presence of frequent references to “child” or “children” or “families” in the VNA corpus is not just rhetorical, but also topical or pragmatic, as new generations of children are born with birth defects attributed to contaminated soil at former military bases.

Table 14. Top 5 Descriptive Phrases per Stakeholder for Cross-section of Corpus: “Remuneration.”

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Claims Process</td>
<td>Veterans Affairs</td>
<td>Investigate the claim</td>
<td>Disabled child</td>
</tr>
<tr>
<td>Disability Claim</td>
<td>Exposure to Agent Orange</td>
<td>Education Benefits</td>
<td>Disabled people</td>
</tr>
<tr>
<td>Regional Office</td>
<td>Vietnam War</td>
<td>Veterans claiming for the first time</td>
<td>Children with disabilities</td>
</tr>
<tr>
<td>Veteran Benefit</td>
<td>Disability Claim</td>
<td>Laugh out of court</td>
<td>People with disabilities</td>
</tr>
<tr>
<td>Veterans Affairs</td>
<td>Disabled Veteran</td>
<td>Reluctance to consider claim</td>
<td>AO victims</td>
</tr>
</tbody>
</table>

In a cross-section of descriptive phrases related to remuneration, we would expect to find administrative talk like “regional office”, “veteran benefits” and “disability claims” from the Congressional text but it is worth noting that talking about veterans as “disabled” is a frame. Drilling deeper into this section of the corpora reveals that though disability is a frequently
occurring term, the ways Congress describes disability is different from how the VNA or Times describe it. There are service-connected disabilities and the “totally disabled veteran”; arguments over the “degree of disability” and “potential claimants.” Comparatively, the term does not appear in the Times editorial corpus at all.

The more interesting sections of this table come from the editorial section of the Times. This section of the table demonstrates in no uncertain terms the difficulty veterans, families, and their advocates faced and continue to face to win recognition and the benefits of “presumptive exposure”, as indicated in frequently occurring phrases like “laughed out of court” and “reluctance to consider claims.”

**Table 15. Top 5 Co-Occurrences per Stakeholder for Cross-section of Corpus: “Remuneration.”**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Record</td>
<td>Claim</td>
<td>Claim</td>
<td>Disabled</td>
</tr>
<tr>
<td>Claim</td>
<td>Veteran</td>
<td>Benefit</td>
<td>Child</td>
</tr>
<tr>
<td>Veteran</td>
<td>Benefit</td>
<td>Compensation</td>
<td>People</td>
</tr>
<tr>
<td>Benefit</td>
<td>Compensation</td>
<td>Feinberg</td>
<td>Disability</td>
</tr>
<tr>
<td>Compensation</td>
<td>Record</td>
<td>Court</td>
<td>Victim</td>
</tr>
</tbody>
</table>

Remuneration is an interesting thread running through the Agent Orange story. Americans have long since been united in the belief that veterans deserve honor and respect, but they have also been somewhat divided when it comes to answering the question of what a government owes its soldiers. One interesting finding observed in this section of co-occurrence data is that the Congressional record, the Times articles and editorials all use the same bureaucratic or technical jargon to talk about remuneration. There are far fewer human actors in this narrative than in the VNA discourse, which has terms like “child”, “people”, and “victim” all co-occurring with remuneration frequently, as opposed to “record,” “claim”, and “benefit”, which co-occurs more frequently with “remuneration” in the Congressional and Times corpora.
Part of what makes the remuneration debates so rhetorically challenging for victims and advocates in Vietnam, is that the debates about remuneration in the States have long since been tortured too. In a 1946 Gallup poll, a sample of American veterans were asked if they believed the U.S. government had compensated them fairly. A majority of veterans who fought in the first and second world wars believed they had (75% and 69% respectively), despite the fact that they received almost nothing compared to what today’s veterans receive in housing, tuition assistance, and medical benefits.

In 2011, the question was asked again of a new veteran class and again 61% agreed that the U.S. government had indeed cared for its veterans well. The general public however, disagreed. About a third of the public polled in the 1947 survey thought benefits were not adequate, but nearly 60 percent thought they were inadequate by 2012. This divide is even more pronounced for Vietnam veterans. A 1979 poll showed that close to one-third of Americans believed Vietnam veterans had been denied just and fair compensation, or had been shorted benefits compared to veterans of the world wars and Korea, even though in practice, they had been afforded more than their predecessors. And, a 1990 Gallup poll found 64% of those surveyed believed Vietnam veterans had been severely mistreated since the war’s end (Roper Center, 2014). Much of this sentiment surrounding the treatment of Vietnam veterans certainly has to do with the way Agent Orange remuneration was handled or rather, mishandled. Though there is no polling data to document sentiment or belief about remuneration to veterans suffering AO-dioxin-related illnesses, previous chapters show that the Agent Orange Act of 1991 was a political decision rather than a decision based on scientific certainty. The Act most certainly would not have passed were it not for the public pressure placed on Congressional leaders to rectify what was perceived as a debt owed the men and women they sent off to war.
The Agent Orange Act of 1991 is the only significant piece of legislation underwritten to address the needs of veterans presumed to suffer the effects of dioxin exposure. It was sponsored by Representative Sonny Montgomery (D) through the House Veteran’s Affairs Committee (Public Law No: 102-4) and presumes several diseases like non-Hodgkins lymphoma and chloracne were the result of exposure to the dioxin in the weaponized version of the herbicide Agent Orange and it took more than 20 years after Paul Reutershan took his story to the Today Show to pass into law.

Ways of Describing Agent Orange

Table 16. Top 5 Keywords per Stakeholder for Cross-section of Corpus: “Agent Orange.”

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Study</td>
<td>Vietnam</td>
<td>Vietnam</td>
<td>Victim</td>
</tr>
<tr>
<td>Exposure</td>
<td>Veteran</td>
<td>Veteran</td>
<td>Child</td>
</tr>
<tr>
<td>Veteran</td>
<td>Defoliant</td>
<td>Expose</td>
<td>Dioxin</td>
</tr>
<tr>
<td>Effect</td>
<td>Chemical</td>
<td>Chemical</td>
<td>War</td>
</tr>
<tr>
<td>Herbicide</td>
<td>Study</td>
<td>Study</td>
<td>Support</td>
</tr>
</tbody>
</table>

Throughout this dissertation I have argued that what we call subjects and objects directly affects our legislative, political, and practical treatment of them. If that is true, then perhaps the most important word to analyze in the debates about Agent Orange is “Agent Orange.” Looking closely at the table above shows that, as hypothesized, different stakeholders name and thus frame Agent Orange in different ways. In the Congressional corpus, Agent Orange is most often described as an herbicide; in the Times article corpus, it is a defoliant; in the editorial corpus it is a chemical, and in the VNA corpus it is a dioxin. Because of the frequency in which these terms appear in the corpora, one cannot assume these linguistic choices were merely pragmatic or arhetorical because all of them carry different descriptive or metaphorical charges and psychological weights. The term “herbicide” is somewhat neutral—a term with which Americans
are somewhat comfortable since they use it on their gardens and laws. The term chemical carries a far more negative emotional charge. Chemicals are often non-natural, they are produced by big business and they are responsible for significant environmental corruption and damage. Worst of all though is the term dioxin, a known carcinogen and teratogen.

As already demonstrated, when political and military leaders talked about Agent Orange, they talked about veterans, exposures, effects, and studies. In *Times* articles and editorials veterans, studies, and chemical consequences were keywords, making up the bulk of the content, and in the *VNA* corpora, Agent Orange is talked about in relationship to the victims, children, and the war that created it.

**Table 17.** Top 5 Descriptive Phrases per Stakeholder for Cross-section of Corpus: “Agent Orange.”

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposure to Agent Orange</td>
<td>Vietnam War</td>
<td>Exposure to Agent Orange</td>
<td>Victims have Agent Orange</td>
</tr>
<tr>
<td>Health effects</td>
<td>Exposure to Agent Orange</td>
<td>Birth Defects</td>
<td>Vietnamese Victim</td>
</tr>
<tr>
<td>Agent Orange Study</td>
<td>The defoliant Agent Orange</td>
<td>The Truth about Agent Orange</td>
<td>Affected by Agent Orange</td>
</tr>
<tr>
<td>Ranch Hand</td>
<td>Chemical Company</td>
<td>Chemical Companies</td>
<td>AO Victim</td>
</tr>
<tr>
<td>Veterans exposed to Agent Orange</td>
<td>Birth Defect</td>
<td>Vietnam War</td>
<td>American War</td>
</tr>
</tbody>
</table>

The ways of describing Agent Orange and thus framing larger debates about it tend to follow rhetorical patterns already established. Congress discusses exposures and the science of effects; it describes studies and exposure levels and conditions, the Ranch Hand operation and administration of benefits. It does not talk about responsibility or justice, and there is never an apology.
The *Times* also talks about veteran exposures, and also attempts to talk about what’s in the rhetorical silences—birth defects and the role and responsibilities chemical manufacturers had/have to veterans and the Vietnamese. The *Times* editorials also highlight the links between exposure and birth defects, but minimally. They too are concerned with holding chemical companies responsible, but it seems, at least from the above table, what they want more than compensation is validation. The “truth” about Agent Orange was a halting and haunting find.

When the *VNA* talk about Agent Orange they talk about its “victims”—both past and present—and they treat Agent Orange as communicable, something one can “get.” Little of the uncertainty that afflicts the Congressional corpora is present in the *VNA* text, though they too are concerned with how victims are affected by the chemical that stripped their jungles bare during the “American War.”

**Table 18. Top 5 Co-Occurrences per Stakeholder for Cross-section of Corpus: “Agent Orange.”**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Study</td>
<td>Veteran</td>
<td>Vietnam</td>
<td>Victim</td>
</tr>
<tr>
<td>Veteran</td>
<td>Vietnam</td>
<td>Veteran</td>
<td>Child</td>
</tr>
<tr>
<td>Expose</td>
<td>Defoliant</td>
<td>Expose</td>
<td>Dioxin</td>
</tr>
<tr>
<td>Vietnam</td>
<td>Chemical</td>
<td>Chemical</td>
<td>War</td>
</tr>
<tr>
<td>Exposure</td>
<td>Study</td>
<td>Study</td>
<td>Support</td>
</tr>
</tbody>
</table>

Little can be said of the above co-occurrence table that has not already been observed and explored elsewhere, except that where the Congressional and U.S. media stakeholders talked about Agent Orange as a chemical to be studied, the *VNA* talk about Agent Orange as a chemical that victimized and these victims, many of them children, need support, not more study.

To conclude this chapter, I have included four network visualizations of each, complete corpora for each stakeholder beginning with Congress. In the network representations below, the size of the node denotes its relative importance in the network and within its cluster or
“community” of contextual terms. It is important to note that nodes are not linked because they are next to each other in a sentence, but because they are central to a “window” of context-dependent words. This “node-edge” structure is encoded and visually represented as a graph using the open source network visualization tool Gephi. These and similar network visualizations depend on Gephi metrics and community detection algorithms.

The size of the nodes in the network graphs depends on the network calculation called “betweeness centrality,” a standard network measure that is equal to the number of shortest paths from all vertices to all others that pass along the path through to the node. A node with high betweenness centrality is influential in a network because one assumes that item (which could be knowledge, semantic meaning, etc.) transfer follows the shortest path (Freeman 1977). In semantic network analysis, betweenness centrality is an important measure because those words that measure high on this scale often appear at the junctures of meaning. A “juncture of meaning” is an area of the graph that visualizes a significant relationship—in rhetorical terms, an exigency. In other words, the nodes at a juncture of meaning do not just occur frequently, they are also influential. Nodes can have high betweenness centrality within clusters of concepts too, making them influential terms within a specific context or subsystem of ideas. As stated in Chapter 3, the following network visualizations not only help to “distantly read” each corpora, but also help to confirm the quantitative findings.

Though I did not calculate or include network measures, semantic network visualizations like these are still helpful to those who conceptualize and sense make visually. It is further helpful to visualize texts as networks and look for outliers, those words or phrases that appear in the network, but not in frequency tables. Below (Figure 5) is the network visualization representing the Congressional corpus, which, even at the highest level of visualization, as
featured here, demonstrates and tells the dominant narrative running through the Congressional corpus. As has been shown elsewhere, the Congressional narrative is one wherein the Vietnam Veteran was exposed to herbicides and the main controversy was over studying these exposure effects.

Figure 5. Congressional Corpora Network Representation
The entire network of the Congressional corpus represented as a network using Textexture. The relative importance of the node is denoted by size. Nodes of the same color, or nodes connected to same-colored nodes, denote a context.

Figure 6 (next page) displays the network from the New York Times, and also supports the findings of this chapter with the most influential keywords in the Times corpora being, “Vietnam”, “veteran”, “Agent Orange”, “chemical” and “defect” (as in birth defect).
Figure 6. New York Times Corpora Network Representation

The entire network of the News York Times article corpus represented as a network using Textexture. The relative importance of the node is denoted by size. Nodes of the same color, or nodes connected to same-colored nodes, denote a context.

Again, with the New York Times editorial corpus, the visualization maps to findings with the most influential keywords in the Times editorial corpora being, “Vietnam”, “veteran”, “Agent Orange”, “compensation”, “court” and “claim.” This visualization also highlights a potential outlier, or a word that did not seem to appear frequently in the corpora and that’s the term “year.” Drilling deeper into the text, one finds in fact that, though it wasn’t counted and though it does not occur in any other corpora, there is a theme of “time and waiting” that appears in the editorial texts. This theme signals the frustration experienced by veterans and their families with Congressional inaction and the years it took to win recognition and compensation. Time is talked about literally, but also metaphorically. Remember that veterans and many veteran groups called Agent Orange “the wasting disease” and throughout the editorial corpora there is a kind of metaphorical ticking of the clock, evidenced in phrases of low frequency but phrases that occur
nonetheless like “waiting for an answer;” “waiting for the judge to decide”; “waiting for an army to die”; “years in the making”; “waiting for years” and so on.

Finally and perhaps most stunningly, Figure 8 on the next page represents the network of texts included in the Vietnamese News Agency corpora. Further validating and visualizing the findings of this chapter, all roads—metaphorically speaking—lead to victimization. Other keywords of import include “American war”, “Vietnamese”, “Vietnam”, and Agent Orange.”

In the following and concluding chapter to this work, I offer a summative assessment of the most significant findings, as well as ways to rework the data to its highest and best use, and several opportunities for collaboration and extension of this method on texts unfolding in real time.
The entire network of the Vietnamese News Agency corpus represented as a network using Textexture. The relative importance of the node is denoted by size. Nodes of the same color, or nodes connected to same-colored nodes, denote a context.
CHAPTER 5:
INTERVENTION, VISUALIZATION, AND THE FUTURE PROSPECTS OF BIG DATA
AND AGENT ORANGE

Intervention & Social Change

Throughout this thesis, I have argued that words matter; that words shape the values, attitudes, and beliefs of the rhetors that use them to talk about cultural and political objects and issues. What I have not discussed is what happens after the ways we describe and frame are discovered. As activists, what do we do with this data once it’s shown that the various stakeholders in the debate over Agent Orange privilege different subjects and describe and frame those subjects is different and perhaps conflicting ways? The answer to this question is not uncomplicated.

The first part of the answer is actually another question: if key stakeholders were conscious of the effects their linguistic choices might have over policy and practice, would they have made different choices? If we cannot change what we do not recognize as needing change, then this study was the first and necessary step in visualizing the similarities and differences in what key stakeholders talk about when they talk about Agent Orange. Mapping these phrases, positions, and frames alone is helpful, but the results of this study can, with minor tweaks to the research design, have more lasting impact and serve as a more meaningful intervention. A study by Richard Rogers and Anat Ben-David might be instructional here.

In the debates over security between Israel and Palestine, various words are used to describe what appears to be nothing more than a physical object—a “fence.” But, in investigating
the different ways stakeholders use this term, and the different phrases used to describe it, the investigators found that these words were not merely linguistic or pragmatic, but were rhetorical—they were covert ways of expressing political positions. The term for the fence in the official Israeli discourse, in example, is “security fence” but in the Palestinian discourse it is called the “apartheid wall.” This maps somewhat similarly to the Agent Orange discourse where Congress adopted the term “herbicide” to describe Agent Orange and the Vietnamese adopted the term “chemical”—both of which subtly suggest a position and judgment. Herbicide feels less invasive than the term chemical, which, especially applied in the context of war, has seriously negative connotations. “Herbicide” and “chemical” as well as “apartheid wall” and “security fence” are contextually charged notions that literally describe a defoliant and a construction project, but through the magic of language, also subtly describe a position.

The “wall” study used data gathered from Google News, which includes official, nongovernment organization and news stakeholder data, and compared the ways and situations in which various terms were used, to create a kind of “conflict indicator” based on which phrase or term was adopted and used in which discourse at what time. Researchers found the Palestinians and Israelis use their words differently with Israelis consistent but virtually singular in their term use and the Palestinians adopting different phrases based on context or setting. The study demonstrated that in large text mining projects like this, “setting” is a critical variable to trace alongside content. The significant takeaway from “the wall” study is that Israeli and Palestinian actors “come to terms”—meaning they are most diplomatically conciliatory—when the term “separation wall” is used. This term couples the Israeli left-of-center adjective “separation” with the Palestinian noun “wall” and implies, according to the authors, a peace-related arrangement that is unique from either side’s official position.
I use this study as an example for two reasons: first, the study demonstrates the power of a similar text mining method to affect social change. Though offering an interventionist strategy or suggesting how to reframe is really beyond the scope of the study, it is implied that a deliberate adoption and use of the term “separation wall” may provide a path out of the political stalemate that has beset peace in the region for decades. The history of Agent Orange public policy debates and the peace talks between Israel and Palestine are not too dissimilar. This study is further evidence that the method used does not just reveal counts, but that inferences made about those counts—particularly when attached to other meaningful variables like “place”—might be of considerable value in campaigns for social change where communication and language play a central role.

The second reason I mention this study particularly is because, like mine, it is a mixed methods study that borrows from text analytics and uses network analysis tools to visualize findings. Rogers and Ben-David’s study is more sophisticated and better constructed than the present Agent Orange study, and the network visualizations are used to visualize the network of connections from actors to words instead of using visualizations as a kind of validity counterbalance. As noted elsewhere though, my study in its current form is limited because, unlike the political conversations about the “wall”, which are ongoing and voluminous, conversations about Agent Orange remediation and remuneration are comparatively scant and distributed.

The second part of the answer to the question of intervention has to do with the work of professional writing and marketing/advertising. In 2001, the National Research Development Council commissioned research from a nonprofit organization called the FrameWorks Institute to examine public beliefs and sentiment about the use of chemicals in the average American’s home and garden; one of these chemicals was a common herbicide. The research included a
review of the existing literature, a series of public interviews, rapid ethnographic assessment, public polling, and focus grouping. The data collected were analyzed and the findings applied in a pilot communications campaign launched in two test markets. These campaigns used a combination of paid and in-kind professionally designed and placed advertising with key messages tailored to reinforce values, attitudes, and beliefs expressed during data collection. Ongoing testing will measure the impact these campaign messages have had on the public’s preference for nontoxic home and garden products, as well as their understanding of environmental health risks related to domestic herbicide and chemical use.

As I hope this brief example illustrates, using data to affect social change is a complex and multifaceted process that starts by collecting and studying data using a variety of collection and analysis methods and tools. The work of intervention can take many forms. It can look like the reframing campaigns the Ford Foundation recently initiated, which encourage advocates and NGOs to use key terms and phrases that are more inclusive, just, and scientific. Or, it can look like the more overt marketing campaigns on domestic toxins like that conducted by the FrameWorks Institute. It can look like recommendation reports on policy delivered to congress before an appropriations bill vote, or a series of letters to the editor. As I have argued elsewhere, reframing alone cannot solve complicated techno-socio-political controversies like Agent Orange. Neither will single interventionist strategies. But, identifying what is privileged by the powerful, how the powerful describe those subjects and issues, and then critically analyzing those descriptions and frames offers activists access to a seat of power because language is power.
The Value and Possibilities of Network Analysis & Visualization

One of the more powerful aspects of the Rogers and Ben-David “wall” study is its use of network analysis tools to visualize results. Below is a visualization from their work that features the various terms used to describe the construction project in official and unofficial discourses (2010, p. 18).

Figure 9. Sources in Media

Term uses of official and unofficial sources in the media space.

Throughout this dissertation I have argued that one of the advantages of humanists borrowing tools and methods from other disciplines, specifically social network analysis, is that doing so allows us to visualize our findings in more persuasive and compelling ways, which buys credibility among those with whom we may not have always been credible. As technical communicators, data visualization methods are integral to our field already. The increasing amount of digital information has ushered in a golden age for data visualization. More data allows for more frequent exploratory analysis to explain key scientific, social, cultural and economic phenomenon. Of course, access to data alone is insufficient. In shifting our perception
of what “counts” as a research site, and in learning new theories, tools, and methods with which to study these new sites, we must also cultivate and extend our abilities to visualize our results and to understand what the patterns embedded in data say. We must learn to look for patterns, connections, outliers, and gaps. Modern databases of text, data, and figures are simply too large to examine without machine assistance. Beyond merely processing and computing data, programs can help us better visualize our findings too.

Presenting data through visualization is an effective way to take advantage of our strongest perceptual system—the eyes. Of course, choosing an effective presentation system is difficult and often requires special knowledge and training. Not all visualizations are created equally. The network visualizations included in this thesis, in example, are not true network visualizations in my opinion because they do not highlight the patterns, gaps, and outliers that make conclusions more obvious or, as Tukey suggested (1977), that force us to see “what we never expected to see.”

Despite the fact that networks are difficult to visualize, they are appealing because they focus on relationships rather than individual words of a text. Even as we found in this study, it is the relationship of words (the term toxic co-occurring to chemical, in example) that is most insightful. The same can be said of network visualizations that emphasize relational data. These relationships often explain cultural, political and economic phenomena and these connections are often as important as the elements themselves (Freeman 2004).

**Distant Reading through Network Visualization**

Distant reading is the idea that content or information in a text can be processed—subjects, themes, actors, places, pub dates, authors, titles—across a large number of cases without actually reading the text. This distant “reading” then is a form of mining that allows
information in the text to be analyzed at a glance. Some argue that distant “reading” is a misnomer because it isn’t reading so much as statistical processing. Proponents of distant reading however suggest that the computational statistics and visualizations expose aspects of text at scale otherwise impossible for human coders or readers to see, which provides new points of departure for research. This is my argument as to the value and utility of distant reading through network visualization techniques. I used such visualizations as a way into the discourses and as a way to validate and analyze what the computational programs calculated. Distant reading can highlight patterns in changes in vocabulary, moods, themes, and other topics and larger social, technical, or political questions can be asked about why those words or themes or moods were included or excluded of traditional studies of literary or historic materials.

Distant reading and network visualizations must still ask and answer questions about what can be measures, counted, sorted, displayed and interpreted, all of which affect the outcome of a research project, and results must always be “read” in relation to these decisions.

Were I conducting a true network analyses, I would determine the unit of observation, which entity to measure, what measurements to use (actor, pair of actors, relational ties, events, in example); I’d model the unit (actor, dyad, triad, subsets or network), and then quantify the relationship (directional or non-directional, in example). I did not conduct a true network analysis because the data did not support this kind of research study. Semantic or social network analysis is good for representing the formal properties of social configurations and the patterns through which perceptions of relationships are structured.

The skills technical communicators need to analyze network data include the ability to analyze the nodes, relationships between lines and direction (arrows), positions—whether a node is central, peripheral, isolated, or groups and subgroups/clusters and what statistics best calculate
key factors like centrality and betweenness. Networks also have to be “drawn”, or rather, untangled, as most networks render as a “hairball” of conjoined and overlapping nodes, links, colors, and sizes. Untangling a network visualization requires the use of programs like UCINET, Gephi, Pajek, ORA. Texteture, the program used for the “distant reading” of the texts in each corpora can be exported into files supported by these programs.

There is an entire subfield that studies the principles for drawing networks, and a detailed best practices of is beyond the scope of this dissertation, but general visualization tips include focusing on the distance between vertices, making sure they express the strength or number of ties as closely as possible; drawing (and coloring) connecting nodes closer together than those that are unrelated; assigning proportional line length to value; and, minimizing crossing lines.

Close reading and critical discourse analysis on small sets of data has been the work of rhetoricians for years, but there is an imbalance of research methods used in rhetoric and technical communications (Charney, 1996) that can be righted through the use of new, complex, and quantitative empirical methods. Though textual analysis will likely remain a core component of our work, this thesis attempts to demonstrate how quantitative content analyses can make our studies even more rigorous, our results more defensible, and our visualizations more persuasive and compelling with new or powerful audiences.

As stated earlier in the methods chapter, content analysis analyzes the words, phrases, or relationships between them. It is a “research technique for making replicable and valid inferences from texts and other meaningful matter in the contexts of their use” (Krippendorff K. , 2004, p. 18). Content can be analyzed quantitatively by identifying meaningful and valid measurements and making inferences based on statistical methods (Riffe, Lacy, & Rico, 1998). A corpus can be represented visually in using network analysis tools like ORA and GEPHI and Texteture. A
recent 2014 article by Jordan Firth demonstrates additional benefits of using social network analysis to research professional practices in Technical Communication (Frith, 2014). As Firth, Nuendorf, Krippendorff, Wodak, Graham, Pfeffer and others have shown, machine-assisted quantitative content analysis is particularly well suited to historical and longitudinal studies of large, digital corpora, enabling the modern researcher to study ever-smaller units of texts across ever-larger spans of time and quantity, and to visualize statistical relationships in novel and compelling ways.

**Overcoming Limitations Specific to this Project**

There are two areas of the study that need to be significantly redesigned moving forward. The first is in the kind of data collected, and the second is in the kinds of variables studied. One of the gravest disappointments in this project was a lack of thick description. The texts that composed the corpora were selected based on availability, but they were not the kind of “thick description” ethnomethodological studies require and the corpora did not truly constitute a discourse community as Berlin defines it. The lack of access to good textual data limited my ability to make claims about intentionally. I took this chance because it was extremely important to me to include voices that had been historically dismissed in the debates about Agent Orange, but going forward, it is clear to me that my texts need to be more clearly delimited by genre and variable (speaker, context, date, etc.)

Though individual researchers like Tine Gameltoft and Diane Niblack Fox, both of whom I have met or spoken to, have excellent sources of thick description collected for their own qualitative research projects, it was impossible to gather the necessary IRB approval from these subjects retroactively. I believe the story of Agent Orange will remain incomplete so long as
certain voices, specifically American and Vietnamese victims and veterans—remain excluded from political, technical and medical discourses.

To overcome this lack of thick description, I am presently writing a grant that, if funded, would support the construction of a website that will serve as a digital archive used to collect and digitized narratives of Agent Orange exposure. Enabled with recording modules, this site will allow victims to speak into an electronic device (mobile, desk top, iPad etc), and record their AO experience. Conceptually, the site will also automatically transcribe these audio recordings into machine-readable text files, which would then be sorted and archived according to front-end data the veteran, friend, or family member entered like gender, base assignment, years of service, etc. As discovered in this project, such variable data is supremely important, as variables allow for deeper and more meaningful statistical correlations. Ideally, this site will complete the failed government attempts to “map” military units (both American and Vietnamese) to time, place, and duration of exposure alongside exposure and health narratives of both the veteran and affected family members. Such data points would allow me to compare ways of describing and framing, and to conduct true social network analyses.

The design, development, and distribution of such a project is admittedly ambitious, but the value of a project of this scale and scope is manifold. The narrative data or “thick description” collected from veterans and victims can be analyzed and used to improve the outcome of advocacy, coalition building, and marketing efforts, as well as medical and health communication practices. Further, and perhaps most meaningfully, simply the ability to be heard—for the victim’s exposure narratives to be validated as a meaningful form of evidence—might at last relieve the weight of this thing they still carry. Clearly, narratives cannot undo the physical trauma of exposure, but telling one’s story can sometimes heal the mind, as narrative
projects in South Africa and at Holocaust Centers suggest (Young S., 2004). Perhaps in the
telling we can satiate the hungry ghosts that still haunt the discourse of Agent Orange.

As we develop more “ethnomethodological” investigations in Rhetoric and Technical
Communications we will undoubtedly apply these methods to sites where culture is being written
in real time, sites like Facebook, Twitter, media blogs, snapchats, and yikyak feeds. It is in new
sites of text creation and cultural production where methods like network analysis and tools like
RSS feeds, web-scrappers, Python, and corpus analytics software like WordStat and NVivo hold
the most promise for technical communicators.

**Studying Emerging Sites of Political and Scientific Controversy**

Throughout this dissertation, I have argued that words are never innocent. Let me borrow
from recent Florida headlines to support this claim. The state of Florida is one of the regions in
the country most susceptible to the effects of global warming. Sea level rise threatens 30 percent
of the state’s beaches alone, which could have significant effects on tourism, infrastructure, and
public safety. It is a serious threat and yet, the Florida Department of Environmental Protection,
the state agency responsible for planning for and safeguarding against environmental
catastrophe, can no longer use the terms “climate change” and “global warming” in any official
or technical communications. Though the policy is unwritten, the tactics and motives are clear:
Governor Rick Scott is subtly influencing legislation and public policy through language use.
More specifically, through the descriptive terms used to talk about controversial environmental
science.

Climate change and global warming are terms used to describe a body of scientific
evidence that shows the earth’s environment is warming because of human influence. Burning
fossil fuels and deforestation are two particularly dangerous human actions with long-range
climate effects. Climate change and global warming is accepted science all over the world. The Intergovernmental Panel on Climate Change (IPCC), established by the United Nations, wrote in a 2014 report that human influence and anthropogenic emissions of greenhouse gases are the highest in history and these changes have had “widespread impacts on human and natural systems” (Intergovernmental Panel on Climate Change, 2014).

Even so, many conservative U.S. politicians disbelieve this science, charging it is not conclusive, and they refuse to work on legislation that would address this issue without conclusive, correlative evidence. Established industries that have a financial stake in lobbying Congress have, so far, been effective in stopping proposed reforms like carbon taxes.

Like other savvy politicians before him, Scott recognizes in words and descriptive phrases a subtle magic—the magic to influence one’s attitudes and beliefs, to reinforce values. Were that not so, he would have merely banned the use of the terms, but he didn’t just do that; he also replaced them. Instead of sea-level rise, in example, Scott demands DEP employees and technical writers use the term “nuisance flooding” (Rice, 2015).

The similarities between Agent Orange and this most recent policy directive by Governor Scott are eerily similar. It took more than 40 years after the end of the Vietnam War to make significant environmental progress on remediating “hot spots”—former air bases contaminated with dioxin remnant left over from the use of weaponized herbicides in the war—and more than 20 years for veterans to win “presumptive exposure” so the medical costs of exposure-related diseases were covered. Still today, certain classes of veterans fight for inclusion under the Agent Orange Act of 1991, (see Bluewater Veterans and Agent Orange Exposure) and only this year did Vietnamese AO-dioxin victims see significant improvement in access to medical care through partnerships with VAVA and USAID. In fact, so contentious remained the issue of

15 http://www.publichealth.va.gov/exposures/agentorange/locations/blue-water-veterans.as
Agent Orange that it was only in 2014 that the first elected representative from Congress visited the home of a dioxin “victim.” In the meantime, tens of thousands of veterans and victims suffered, natural resources were devastated, and scientific progress stalled. We have to do better.

Like many legislators during the 1970s and 80s, who were skeptical of veterans’ claims of a link between Agent Orange-dioxin and disease, Governor Scott and other climate deniers remain skeptical about the science of climate change and global warming. In 2010, during his first bid for governor, Scott told reporters who asked his views on the subject that he “had not been convinced” of climate change. There are serious consequences to such denialism, then and now, which is what makes studying the metaphorical and descriptive phrases and frames used in public policy debates exigent.

Scholars like Lakoff, Rose & Baumgartner, Koerber, and Sontag among others have all demonstrated that the words and phrase with which we compose our technical, scientific, and political documents do not merely describe the objects, subjects, or processes contained therein, but also frame the greater political debates about those object, subjects, and processes, and thus are of significant consequence. Because language works subtly to influence behaviors, values, attitudes, and beliefs it is especially important to conduct research studies like this and others, which aim to make visible the invisible power of language on our public policies and scientific practices.

Turning back to Florida environmental policy for a moment, it is interesting to look at how descriptive phrases have framed key debates in policy negotiations about the environment. Consider for a moment the images and other words the term Everglades “restoration” versus “clean up” triggers in one’s mind. “Clean up” frames the debate in terms of negligence and accountability. Restoration is a much more neutral term. This is similar to the ways that
Congressional leaders insisted on describing AO as an herbicide, rather than a chemical or toxin, which may imply culpability and responsibility. Similarly, the word “pollution” has a far different charge to it than “runoff” or “outfall.” The words we choose to describe objects and subjects reflect and reify our values, attitudes and beliefs about those subjects. Similarly, the specific language used to describe a discourse is as important as the content itself. Therefore, mapping both what a discourse community with competing interests talks about and how they talk about it offers illuminating insights into how stakeholders advance positions and frame key aspects of complicated issues. I’ve included this vignette on Scott and Florida environmental politics as evidence of an interesting and unfolding modern scientific controversy unfolding in digital space that might serve as a site of new application for this method.

In studies where text is the primary means of data, these sites, methods and tools provide both strategy and the tactics necessary to conduct thorough, rigorous, and generalizable research. Relational studies, like this one, where themes and words are compared and their relationships analyzed, are more strongly supported with quantitative data. When that data can stream from sources in real-time, researchers have an opportunity to construct new knowledge—to see and perhaps even shape culture in the making—instead of as I have here, confirm long-held suppositions. The affordances of the tools used in this method include the ability to study new sites of discourse and discreet elements of discourse at scale. I am engaging in one such project presently, where I have scraped and am pulling the feeds of a private Facebook group called “Children of Agent Orange.” The aim of this project, which will be the focus of a presentation delivered in November at the Approaches in Digital Discourse Analysis conference in Valencia Spain, is to use digital methods to measure the ways that ethos is built in an online lay community that exchanges complicated scientific, technical, or medical data. Further, the
measurements as statistical representations offer quantitative text analysis distinct advantages over purely qualitative approaches alone. Research questions can be answered more conclusively and hypotheses more easily confirmed, rethought and visualized.

Summary of Results

The results of this study suggest that, as expected, the various stakeholders in the debate about Agent Orange talk about, describe, and frame the debate differently. For example, in talking about the environment, it is interesting to note that the term “dioxin” is used persistently in the Times article and VNA corpus, but is not frequently used in the Congressional corpus. In fact, the term occurs in the Congressional corpus 1,566 times, compared to 2,039 times in the Times corpus and 1,982 in the VNA corpus. That the term occurs 426 times more in the VNA corpus is particularly meaningful given that the Congressional discourse was significantly larger and thus there was more opportunity for the word to occur.

Other general findings suggest that the Congressional corpus is geared toward action, while the Times corpus tends to highlight what’s at stake or controversial; the editorial corpus is almost exclusively focused on veterans’ battles in court; and the VNA corpus is uniquely focused on the people, specifically the Vietnamese people, who are affected by dioxin.

In the environmental cross-section, the key themes are studying environmental effects, protecting the environment and cleaning up the environment. In the uncertainty cross-section, the Congressional corpus is the most uncertain, focused on probabilities of exposure and mitigating risk. The Congressional corpus carries in it the most probabilistic language of all the corpora. The Times corpus focuses on uncertainty in different ways, and is particularly interested in the effects of dioxin and the uncertain outcome of the multiple court cases. Language that marked uncertainty was almost nonexistent in the VNA and Times editorial corpuses. Though there could
be many explanations for this, the most likely explanation is two-fold. First, to focus on the uncertainty of Agent Orange/dioxin science is to shift the attention away from the rhetorical goals of advocating for financial support and remediation. Second, the absence of hedging or uncertain language is in itself a way of framing the issue. To the Vietnamese, the links between Agent Orange/dioxin and their physical maladies is not in question. Spending time debating the science of exposure defeats the rhetorical aim of securing medical, financial, and environmental relief and remediation from those they hold responsible for their illnesses and ecological imbalances.

The term “victim” is particularly prevalent in the VNA corpus, and acts as a key frame through which the Vietnamese seem to talk about all experiences related to Agent Orange. This is best visualized by the network image of the entire VNA corpus, which literally has virtually all nodes linked to the term “victim.” In all of the corpora, the most closely associated word with victims is “suffer”, which suggests semantic preference, if not also prosody, but what victims suffer or how to alleviate that suffering is different depending on the stakeholder. For Congress, relief comes through the better management of health care; in the Times suffering should be compensated, and includes birth defects; for the Times editorials, suffering is chloracne and job loss; and for the Vietnamese, suffering is from the disease “Agent Orange” for which chemical companies are responsible.

Remuneration is a particularly interesting and large part of the corpus. Here, the term “disabled” and “disability” appears and is closely associated with the term “claim” suggesting a rhetorical move on behalf of the Vietnamese to describe AO victims as disabled in an effort to win compensation claims, as were won in the United States. The disabled are not just veterans or even adults, but children too. This cross-section of the corpus shows how difficult it was for
veterans to win presumptive exposure. It also dramatically highlights the different tone and focus of each discourse with Congress emphasizing *process* and the VNA emphasizing *people*.

Finally, the ways the different stakeholders describe Agent Orange is particularly insightful. Congress more often describes Agent Orange as an “herbicide”, and almost never uses the term chemical or the phrase toxic chemical. The *Times* more often uses the term “defoliant” to describe Agent Orange, but nearly equally describes it as a “chemical.” In the editorial corpus it is almost always described as a “chemical” and in the *VNA* corpus it is associated most frequently with the term dioxin.

Though this study looks at historical textual data, broader applications of mixed-methods content analyses similar to this study have been used to evaluate transcripts from usability studies that help identify mental models; assess bias in government publications; codify images used in marketing collateral; and assess rhetorical strategies of successful grant applications. The aim of this study was not simply to investigate an interesting data set, but to also advance our discipline’s methods beyond the traditional use of frequency counts into more “ethnomethodological” investigations (Boettger & Palmer, Quantitative Content Analysis: Its Use in Technical Communication, 2010, p. 348).

In addition to the findings detailed above, which are of value in themselves, methodologically, I hope to have shown that large-scale content analysis studies are best approached not only through mixed methods, but also by combining proprietary coding processes, as developed in Python for this project, and commercially-available content analysis software, like WordStat. Using custom scripts like Python affords the research team a greater level of flexibility to parse textual material into meaningful units that can then be analyzed, quantified and visualized using commercial products that do those things well.
Conclusion

To restate, quantitative methods need not replace more qualitative methods, but can be used to enhance traditional, qualitative studies adding valuable methods of triangulation and proof. As our field continues to define its boundaries and the specific practices and insights technical communications offers to other fields and disciplines, strengthening the research we produce is supremely important. Using a diverse cache of methods and theoretical frames from other fields, including Sociology, Linguistics, and Computer, Network and Cognitive Science and best practices in research design and visualization, technical communicators and rhetoricians will produce work that is qualitatively rich, narratively satisfying, methodologically defensible, and practically applicable. These methods are particularly helpful in cracking open the black box of “settled” issues like Agent Orange (Latour, 1999), offering new ways to explore, analyze, criticize, visualize, and advocate for change on issues of technical complexity and significant social consequence.
REFERENCES


----(2009, July NA). Metaphorically Speaking. NA.


http://www.nap.edu/openbook.php?record_id=2141&page=33


Figure 10. Frequency List for Congressional/Environment 1960–1969
*Top ten most frequently occurring keywords in the Congressional corpus on the Environment, 1960-1969.*
Figure 11. Descriptive Phrases, Congress/Environment, 1960–1969

Top ten most frequently occurring descriptive phrases in the Congressional corpus on the Environment, 1960-1969. Note: there are only four concepts included here because there were only four concepts that occurred with frequency over 10. This was common when the corpus was very small, as is the case for this small cross-section of texts representing only 10 years of texts, but contained in only a few documents.

Figure 12. Co-Occurrence Congress/Environment, 1960–1969


Figure 13. Frequency List, Congress/Environment
Top ten most frequently occurring keywords in the Congressional corpus on the Environment (full corpus).
Figure 14. Descriptive Phrases, Congress/Environment
Top ten most frequently occurring descriptive phrases in the Congressional corpus on the Environment (full corpus).

Figure 15. Co-Occurrence List, Congress/Environment
Top ten most frequently co-occurring words in the Congressional corpus on the Environment (full corpus).
Figure 16. Frequency List, New York Times/Environment
Top ten most frequently occurring keywords in the New York Times corpus on the Environment (full corpus).

Figure 17. Descriptive Phrases, New York Times/Environment
Top ten most frequently occurring descriptive phrases in the New York Times corpus on the Environment (full corpus).
Figure 18. Co-Occurrence List, New York Times/Environment
Top ten most frequently co-occurring words in the New York Times corpus on the Environment (full corpus).

Figure 19. Frequency List, Vietnamese News Agency/Environment
Top ten most frequently occurring words in the Vietnamese News Agency corpus on the Environment (full corpus).
Figure 20. Descriptive Phrases, Vietnamese News Agency/Environment
Top ten most frequently occurring descriptive phrases in the Vietnamese News Agency corpus on the Environment (full corpus).

Figure 21. Co-Occurrence List, Vietnamese News Agency/Environment
Top ten most frequently co-occurring words in the Vietnamese News Agency corpus for key theme “Uncertainty.”
Figure 22. Frequency List, Congress/Uncertainty
*Top ten most frequently occurring words in the Congressional corpus for key theme “Uncertainty.”*

Figure 23. Descriptive Phrases, Congress/Uncertainty
*Top ten most frequently occurring descriptive phrases in the Congressional corpus for key theme “Uncertainty” (full corpus).*
Figure 24. Co-Occurrence List, Congress/Uncertainty
Top ten most frequently co-occurring words in the Congressional corpus for key theme “Uncertainty” (full corpus).
Figure 25. Frequency List, New York Times/Uncertainty
Top ten most frequently words in the New York Times articles corpus for key theme “Uncertainty” (full corpus).

Figure 26. Descriptive Phrases, New York Times/Uncertainty
Top ten most frequently occurring descriptive phrases in the New York Times articles corpus for key theme “Uncertainty” (full corpus).
Figure 27. Co-Occurrence List, New York Times/Uncertainty
Top ten most frequently co-occurring descriptive phrases in the New York Times articles corpus for key theme “Uncertainty” (full corpus).

Figure 28. Frequency List, Congress/Victimization
Top ten most frequently occurring words in the Congressional corpus for key theme “Victimization” (full corpus).
Figure 29. Descriptive Phrases, Congress/Victimization
Top ten most frequently occurring descriptive phrases in the Congressional corpus for key theme “Victimization” (full corpus).
Figure 30. Co-Occurrence List, Congress/Victimization

Top ten most frequently co-occurring words in the Congressional corpus for key theme "Victimization" (full corpus).
Figure 31. Frequency List, New York Times/Victimization
Top ten most frequently occurring words in the New York Times article corpus for key theme “Victimization” (full corpus).

Figure 32. Descriptive Phrases, New York Times/Victimization
Top ten most frequently occurring descriptive phrases in the New York Times article corpus for key theme “Victimization” (full corpus).
Figure 33. Co-Occurrence List, New York Times/Victimization
Top ten most frequently co-occurring words in the New York Times article corpus for key theme “Victimization” (full corpus).
**Figure 34.** Frequency List, New York Times Editorial/Victimization
Top ten most frequently occurring words in the New York Times editorial corpus for key theme “Victimization” (full corpus).

**Figure 35.** Descriptive Phrases, New York Times Editorial/Victimization
Top ten most frequently occurring descriptive phrases in the New York Times editorial corpus for key theme “Victimization” (full corpus).
Figure 36. Co-Occurrence List, New York Times Editorial/Victimization
Top ten most frequently co-occurring words in the New York Times editorial corpus for key theme “Victimization” (full corpus).

Figure 37. Frequency List, Vietnamese News Agency/Victimization
Top ten most frequently occurring words in the Vietnamese News Agency corpus for key theme “Victimization” (full corpus).
Figure 38. Descriptive Phrases, Vietnamese News Agency/Victimization

*Top ten most frequently occurring descriptive phrases in the Vietnamese News Agency corpus for key theme “Victimization” (full corpus).*
Figure 39. Co-Occurrence List, Vietnamese News Agency/Victimization
Top ten most frequently co-occurring words in the Vietnamese News Agency corpus for key theme “Victimization” (full corpus).
Figure 40. Frequency List, Congress/Remuneration
Top ten most frequently occurring words in the Congressional corpus for key theme “Remuneration” (full corpus).

Figure 41. Descriptive Phrases, Congress/Remuneration
Top ten most frequently occurring descriptive phrases in the Congressional corpus for key theme “Remuneration” (full corpus).
Figure 42. Co-Occurrence List, Congress/Remuneration
Top ten most frequently co-occurring words in the Congressional corpus for key theme “Remuneration” (full corpus).

Figure 43. Frequency List, New York Times/Remuneration
Top ten most frequently occurring words in the New York Times article corpus for key theme “Remuneration” (full corpus).
**Figure 44.** Descriptive Phrases, New York Times/Remuneration
Top ten most frequently occurring descriptive phrases in the New York Times article corpus for key theme “Remuneration” (full corpus).

**Figure 45.** Co-Occurrence List, New York Times/Remuneration
Top ten most frequently co-occurring words in the New York Times article corpus for key theme “Remuneration” (full corpus).
Figure 46. Frequency List, New York Times Editorial/Remuneration
Top ten most frequently occurring words in the New York Times editorial corpus for key theme “Remuneration” (full corpus).

Figure 47. Descriptive Phrases, New York Times Editorial/Remuneration
Top ten most frequently occurring descriptive phrases in the New York Times editorial corpus for key theme “Remuneration” (full corpus).
Figure 48. Co-Occurrence List, New York Times Editorial/Remuneration
Top ten most frequently co-occurring words in the New York Times editorial corpus for key theme “Remuneration” (full corpus).

Figure 49. Frequency List, Vietnamese News Agency/Remuneration
Top ten most frequently occurring words in the Vietnamese News Agency corpus for key theme “Remuneration” (full corpus).
Figure 50. Descriptive Phrases, Vietnamese News Agency/Remuneration
Top ten most frequently occurring descriptive phrases in the Vietnamese News Agency corpus for key theme “Remuneration” (full corpus).

Figure 51. Co-Occurrence List, Vietnamese News Agency/Remuneration
Top ten most frequently co-occurring words in the Vietnamese News Agency corpus for key theme “Remuneration” (full corpus).
Figure 52. Frequency List, Congress/Agent Orange
Top ten most frequently occurring words in the Congressional corpus for key theme “Agent Orange” (full corpus).

Figure 53. Descriptive Phrases, Congress/Agent Orange
Top ten most frequently occurring descriptive phrases in the Congressional corpus for key theme “Agent Orange” (full corpus).
Figure 54. Co-Occurrence List, Congress/Agent Orange
Top ten most frequently co-occurring words in the Congressional corpus for key theme “Agent Orange” (full corpus).

Figure 55. Frequency List, New York Times/Agent Orange
Top ten most frequently co-occurring words in the New York Times articles corpus for key theme “Agent Orange” (full corpus).
Figure 56. Descriptive Phrases, New York Times/Agent Orange
Top ten most frequently descriptive phrases in the New York Times articles corpus for key theme “Agent Orange” (full corpus).

Figure 57. Co-Occurrence List, New York Times/Agent Orange
Top ten most frequently co-occurring words in the New York Times articles corpus for key theme “Agent Orange” (full corpus).
Figure 58. Frequency List, New York Times Editorial/Agent Orange
Top ten most frequently occurring words in the New York Times editorials corpus for key theme “Agent Orange” (full corpus).

Figure 59. Descriptive Phrases, New York Times Editorial/Agent Orange
Top ten most frequently occurring descriptive phrases in the New York Times editorials corpus for key theme “Agent Orange” (full corpus).
Figure 60. Co-Occurrence List, New York Times Editorial/Agent Orange
Top ten most frequently co-occurring words in the New York Times editorials corpus for key theme “Agent Orange” (full corpus).

Figure 61. Frequency List, Vietnamese News Agency/Agent Orange
Top ten most frequently occurring words in the Vietnamese News Agency editorials corpus for key theme “Agent Orange” (full corpus).
Figure 6.2. Descriptive Phrases, Vietnamese News Agency/Agent Orange
Top ten most frequently occurring descriptive phrases in the Vietnamese News Agency editorials corpus for key theme “Agent Orange” (full corpus).

Figure 6.3. Co-Occurrence List, Vietnamese News Agency/Agent Orange
Top ten most frequently co-occurring words in the Vietnamese News Agency editorials corpus for key theme “Agent Orange” (full corpus).