

**Actor-Role Analysis:
Ideology, Point of View, and the News**
by
Warren Sack

B.A. Computer Science and Psychology
Yale University
New Haven CT
1985

Submitted to the Program in Media Arts and Sciences
School of Architecture and Planning
in partial fulfillment of the requirements of the degree of

Master of Science in Media Arts and Sciences
at the
Massachusetts Institute of Technology
September 1994

© MIT 1994
All Rights Reserved

Signature of Author

.....
Program in Media Arts and Sciences
August 19, 1994

Certified by

.....
Kenneth W. Haase, Jr.
Assistant Professor of Media Arts and Sciences
Program in Media Arts and Sciences
Thesis Supervisor

Accepted by

.....
Stephen A. Benton
Chair

Departmental Committee on Graduate Students
Program in Media Arts and Sciences

MASSACHUSETTS INSTITUTE
OF TECHNOLOGY

OCT 12 1994

**Actor-Role Analysis:
Ideology, Point of View, and the News**

by
Warren Sack

Submitted to the Program in Media Arts and Sciences
School of Architecture and Planning
on August 19, 1994
in partial fulfillment of the requirements of the degree of

Master of Science in Media Arts and Sciences
at the
Massachusetts Institute of Technology

Abstract


A theory of ideology and point of view is articulated and a method for detecting the point(s) of view represented in a news story is described. A version of the method, *actor-role analysis*, is encoded in a computer program, *SpinDoctor*, which can automatically detect the point(s) of view represented in some news stories. Results obtained by testing SpinDoctor on a corpus of news stories are reported. Related work in the fields of natural language processing, information retrieval, and the theory of ideology is described and contrasted with the work reported in this thesis.

Thesis Advisor:
Kenneth W. Haase, Jr.
Assistant Professor of Media Arts and Sciences

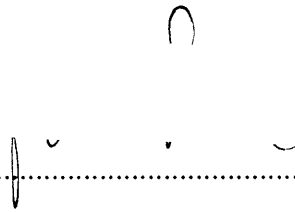
**Actor-Role Analysis:
Ideology, Point of View, and the News**
by
Warren Sack

The following people have served as readers for this thesis:

Reader


.....
Edith K.E. Ackermann
Associate Professor of Media Arts and Sciences
Program in Media Arts and Sciences

Reader


.....
Teun A. van Dijk
Professor of Discourse Studies
University of Amsterdam

Acknowledgments

This thesis is a part of a larger research project on news analysis that germinated during the many discussions I had with my friends and colleagues in the Cultural Studies of Science and Technology and Graduate Students Against the War groups at the University of California Santa Cruz. During the long months of the Gulf War we were analyzing the news media's coverage. Some of the reflections of our discussions can be seen in the *Journal of Urban and Cultural Studies*, Volume 2, Number 1. I learned a great deal that winter and spring of 1991.

The following fall I wrote essays to apply to graduate school and to apply for a Fulbright to go to the University of Amsterdam. Those first essays on proposed news analysis work benefited from advice and critical comments from Joost Breuker, University of Amsterdam; Mark Driscoll, UCSC; Joe Dumit, UCSC; Jennifer González, UCSC; David Levin, Columbia; Tom Levin, Princeton; Maureen O'Malley, UCSC; Vivian Sobchack, UCLA; Elliot Soloway, University of Michigan, Ann Arbor.

During the application process I met the three key people who supervised this thesis. Ken Haase, my advisor accepted me into the Media Lab to pursue my studies here. He has fed me, nurtured me intellectually, provided me with all kinds of equipment, nudged me forward at critical times, and, most importantly, turned me loose to explore the fields of research in whatever areas I felt a need to go. I could not ask for a better advisor than Ken.

Edith Ackermann's critical yet sensitive eye has always been able to see things in my work that I am blind to. Sometimes the things that she sees are gaping holes, but usually she spares me that and then finds delightful undercurrents that, once they are explained to me, bear me off into new, interesting, challenging directions.

Teun van Dijk and I met through email during my Fulbright application process. Although, in the end, I decided not to go to

Amsterdam, Teun has provided me with copious comments on my work and has shared his thoughts about his own on-going work. I have no way of explaining how I ever could have merited the enormous amount of time and energy that Teun has given me other than by noting that I certainly did not merit it; rather, Teun is a very generous man who has given me a myriad of ideas about the work in this thesis and the work that I hope to accomplish in the following years.

When I arrived at the Media Lab in the fall of 1992 I had the opportunity to write an essay on news audiences (Sack, 1994a) for a class taught by Edith Ackermann and Mitchel Resnick which allowed me to start sketching out some of the territory upon which this thesis rests. The following spring Abbe Don and I wrote a browsing and resequencing tool for digital video (Splicer/Suture-Self-Video) for a class taught by Henry Lieberman and Hal Abelson. The discussions I had with Abbe about point of view helped clarify my interest in ideology and gave me a new perspective on news video and documentary film. Classes and discussions with Hayward Alker, Political Science Department and Henry Jenkins, Literature Department have also been formative. I would like to thank them for the interest and concern they have shown in my writing.

Last summer I wrote a first draft of my thesis proposal and a program to illustrate it. I also wrote a synopsis of the research for the International Joint Conference on Artificial Intelligence Workshop on Models of Teaching and Models of Learning (Sack, 1993). I would like to thank the workshop organizers for allowing me to present my work there, nascent as it was, especially Larry Birnbaum, Northwestern University; Roger Schank, Northwestern University; and Elliot Soloway, University of Michigan, Ann Arbor.

Janet Cahn's meticulous critique of that first draft and Ken Haase's review of the program helped prepare me for "Thesis Prep" a seminar run by Corrine Bickley. Corrine brought all of the master's thesis students together to discuss and critique our thesis proposals. I gained a great deal from my classmates by learning more about their projects and gaining the benefit of their thoughts

on mine. I especially enjoyed my discussions with Gilberte Houbart whose project has certain affinities with this thesis; we have been able to hash out some interesting similarities and differences.

Walter Bender organized a weekly meeting for all of us sponsored at the Media Lab by the News in the Future (NiF) Consortium and provided us with a wonderful forum in which to eat sushi, discuss common issues, and learn about each other's work. I have learned a lot from Walter and the rest of the NiF crew.

Throughout my time at the Media Lab, Marc Davis has been teaching me about digital video and pushing me to think about whether my work has relevance in this age of television. My work with him on video story generation (Sack and Davis, 1994) made me realize that I needed to know more about news analysis before I can go on to do more work in generation. Marc and Wendy Buffett have given me a room in their house in San Francisco to finish the thesis and Marc encouraged me to present preliminary results of the thesis at the American Association of Artificial Intelligence Workshop on Indexing and Reuse of Multimedia (Sack, 1994b) which he co-chaired.

I would like to thank my closest colleagues at the Media Lab, the group called hmm (for Haase, Minsky and students): Janet Cahn, Anil Chakravarthy, Marc Davis, Sara Elo, Ken Haase, Marvin Minsky, Alan Ruttenberg, and Mike Travers. Anil and Mike have been my officemates and so have probably heard more about this thesis than they ever wanted to, but the others have had an earful too and I thank them for their patience, ideas, and friendship.

Penultimately, I need to make a blanket thank you for everyone that I have inevitably forgotten to mention. Of course with an acknowledgments this long and detailed I really should be able to thank everyone by name, but I realize that the list would be like a role call for the Media Lab and beyond. For example, I forgot to mention "Wave," Michael "Wave" Johnson, that is; Wave who always wanders up from the basement at one o'clock in the morning to find me and tell me a thousand things about design, animation, motion blur and distributed computing; Wave who

enjoys the gourmet delights of Pizza Hut as much as I do; Wave who always so graciously asks “Aren’t you done with that thing yet? Just right it up man. It’s only a master’s!” I am not going to thank people like Wave, nor my friends in IC, the VLW, Tod Machover’s group, and Neil Gershenfeld’s group, and those in E&L, and in Pattie Maes’ group, the people in Narrative Intelligence, and my exercise buddies in Club Media. The reason why I am not going to thank everyone is because I am trying to keep these acknowledgments focused only on those who made a direct contribution to the thesis. OK?

Finally, one person whom I must thank is my partner in everything, Jennifer González. She is the one who asks the most difficult questions of my work; the sorts of questions that only one who loves you deeply can ask: Why is your work worthwhile? What is interesting about it? Why can’t you explain the whole of it to me now, succinctly, and without omission? Someday Jennifer, I hope to have some good answers to at least some of your questions.

Table of Contents

| | |
|---|-----|
| Acknowledgments | 4 |
| Table of Contents..... | 8 |
| Chapter 0: Introduction: Actor-Role Analysis and Electronic News | 9 |
| Chapter 1: Ideology as Closure | 12 |
| Chapter 2: Actors, Roles and Points of View as a Representation of Ideology ... | 25 |
| Chapter 3: A Computational Representation of Actor-Role Analysis..... | 33 |
| Chapter 4: An Extended Example of Actor-Role Analysis | 52 |
| Chapter 5: Related Work | 69 |
| Chapter 6: Conclusions..... | 79 |
| References..... | 80 |
| Appendix A: Actor Patterns | 86 |
| Appendix B: Role Patterns | 90 |
| Appendix C: Point of View Patterns | 94 |
| Appendix D: Ten News Stories from the MUC-3 Corpus | 95 |
| Appendix E: Ten of SpinDoctor's Actor-Role Analyses | 108 |

Chapter 0: Introduction: Actor-Role Analysis and Electronic News

This thesis develops a methodology for critically reading the news. The methodology, actor-role analysis, was developed to study the issues of news editing and filtering. As many media experts have pointed out (e.g., Herman and Chomsky, 1987) the stories in newspapers and broadcast radio and television are edited and filtered to support certain points of view, certain ideologies. This observation, in the age of mass media, can only be read as a critique of media producers, i.e., as a critique of journalists, editors, and television executives. In the age of electronic media, this critique will also apply to news recipients as well as news producers.

With electronic media (e.g., electronic mail) recipients can also be producers and so the distinction between the positions of recipient and producer becomes blurred. Furthermore, recipients of electronic media can be *selective* recipients in a qualitatively different manner than recipients of mass media. With mass media, the recipient typically either takes it or leaves it: the recipient either subscribes to a particular newspaper or does not; either watches a given television broadcast or does not. Of course, cracks in this model of the audience have occurred with the introduction of new technologies (Ang, 1990). For example, the remote control channel changer, in combination with the storage facilities of a video cassette recorder, allows the television viewer to watch shows at a different hour than their broadcast, "zip" commercials, and "zap" through several channels.

However, a whole new set of digital technologies are on the horizon that will make it practically impossible to continue to characterize the audience as a passive take-it-or-leave-it group. The new digital technologies will allow users to find, filter, and reformat their news and entertainment. Pieces of software designed to automatically ferret out an interesting story by crawling through computer networks, pieces of software to filter by

applying a set of automatic, critical reading strategies to found stories, and pieces of software to assemble summaries and collage and splice together new presentations of found material will be available to the selective recipient of electronic media. Consequently, not only producers, but also recipients will be filtering and editing news to suit their own points of view, their own ideologies.

To many this view of the future looks insidious. Yet, how can this future be more insidious than the present? In the present a small minority filters, edits, and broadcasts the news to the majority. In the future it will hopefully not be a small minority, but rather a great majority, filtering, editing, and broadcasting the news to each other. Nevertheless, this view of the future carries with it an intellectual and moral imperative: computational techniques and tools need to be developed to allow producers and recipients to thoughtfully and responsibly create, find, filter, and reformat the news. The goal of this thesis is to create a practical theory of ideology and point of view which can be applied by a user to critically read and filter electronic news.

During the Gulf War, the linguist George Lakoff wrote an essay on the metaphors used by journalists to justify the war (Lakoff, 1991). In the essay he points out how the war has been cast in the American media in a fairy tale-like format where the US plays the role of hero, Iraq the role of villain, and Kuwait is the maiden in distress.

One way of viewing this thesis is as an attempt to elaborate answers to some of the key questions Lakoff's essay raises: (1) What is a news actor, like the US, President Bush, or Iraq? (2) What is a role, like villain or hero? (3) What kinds of distinctions are being drawn when certain actors are paired with certain roles? (4) And finally, can any of this theoretical discussion be represented computationally to help readers of electronic news sort it out in a manner similar to the way in which Lakoff sorted out the US news of the Gulf War?

The work to answer these questions is of two sorts: theoretical and computational. The theoretical work is required to knit

together a theory and methodology, actor-role analysis, that might be acceptable to narrative and media theorists, and that might also provide some means of answering questions 1, 2, and 3 in a principled manner. The computational work involves the implementation of a prototype computer program that supports some semblance of actor-role analysis.

Lakoff sorts the news by ideology when he asserts that certain metaphors and certain pairings of news actors and dramatic roles are used to justify the war in the Gulf. The present investigations starts, in Chapter 1, with an examination of what is meant by "ideology." Chapter 2 discusses the theory of ideology that forms the backbone of the actor-role analysis methodology: actors and roles are combined to form points of views, groups of points of view are said to constitute a representation of an ideology. Chapter 3 is a computational "test" of the actor-role analysis methodology. A computer program, SpinDoctor, has been written which can perform an ideological analysis of a news story and categorize the story according to whose point of view is represented in the story. SpinDoctor has been designed to help explore the issues of news *filtering* and *editing*, namely, Whose point of view?, Whose interests?, are represented in the news? Chapter 4 is an extended example of the use of actor-role analysis to describe points of view in news about El Salvador. This actor-role analysis was coded into the SpinDoctor system and SpinDoctor was used to analyze a corpus of news stories. Some data concerning SpinDoctor's performance are reported in this chapter. Chapter 5 is a review of related work in computational news analysis.

Chapter 1: Ideology as Closure

The term ideology is often used pejoratively. When one says “Oh, that’s just ideology” often one could just as well have said “That’s untrue” or “That’s a distorted picture of the world” or “That’s unscientific.” Even within the specialized social science literature on ideology, “ideology” is often meant to denote “false consciousness.” However, there exist a variety of other useful conceptions of ideology. Eagleton (1991) lists several of them:

- (a) the process of production of meanings, signs and values in social life;
 - (b) a body of ideas characteristic of a particular social group or class;
 - (c) ideas which help to legitimate a dominant political power;
 - (d) false ideas which help to legitimate a dominant political power;
 - (e) systematically distorted communication;
 - (f) that which offers a position for a subject;
 - (g) forms of thought motivated by social interests;
 - (h) identity thinking;
 - (i) socially necessary illusion;
 - (j) the conjuncture of discourse and power;
 - (k) the medium in which conscious social actors make sense of their world;
 - (l) action-oriented sets of beliefs;
 - (m) the confusion of linguistic and phenomenal reality;
 - (n) semiotic closure;
 - (o) the indispensable medium in which individuals live out their relations to a social structure;
 - (p) the process whereby social life is converted to a natural reality.
- (Eagleton, 1991: 1-2)

It is possible to identify one or more theorists of ideology who have developed each of the definitions of ideology listed by Eagleton. For example, it could be said that the hermeneuticist Paul Ricoeur, with his notion of actions as texts (Thompson, 1984: 189) employs something like item (l) as do the artificial intelligence

theorists Abelson (1967) and Carbonell (1979), with their analysis of texts as actions.¹ The theory of ideology that I will be advancing in this thesis is, essentially, item (n) *semiotic closure*, a theory discussed by, among others, Fredric Jameson (1981).

In my discussion I will not assume, as many theorists do, that an ideology is equivalent to a system of beliefs. I submit that ideology is a special kind of system of beliefs; it is a system of beliefs about social, economic, and political relations; beliefs that regulate and represent individuals, institutions, groups, and relationships between them. Thus, in principle, any belief could be a part of an ideology because any sort of belief might affect some relationship. As many historians of science have shown, even beliefs about obscure scientific technicalities can be shown to be ideological (e.g., Haraway's (1991) studies of primatologists). However, in practice, only a subset of all beliefs is of significance in a given study of ideology. Which beliefs are of interest depend upon the individuals, institutions, and groups studied. Thus, beliefs about the number and type of subatomic particles might clearly partition some set of physicists at a given historical juncture, but would be of no use in exploring the social structure of another group, e.g., a group of professional baseball players, at the same moment in history.

Ideology and News Texts

Literally, *ideology* means the study of ideas. However, since ideas can only be studied in some material form, it is necessary to choose a medium within which to observe the expression of ideology. The current thesis will focus on ideology and its textual instantiations. In particular, I am concerned with news texts and the means by which ideological differences are inscribed in news texts. Thus, the phenomena of interest for this thesis are the lexical choices and phrasal arrangements that mark a text as under the influence of a given ideology.

¹ See (Van Dijk, 1981: 14) for a critique of the widespread practice in artificial intelligence whereby researchers do not distinguish between stories and actions.

Ideology can affect the text of both partisan and “objective” news:

- *Partisan News*: When the ideological roots of a news product are admitted by the publisher or broadcaster, the ideological functions of the news product, i.e., the promotion of certain social, political, and economic relationships favorable to the news producer and its public, are openly supported and discussed in news texts. For example, in France, until the 1830's newspapers were run as political party organs. Party members subscribed to their party's paper and so the papers were unabashedly partisan.
- *“Objective” News*: When a news producer attempts to create an “objective” news product, the worst insult that can be leveled is that the news product is “ideological” rather than “objective.” Thus, for this sort of news, ideology is important because textual strategies are employed by writers and editors to phrase the stories so that their audience does not find them ideological. In other words, under this conception of the news in which news is considered to be the objective reporting of facts, ideology constantly lurks in the background and is actively policed against by writers and editors. Various researchers have suggested that “objective” news was a marketing strategy to sell newspapers to groups larger than a single political party. For example, in France, after 1830, several entrepreneurs (notably in Paris, Moise Millard, the founder, in 1863, of *Le Petit Journal*) radically changed the partisan news market by selling the newspaper as a commodity. Prices were slashed and “objectivity” in journalism was emphasized to promote the commodity newspaper to a large, multi-party audience (Terdiman, 1985: 117-146).

In either case (of partisan or “objective” news) what can be studied, and what I attempt to study in this thesis, are the textual

strategies employed by writers and editors of news products to both explicitly express and hide ideological affiliations.

Even in “objective” news these textual strategies, or clues, which attempt to hide ideological bases of news production are generally not very subtle. “Modalization” is one such strategy of hiding. For example, objecting to strongly “subjective” language of a given journalist, an editor might tell the journalist to rewrite a phrase like “Mr. Giovanni, a heinous murderer and well-known figure in the Mafia” and then accept as a replacement the following, more “objective” phrase that conditionalizes the description of the news maker: “Mr. Giovanni, *allegedly* a heinous murderer and well-known figure in the Mafia.”

Ideology and Groups

Textual expressions of ideology reflect social, economic, and political beliefs that function to bind or partition groups of people; e.g., ideology separates radicals from liberals from moderates from conservatives from the ultra-right. This is the conception of ideology that I will work with in this thesis: ideology is a set of distinctions (in particular, textual distinctions) that partition the universe of peoples into smaller, opposing and contrasting groups.

The sociologist Bourdieu cautions academics and intellectuals against seeing the groups they write about (e.g., “liberals” and “conservatives”) as actual, mobilized groups, as opposed to theoretical constructs (e.g., Bourdieu, 1991: 231). I will address Bourdieu’s point more carefully after first presenting an explanation of what it might mean, even theoretically, to partition people into groups.

Ideologies (e.g., the various forms of Marxism) often partition people into opposing groups (e.g., “the working class” versus “the bourgeoisie”). Often, an ideology privileges one of the groups of the partition and then names the other groups using derogatory terminology. For example, the radio and television commentator Rush Limbaugh labels a variety of his opponents “feminazis.” Obviously, no such group called “the feminazis” exists. The name

and the grouping of “feminists” with “Nazis” is simply a reflection of Rush Limbaugh’s ideology; his language functions to partition people into groups and then, also, to vilify certain of the groups dictated by his partition. Of course, not all ideologies vilify opposing groups, but most ideologies will privilege one of, or a small set of, the groups. I will call this conception ideology-as-semiotic-closure.

Under this interpretation, ideology is "closed" insofar as certain groups and certain relations between certain groups are central to a given ideology. The number of groups and relations outside of the central ones are therefore limited and constitute the margins or horizon of an ideology. Thus, for example, a racist ideology might categorize people according to skin color and then privilege one color; other means of classification, e.g., according to a person's age, might play only a margin role in the discriminations made by a racist. Consequently, a racist ideology could be closed semiotically to the extent that its division of the world of people into groups is limited to the number of races recognized. By contrast an ageist ideology is also limiting and thus semiotically closed, but in a different manner; the set of groups recognized within an ageist ideology would be limited to a different set of people (e.g., children, privileged youths, middle-age adults, older adults) than the set enumerated by a racist.

Semiotic Squares as Models of Ideology

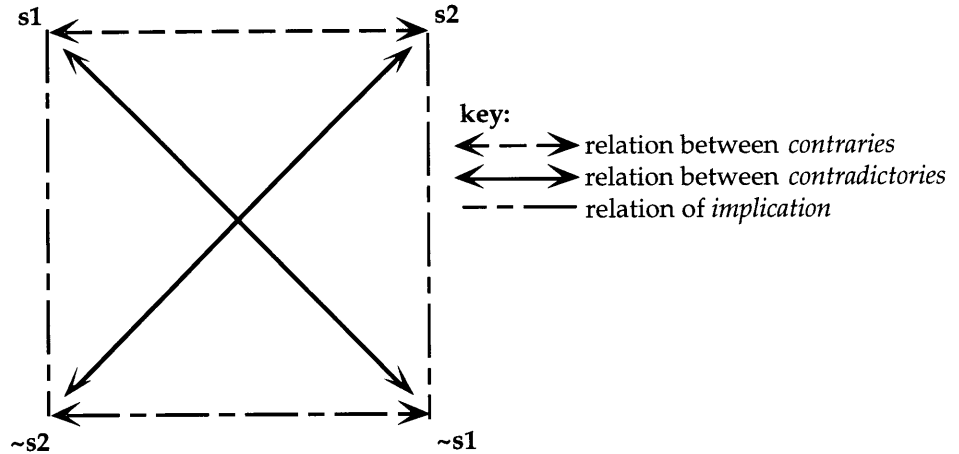
Fredric Jameson, in his studies of ideology, uses a formal tool from narrative analysis to model how relationships between groups can be mapped out within an ideology (Jameson, 1987; Jameson, 1982). The semiologist Algirdas Greimas developed the tool for purposes of narrative (and ideological) analysis. In the theory of literature, the tool is usually referred to as a *semiotic square*. Jameson is not unique in his use of Greimas’ tool: the semiotic square is widely used in literary theory and other parts of the humanities. Nevertheless, Jameson is a central figure within literary theory and his work on ideological analyses of texts and

other media are considered seminal by many. Thus, Jameson's use of the semiotic square as a tool for ideological analysis deserves special attention and will be critiqued in this thesis.

This is Jameson's short description of the semiotic square and its possible utility in ideological analyses:

I have suggested that ideology, in some more comprehensive sense, can be grasped as a twofold or amphibious reality, susceptible of taking on two distinct and seemingly incompatible forms at will, which are very precisely our old friends the narrative or the cognitive. That "ideology" in the narrower sense is a mass of opinions, concepts, or pseudoconcepts, "worldviews," "values," and the like, is commonly accepted; that these vaguely specified conceptual entities also always have a range of narrative embodiments, that is, indeed, that they are all in one way or another buried narratives, may be less widely understood and may also open up a much wider range of exploration than the now well-worn conceptual dimension of the ideology concept. Yet it was not to replace the cognitive by the narrative that my proposal was made but rather to coordinate both by way of a definition that insisted on their necessary alternation: *Ideology is whatever in its very structure is susceptible of taking on a cognitive and a narrative form alternately.* [Emphasis added]

The relevance of these proposals for Greimas's work turns on the whole matter of the so-called elementary structure of signification, or, in other words, the famous "semiotic square," for many of us the supreme achievement of Greimassian semiotics. Here finally we find opened up the "black box" through which narrative is somehow "converted" into cognition and vice versa: Finally we have the equations, we can witness the processes of transfer, which need no longer be posited mystically since it is "visible" before us ... How this can be so, however, obviously demands yet another simplified exercise in the explanatory capacities of the "square," whose canonical form is herewith reproduced:



The enumeration of the advantages of the square can begin at once with the observation that it is a decisive enlargement on the older structural notion of the binary opposition: s_1 versus s_2 is clearly just a binary opposition, or in the language of philosophical logic a “contrary,” that is, a strong opposition (white versus black, male versus female), but one that the square now reveals to encompass far more than two available positions. It immediately implies, for example, the two supplementary slots of what logic calls a “contradictory,” where $\sim s_1$ and $\sim s_2$ are the simple negatives of the two dominant terms, but include far more than either: thus “nonwhite” includes more than “black,” “nonmale” more than “female.” Meanwhile, the two compound or “synthetic” positions of S and $\sim S$ offer still greater conceptual enlargements, S standing as a complex or utopian term, in which the opposition of “white” and “black” might be transcended (mestizo, for example), whereas $\sim S$ stands as the neutral term, in which all of the privations and negations are assembled (“colorless,” for example). Finally, the transversal axes map the place of tensions distinct from the principal or binary one, while the synthesis hypothetically proposed by uniting the two sides of the square (“white” plus “nonblack”) designates alternative conceptual combinations. The entire mechanism then is capable of generating at least ten conceivable positions out of a rudimentary binary opposition (which may originally have been no more than a single term, e.g., “white,” which proves to be internally defined by a hidden opposition we

articulate by promoting the concealed pole “black” to visibility). I have suggested that other traditions may find this schema interesting if they entertain the hypothesis that it constitutes a virtual map of conceptual closure, or better still, of the closure of ideology itself, that is, as a mechanism, which, while seeming to generate a rich variety of possible concepts and positions, remains in fact locked into some initial aporia or double bind that it cannot transform from the inside by its own means.
(Jameson, 1987: xiii-xv)

I will make use of the semiotic square as a simple model of ideology-as-semiotic-closure. It is a *model* of this type of ideology in so far as it gives one a mechanism for envisioning a set of contrasting positions all of which are more or less defined in relation to one another. It is a *simple* model of the ways in which an ideology can be means of partitioning up the world, of enumerating possible groups of people, because it is a very simple graph structure, a structure with four nodes. Obviously, a more complete representation of positions, signs, or “semes” would allow one to depict more than four. However, I submit that the square gives one a place to start a more complete enumeration; it pushes the modeler to move beyond binary oppositions.

Jameson’s enthusiasm for Greimas’s semiotic square seems to stem from two properties of the square, one a weakness and the other a strength: (1) The strength of the square lies in its ability to allow us enumerate sets of possible positions that take us beyond simple binaries like “conservative” versus “liberal.” (2) The weakness of the square lies in its exclusive use of two forms of contrast, “contradictions” and “contraries.” Once all four of the corners of the square are filled in, it is difficult to think of other possible, contrasting positions. Thus, Jameson claims, the square is a good model of ideology in so far as an ideology can circumscribe the horizons of the number of groups into which one can conceive of partitioning the world (Jameson, 1981: 47-48).

What is less clear about Jameson’s proposal is his notion of how the terms of the square can “alternate” between cognitive and

narrative forms. However, the insight that ideology both shapes narrative forms (e.g., news stories) and cognitive structures is a key insight that can be illustrated by venturing further into Greimas's theory of narrative, as I will soon show. Furthermore, I believe that the means by which ideological structures can function both cognitively and narratively can be illustrated with the computational means of representing ideologies that I have chosen for this thesis (namely, a particular kind of machine readable patterns), even if it is not clear how the semiotic square illustrates the double function of ideological structures, as Jameson claims it does.

Of course, one might argue, that as a model of ideology-as-semiotic-closure, Greimas's square is overkill. Would not the structure of a simple binary opposition work just as well as a "model"? In fact, I have been implicitly employing this binary model in my discussion to this point by using examples like conservative/liberal and workers/bourgeoisie. However, the square gives us something beyond the schematization of two rational possibilities: it allows one to start mapping out why, and in what circumstances, two terms can be posited as oppositions or *contradictories* (as Greimas prefers to say).

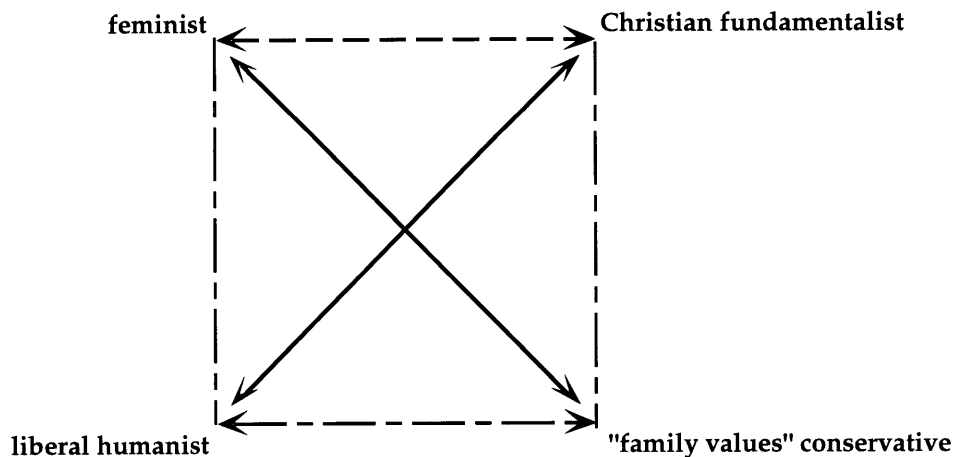
A Model of an Ideology

One can look at the contradictory positioning of two groups of people, like "feminists" and "Christian fundamentalists," and, using the semiotic square, examine a bit more closely the tensions which separate these two groups. For example, given an issue like abortion one can start to map out the ideology which circulates around a "pro-choice" versus "pro-life" opposition. To a large degree, most feminists are "pro-choice"; they see abortion as an issue that concerns women's bodies and whether women are given a say as to what they want to do with their own bodies. Contrary to the feminists' position, the "pro-life" position, is the position of Christian fundamentalists (and Catholics) which see the issue as an issue of life or death of the embryo. Given this binary opposition,

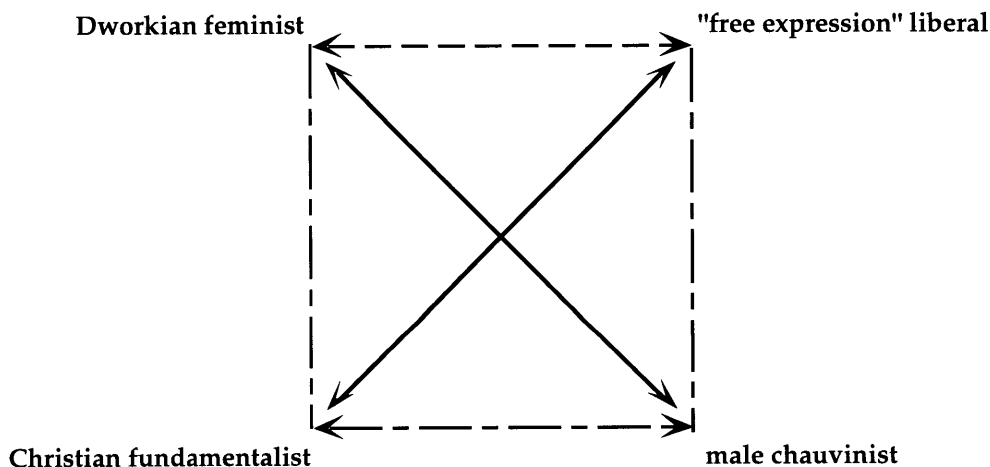
feminist versus Christian fundamentalist, it is easy, using the techniques of the semiotic square to start to enumerate other groups in conflict and/or concordance with the first two.

After negating both of the original terms (to produce “not-feminist” and “not-Christian-fundamentalist”) one can replace the negations with synonymous terms. Immediately, we are faced with a decision about what, for example, does “not-Christian-fundamentalist” mean? This kind of a question is typical of the kinds of questions generated by the use of the semiotic square and it illustrates why the square is a useful tool for thinking beyond binary oppositions. “Liberal humanist” comes to mind as a possible synonym for “not-Christian-fundamentalist” since Liberal humanists are often at odds with Christian fundamentalists on other issues like, for example, whether or not “creationism” or “evolution” should be taught in the classroom.

The fourth term of a square is always the most difficult one to fill in. One must pick a relation which is *implied* by one of the existing positions, which is a *contrary* of another, and which is a *contradictory* of the third term. The difficulty associated with filling in the fourth term is consequently a provocative model of ideology-as-semiotic-closure: once the world has been partitioned into several parts by assigning those parts names, it is hard to imagine other “parts,” other positions, which could fit into the spaces left between the existing positions framed by the semiotic square. “Liberals” are often opposed by “conservatives,” “Christian fundamentalists” often promote, what they call, “Family values,” and popular culture representatives of “feminism” are often cast as opponents to high-profile members of “family value conservatives” (e.g., the real/televisual debate between the sitcom character Murphy Brown and then-vice president Dan Quayle over single-parent families). Thus, perhaps in this case, the fourth term should be “family value conservatives.” Consequently, the completed semiotic square looks like this:



What one example square by itself cannot illustrate is how the four terms depicted (s_1 , s_2 , $\sim s_1$, $\sim s_2$) not only have multiple relations between them which distinguish them from one another (the implies, contradictories, and contraries relations), but also the ways in which these relations can shift given a new context. For example, the following square schematizes one of the ideological complexes which surrounds the issues of pornography:



Notice how three of the four terms from the abortion square have been retained in the pornography square (feminist, liberal, and fundamentalist). However, those three terms now stand in a different relation to one another. A certain kind of feminism which opposes pornography (which I will name "Dworkian feminism" for one of its most vocal advocates (Dworkan, 1981)) is often opposed

to many “liberals” who see anti-pornography work as a form of censorship, and so fight against it. Instead, this group of feminists, at least in American politics, often finds itself in an uneasy alliance with the religious right, the “Christian fundamentalists,” who, in this square again, are contradictories to the liberals. However, in this case, the contradictories of the feminist position is not the “family values Conservatives.” It is highly unlikely that very many Republicans campaigning with a “family values” plank in their platform would support pornography. Instead, the “liberals” also find themselves in an uneasy alliance with a group that habitually opposes the feminists, namely, “male chauvinists.” Thus, I would claim, that one semiotic square may be more useful as a model of an ideology than a simple binary opposition. But, two or more squares gives one a better model of the ideology. It allows one to see how more than four positions can come into play and it also begins to model some of the complicated relationships between the world of positions represented in an ideology.

Ideologies and Points of View

Because the semiotic square does not privilege any of the four terms included, it seems natural to ask *which* ideology is modeled by a square or set of squares? A feminist ideology? A Christian fundamentalist? A liberal humanist? A “family values” conservative ideology? A male chauvinist ideology? In theory -- in the theory of ideology-as-semiotic-closure -- the square models all four of these positions by articulating how each of the four positions gains its meaning by staking out an area that is not covered by the other three positions. This philosophy of representation-by-relation-to-other-entities is a sort of *(post)structuralism* (Saussure, 1959). In another conception of ideology (e.g., in a conception of ideology that Eagleton terms, in the list at the beginning of this section, “action-oriented sets of beliefs”) each of the four positions in the square might be considered an ideology. However, in the theory of ideology that will be used in this thesis, it is the complex of all four positions (or

the five or more positions of two or more squares) which constitutes the model of an ideology. For the purposes of this thesis, each of the positions (feminism, Christian fundamentalism, etc.) will be termed a *point of view*. The term *point of view* will be defined in a more detailed fashion in the following chapter.

Summary

The conception of ideology that I have chosen to investigate in this thesis, ideology-as-semiotic-closure, is only one of many possible, useful definitions of ideology. I have chosen this conception of ideology because it is a useful one for discovering how groups of people and their representations (e.g., newspapers and news weeklies which claim to be looking out for the interests of a certain public or readership) are distinguished from one another. My object of study is the textual evidence of ideologies. I could, of course, have chosen to study the representation of ideology in any other medium (e.g., film or photography), however I have chosen text as a practical concern. Computers are now able to quite easily manipulate, index, and match text, so text is an obvious medium to study with the assistance of computers. Finally, I wish to state that I believe semiotic squares, as they have been employed by Fredric Jameson, and others, can function as interesting models of ideologies. I think the squares give one a means to begin to sketch out the horizons of an ideology. Nevertheless, I will later claim that my work is a significant improvement over the semiotic squares, even though the squares may be the right place to begin articulating a representation of an ideology. In this thesis I will propose a representation of ideologies which are constituted as sets of points of view and which are encoded as patterns which can be matched against texts.

Chapter 2: Actors, Roles and Points of View as a Representation of Ideology

Categorizing news stories according to ideological concerns requires one to compare stories which might have completely different narrative structures. For example, one newspaper might run a long front-page story on a subject of pressing concern to its readership while a rival paper might cover the same subject in a small item buried in the middle of the paper. To compare the two stories it would be necessary to compare a long story to a very short one.

Characteristically, the situation is actually much more complicated: given stories from two different news products, produced by opposing ends of an ideological spectrum, it is often almost impossible to find two stories which could be said to be about the same subject. For example, if one examines the French-American trade agreement talks (on GATT, for example) from the perspective of *Le Monde*, the specifics of the talks might be woven into a larger story about farms and farming in France. The *New York Times*, on the other hand, might include in its story a discussion of US trade imbalances with Japan and the EEC. Thus, the only parts of the stories which might be comparable would be the bits and pieces which directly refer to the talks. The methodological challenge is to find a set of features that can provide the basis to compare stories with very different narrative structures.

Regardless of the narrative structure of the stories there are some story features which remain constant even across ideological divisions. In particular, the people that journalists call the "newsmakers" often appear ubiquitously in the print, radio, and television media. Newsmakers -- people like heads of state, rich businessmen, and famous actresses -- often show up again and again, for months and years on end, in all sorts of news forums.

Sociologist Herbert Gans claims that five types of "Knowns" and five types of "Unknowns" are the predominant actors in American domestic news:

[The "Knowns" are] (1) Incumbent presidents; (2) Presidential candidates; (3) Leading federal officials; (4) State and local officials; (5) Alleged and actual violators of the laws and mores (i.e., "famous" violators, like the Watergate defendants). [The "Unknowns" are] (1) Protesters, "rioters," and strikers; (2) Victims; (3) Alleged and actual violators of the laws and mores; (4) Participants in unusual activities (e.g., "cult" members or child prodigies); (5) Voters, survey respondents, and other aggregates. (Gans, 1980: 8-15).

Those newsmakers which Gans calls "Knowns" and "Unknowns" constitute a relatively stable group of people which are written about and depicted in all kinds of news media. They are one sort of similarity which exists across ideologically differing news products and so they are a stable point around which differences between ideologically differing news stories can be measured. "Knowns" and "Unknowns" in the news will be referred to as news *actors*. The ideological "markings" of a news story include the descriptions a story assigns to a news actor. Consequently, when a news actor, like Oliver North, is described as a "criminal" in one story and as a "patriot" in another, one can easily tell that the stories are significantly different from an ideological standpoint. *Roles* is the term that will be used to refer to the descriptions associated with news actors.

Actors, Roles and Narrative Theory

The question of *what* a tale's dramatis personae do is an important one for the study of the tale, but the questions of *who* does it and *how* it is done already fall within the province of accessory study. (Propp, 1968: 20)

Narrative theorists distinguish between actors and roles principally so that the actors can be abstracted away thereby making it easier for the theorist to compare the narrative structures of different stories. So, for Propp, who studied Russian folktales at the beginning of the century, it did not matter if the hero's role was filled by an actor named Ivan or Vladimir. By contrast, in the news, often the most important aspect of a story is who did what and how. For example, two of the hottest aspects of the Iran/Contra story of 1986 were (1) Reagan's role in the affair; and, (2) whether or not Oliver North did his work with the approval of his superiors. Narrative theorists distinguish actors from roles in order to, essentially, ignore the former while journalists distinguish actors from roles to investigate their overlap. This thesis uses the formalized distinctions of narrative theory for the purposes of analyzing the products of journalism.

Greimas (1987) formalized and fine-tuned many of the distinctions originally introduced by Propp earlier in the century. One of Greimas's refinements was to break roles into two types: (1) actantial roles; and, (2) thematic roles. *Actantial roles* can only be identified by examining how an actor that fills an actantial role interacts with the other actors within a given narrative. For instance, heroes and villains are actantial roles. Heroes and villains come in all shapes and sizes. The only way one can distinguish a hero from a villain is by comparing their actions and interactions within a given narrative. Heroes are distinguishable from villains and vice versa, but outside of a given story it is difficult to say who is a hero and who a villain. Thus, we might guess that heroes will wear white hats and villains black, but we are not confused for long when the author of a narrative ignores or even reverses this convention. We recognize the heroes and villains not necessarily by what they wear but rather by what they do to and with one another and the other actors (like those, for example, who play the roles of victim). Greimas says that actantial roles, like heroes and villains, exist on a *narrative level*.

Thematic roles, on the other hand, exist on what Greimas calls the *discursive level*. That is, thematic roles are part of larger discourses

which necessarily connect together many stories. A fisherman is an example of a thematic role. Associated with fishermen are a set of attributes and activities that are carried over from story to story and which remain the same regardless of how a fisherman interacts with others in a story. For example, fishermen usually have a boat and fishing gear, like nets, lines, and hooks, and they often can be found baiting, catching, and cleaning fish.

Normally, within a story a particular actor is assigned one or more actantial roles and one or more thematic roles. An actor can be both a fisherman and a hero; or, perhaps, a fisherman and a villain. The converse of actor-role assignments is also true: any given role is often assigned to more than one actor. Consequently, stories often include more than one fisherman and two or more villains.

Within this thesis I will be predominantly concerned with the identification of *actor* and *thematic role* combinations because thematic roles (as opposed to actantial roles) can be identified in narratives with extremely different structures.

One reason why the ideas in Lakoff's evocative essay on the metaphors used to justify the Gulf War cannot simply be applied to analyze other news stories is his reliance on various actantial roles to describe the ways in which the US, Iraq, and Kuwait were described as hero, villain, and victim in the US news media. It may be the case, as Lakoff claims, that the Bush Administration pitched the war to the American public by phrasing it in fairy tale-like terms and casting the US in the hero's role. However, since only a handful of newspapers stories seem to have the structure of a fairy tale, it is difficult to decipher exactly which textual clues Lakoff was referring to when he claims to be able to see the thematic roles of hero, villain, and victim in the US news stories.

Actors, Roles and Ideology

It was claimed that Jameson's use of the semiotic square to model ideologies was an interesting method, but that one could do better. One could do better by unpacking some of the internal

structure of each of the points of view (e.g., feminist, fundamentalist, liberal, conservative) in order to represent more specifically why a certain point of view is "contradictory to" or "contrary to" or "implied by" another. Of course the internal structure of a point of view is infinitely complex and one certainly cannot hope to, for example, state *the* feminist position. In fact a single seamless position is a fiction. As soon as one looks more closely the single point of view fractures into many. While a large majority of feminists may support abortion rights, feminists split into a variety of factions over the issue of pornography. Thus, one must pick a modifier like "Dworkian" to specify a particular kind of feminism. Similarly, Christian fundamentalists are more a coalition for issues like abortion and pornography than they are actually a community (see, for example, NYT, Friday, July 22, 1994, page A1). Groups like "feminists," "fundamentalists," "liberals," and "conservatives" are useful for modeling purposes but may indeed not be "real:"

On the basis of knowledge of the space of positions, one can carve out classes in the logical sense of the word, i.e. sets of agents who occupy similar positions and who, being placed in similar conditions and submitted to similar types of conditioning, have every chance of having similar dispositions and interests, and thus of producing similar practices and adopting similar stances. This 'class on paper' has the theoretical existence which belongs to all theories: as the product of an explanatory classification, one which is altogether similar to that of zoologists or botanists, it allows one to explain and predict the practices and properties of the things classified -- including their propensity to constitute groups. It is not really a class, an actual class, in the sense of being a group, a group mobilized for struggle; at most one could say that it is a probable class, in so far as it is a set of agents which will place fewer objective obstacles in the way of efforts of mobilization than any other set of agents. (Bourdieu, 1991: 231-232).

Keeping in mind then that the goal here is to represent, not create, ideologies and points of view, some of the internal structure of points of view can be elaborated. The constructs of actors and thematic roles can provide some of the necessary tools to begin this work.

Consider again the semiotic square shown in the last chapter which details four positions around the issue of abortion: feminist, Christian fundamentalist, liberal humanist, and "family values" conservative. The square models the tension between the feminist and Christian fundamentalist positions by labeling them "contraries." By looking at the difference in actor-role assignments one can tease out a more descriptive representation of the tensions between the two positions. For example, from a feminist perspective a fetus is described as a part of a woman's body while, by contrast, the fundamentalists assign the fetus the role of unborn child. Consequently, from the feminist position the decision to have an abortion is a decision that a woman should be able to make because it concerns her own body. Fundamentalists, on the other hand, consider an abortion a murder because it involves the life of an unborn child.

The liberal humanist's position is "implied" by the feminist position, but has a slightly different focus. Rational decision making by individuals is a central tenant of the liberal humanist point of view. Liberal humanists therefore focus less on the woman's body and more on woman-as-rational-decision-maker. The actor "woman" is assigned the role of rational decision maker and so the liberal humanist probably agrees that the fetus is a part of a woman's body, but the main issue is seen as the woman's right to make a rational choice.

For the "family values" conservative the issue of abortion is dominated by the role played by the woman rather than the role played by the fetus. From a conservative's point of view a woman's place is in the home: she is a homemaker not a decision maker. The conservative might also agree with the fundamentalist that the fetus is an unborn child. Yet, from this point of view the most

serious issue of the termination of a pregnancy is its displacement of a woman from her assigned role of mother and homemaker.

One can do a similar actor-role analysis for the issue of pornography. For the "Dworkian" feminist the publisher of pornography is an advocate of anti-woman violence. The Christian fundamentalist often teams with the Dworkian feminist against pornography, but not because the fundamentalist position has women's well-being at its center. Rather, the focus centers around the products of pornography (e.g., videos and magazines) which are seen as obscene.

The "free expression" liberal differs radically from the Dworkian feminist on the issue of pornography because, to the liberal, a publisher of pornography is simply a publisher and, like all publishers, should be allowed to express their opinion. The male chauvinist probably agrees with the liberal that publishers are publishers and all should be granted the same rights. However, the male chauvinist is, like the fundamentalist, focused on the products rather than the publication process of pornography. To the chauvinist pornographic materials play the role of natural representations of women. The male chauvinist does not acknowledge that pornographic representations of women are highly artificial and stylized. Moreover, since the chauvinist sees the representations of women in pornography as "natural" their relationship to the Dworkian feminists is particularly acerbic. From their point of view, the feminists play the role of "anti-naturalists" who deny the affinity men have for "natural" pornographic representations of women.

The actor-role analyses for the issues of abortion and pornography can be summarized in a set of diagrams which label points of view and depict actor-role assignments with a notation like this: <actor> ——— <role>.

(Dworkian) feminist point of view

fetus ——— part of a woman's body

publisher of pornography ——— advocate of anti-woman violence

Christian fundamentalist point of view

fetus ——— unborn child

pornographic materials ——— obscene materials

(free-expression) liberal (humanist) point of view

woman ——— rational decision maker

"family values" conservative point of view

woman ——— homemaker

male chauvinist point of view

pornographic materials ——— "natural" representations of women

feminist ——— "anti-naturalist"

Like the semiotic square an actor-role analysis can effectively represent an ideology both because it gives one the means to push beyond simple binary oppositions, but also because it illustrates the sort of semiotic closure that can occur. By listing a set of actors and a set of roles and then pairing them off in various ways one can begin to imagine a whole range of possible positions or points of view. Closure is reached when either one cannot imagine some actor-role pairing (e.g., the assignment of fetus to the role of publisher) or when one's enumeration of actors and roles stops. I will claim, however, that actor-role analyses are better models of ideology than semiotic squares because the actor-role notation allows one to unpack some of the internal structure of the various positions and thereby more specifically articulate the tensions and affinities between points of view.

Chapter 3: A Computational Representation of Actor-Role Analysis

The program constructed to illustrate the theory of ideology and point of view presented in this thesis, SpinDoctor, has been designed to dissuade one particular kind of objection that computational scientists might have to the theory. The kind of objection that seems likely is one in which the critic argues that certainly one cannot ever hope to write any sort of computer program to do an ideological analysis of text until all of the other problems identified in text-processing (by researchers in artificial intelligence, computational linguistics, information retrieval, etc.) have been solved because, it is asserted by the critic, "ideology" is some sort of concept far and away much more abstract than, for example, phonetics, morphology, syntax, or even semantics. Contrary to this criticism, earlier in this thesis it is argued that ideology is most certainly not some sort of ephemeral, abstract concept. Quite the contrary: ideology is, as the term's inventor, Antoine Destutt de Tracy, first argued, to be found in the material world and is, as many scholars have contended since the early nineteenth century, both material and fundamental, indeed, much more fundamental than the abstractions of "syntax," "semantics," etc. used in several of the various sub-disciplines of linguistics and artificial intelligence. Of course, this argument about what is "more fundamental" is a long standing one in linguistics and artificial intelligence. My assertion that syntax is not necessarily more fundamental than ideology can be seen as a reflection of, for example, Minsky's (1986: 266) dispute with Chomskian syntax analysis. Consequently, SpinDoctor is designed to be a relatively "simple" (e.g., in that it does not incorporate a sophisticated syntax analysis routine) program which, nonetheless, can perform a semblance of an ideological analysis of a news story.

Two other software implementations of actor-role analysis were constructed before the current SpinDoctor system. The first implementation, Micro-Edit-Works (Sack, 1993), completed in

summer 1993 was a "toy" implementation insofar as it could only analyze one news story. Both Micro-Edit-Works and an earlier version of the SpinDoctor system (Sack, 1994b) included sophisticated syntax parsers. In fact, the earlier version of SpinDoctor included not only a syntax parser (De Marcken, 1991), but also a morphological analyzer (Kimmo/Englex, Anteworth 1991; Anteworth, 1992), and a part-of-speech tagger (Cutting et. al, 1993) all developed elsewhere by teams of computational linguists for purposes unrelated to the development of the SpinDoctor system. SpinDoctor included a specially tailored knowledge representation language, Trind, to store and retrieve words, phrases, sentences, stories, and their analyses and provided facilities for storing the results of the parsers and analyzers into the database. However, after some experimentation with the older version of SpinDoctor, it became clear that a simpler design could be created that would both run faster and allow one to specify actor and role patterns that allowed the user to more easily predict which texts the patterns would match against because the pattern matcher of the current version is a significantly simpler process than the older version which uses the parser and morphological analyzer.

Upon retrospect it does not seem surprising that the current version of SpinDoctor is simpler than the earlier version. Indeed, what does seem surprising is that I tried to build a program to do actor-role analysis "on top of" a set of programs to do syntax, morphology, etc. which are probably considered by their respective authors "simpler" or more "fundamental" than ideological analysis. In the end, however, one tool from computational linguistics was retained, the part-of-speech tagger from Xerox PARC (Cutting et al., 1993).

Work done by researchers at SRI suggests that perhaps not even a part-of-speech tagger needs to be incorporated into a good actor-role analysis system. Researchers from SRI competed in both the MUC3 and MUC4 competitions.² In MUC3 they used their

² MUC stands for Message Understanding Conference. It is a yearly competition sponsored by the Defense Advanced Research Agencies to test state-of-the-art natural language processing systems against one another. MUC-3 and MUC-4

TACITUS system which performs a set of sophisticated computational linguistic analyses including, for example, a syntax parse. According to the SRI researchers "using it [TACITUS] for the information extraction task gave us a high precision, the highest of any of the sites. However, our recall was mediocre, and the system was extremely slow [it took TACITUS 36 hours to process 100 news stories]" (Hobbs et al., 1992: 268). For the MUC4 competition the SRI researchers built a far simpler system that matched a set of simple patterns (cascaded finite-state automata) against the text of the stories. The SRI's MUC4 system, FASTUS did not even use a part-of-speech tagger and yet was one of the best performing systems in the MUC4 competition.

The parallel drawn here between TACITUS-FASTUS and the older, more complicated SpinDoctor and its newer, simpler instantiation is appropriate, not just because both sets of systems were designed to handle the same types of texts, but also to point out that in natural language processing systems more complicated is not always better. SpinDoctor and FASTUS pick out and retain different details of the same texts. Like all natural language processing systems, the systems were designed to privilege certain aspects of a text and to marginalize other parts. What is important about any natural language processor -- or any reading strategy for that matter -- is not how complicated a methodology is employed, but rather what aspects of a text are privileged by a given methodology. FASTUS did very well at filling templates devised by the Defense Advanced Research Projects Agency (DARPA). SpinDoctor has been designed to apply a method of text analysis, actor-role analysis, that has correlates in a variety of liberal and radical media analyses.

The current version of SpinDoctor is implemented with about 3500 lines of Common Lisp (Steele, 1991) code. In addition it also uses a very large package developed in Common Lisp by Xerox

(held in 1991 and 1992 respectively) used as a test corpus a set of news stories from Central and South America. The test the natural language processing systems were asked to perform was to find all of the "terrorist events" in the corpus of stories (MUC3, 1991; MUC4, 1992).

PARC researchers to do part-of-speech analysis (Cutting et al., 1993). About 1500 of the 3500 lines implement the basic data structures and algorithm of the analysis method. The other 2000 lines are devoted to the implementation of a graphical interface for browsing analysis results and the contents of the database and for editing the data structures in the database. The graphical interface will not be described in this thesis. A description of the interface will be reserved for a forthcoming user's manual. The database contains descriptions of actors, roles and points of view. Examples of these can be found in Appendices A, B, and C. For this thesis, SpinDoctor was tested on the first 100 stories of the MUC3 corpus. Appendix D contains the full text of the first ten of the news stories. Appendix E shows SpinDoctor's analyses of these first ten stories.

Analyze-story

Analyze-story is the top-level routine of the SpinDoctor system. Given the following story, analyze-story processes it and then outputs the analysis which follows.

[MUC 15]

Clandestine, 10 Jan 90 (Radio Venceremos) -- [text] On 10 January at 1030, on the 10th anniversary of Radio Venceremos, an Fmln unit commemorated this occasion by ambushing a military vehicle transporting national guardsmen from Sesorí to San Miguel. The ambush took place at El Botijón, in San Sebastián canton, Sesorí, and was carried out using explosives. A few minutes after the fighting began our troops gained control of the situation.

Attention, your attention: The military vehicle was destroyed during the ambush and five national guard soldiers were killed, including a sergeant who was in charge of the national guard troops. During this successful ambush in Sesorí, we seized from the enemy -- the repressive national guard -- three g3-a3 rifles, three sets of military, over 1,000 bullets for g3-a3 rifles, boots, vests, and other military materiel. Two g3-a3 rifles and a prc-77 military

communications radio were destroyed at the site of the ambush. Our people withdrew with no problems.

analysis of the story [MUC3-0015]

GROUP = RADIO-VENCEREMOS

actor: Venceremos *plays the role :* ?source

and has been coreferenced to

actor: Our people *plays the role :* ?military

actor: our troops *plays the role :* ?military

GROUP = ARMY

actor: five national guard soldiers *plays the role:* ?victim

UNGROUPED

actor: enemy *plays the role:* ?enemy

weighted high-level actor-role pairs:

(guerrillas ?source 1)

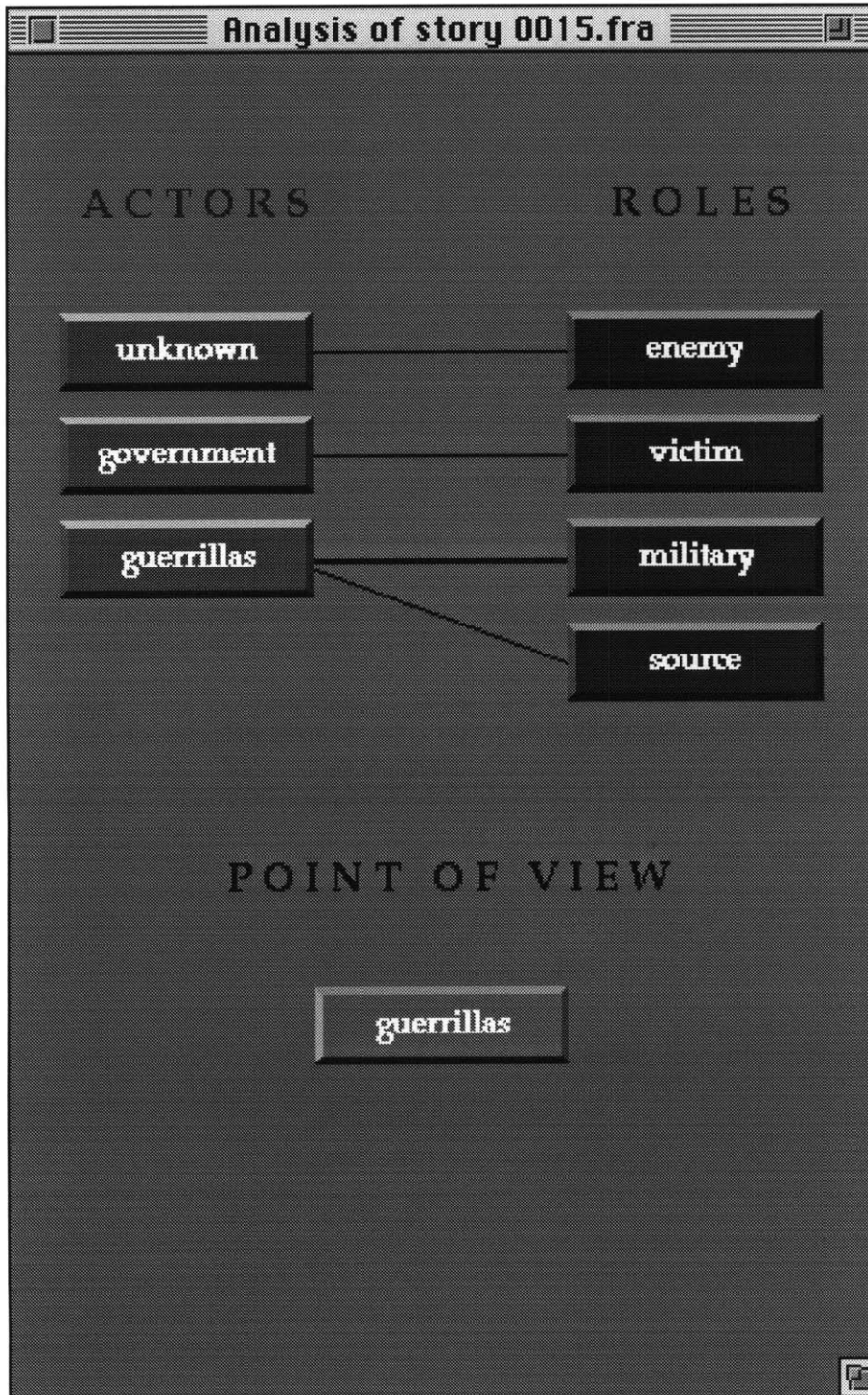
(guerrillas ?military 2)

(government ?victim 1)

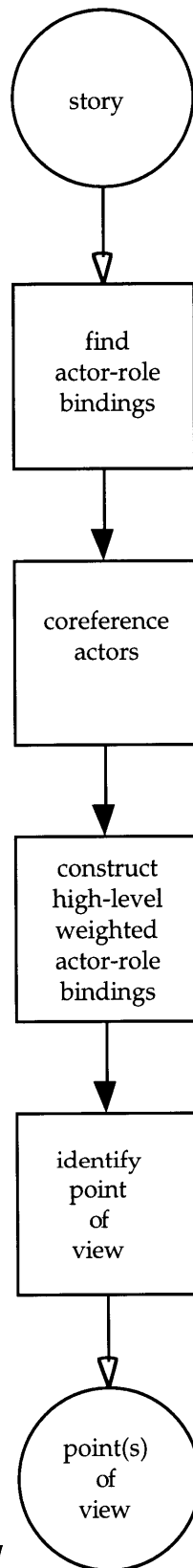
(unknown ?enemy 1)

=====
POVs = (guerrillas)
=====

The system also provides a graphical version of the above analysis that can be browsed by the user by pushing buttons and selecting items from a series of menus. The graphical interface generates a window to illustrate the analysis results (a window like the one shown in the next figure).



Graphical Presentation of an Analysis by SpinDoctor



Flow diagram for *analyze-story*

The analysis process followed by analyze-story (diagrammed in Figure 2) is as follows: (1) Given a newspaper story analyze-story calls the subroutine *find-actor-role-bindings* to find which noun phrases (i.e., actors) play which roles in the story. (2) Then, it determines which actors are of a similar group or are identical to other actors. The subroutine which performs this task is called *coreference-actors*. Coreference-actors does both (a) anaphoric resolution (e.g., determining who or what is being referenced when pronouns like "he," "she," and "it" are used); and, (b) actor grouping (e.g., inferring that "an army spokesman," "Maj. Rodriguez," and "Gen. Bustillo" are all a part of the same group (e.g., the armed forces). (3) After coreferencing the actors into groups, analyze-story uses the subroutine *construct-high-level-weighted-actor-role-bindings* to construct a profile of the analyzed story which describes how often (within the given story) different (groups of) actors (e.g., the armed forces, the government, the FMLN) were assigned different roles. (4) Using weighted actor-role bindings allows the final subroutine, *identify-pov*, to distinguish, for example, stories in which the government is cited as a source once and the FMLN is cited as a source ten times from stories in which the FMLN is cited as a source less often than the government. Points of view are represented as lists of actor-role pairs. Identify-pov matches the weighted actor-role bindings against the point of view definitions, determines which points of view match best and outputs the name of the one (or more) best-matching point(s) of view.

Find-actor-role-bindings

The purpose of the find-actor-role-bindings function is to determine which noun phrases (i.e., actors) that occur in a story fill variables in the role patterns matched to the same story. Thus, the input to find-actor-role-bindings is a news story and the output is a set of actor-role bindings like the following:

| | |
|---|-----------------------------------|
| <i>actor</i> : Venceremos | <i>plays the role</i> : ?source |
| <i>actor</i> : Our people | <i>plays the role</i> : ?military |
| <i>actor</i> : our troops | <i>plays the role</i> : ?military |
| <i>actor</i> : five national guard soldiers | <i>plays the role</i> : ?victim |
| <i>actor</i> : enemy | <i>plays the role</i> : ?enemy |

Find-actor-role-bindings first calls the subroutine *find-noun-phrases* (to identify the noun phrases in the story) and then calls the subroutine *find-roles* (to match all of the defined role patterns against the text of the story).

Find-roles

For each role pattern, find-roles output either a list of lists of triplets or else NIL if the pattern does not match against the text of the story. For example, given a role pattern like this:

(?victim was assassinated by ?criminal)

find-roles might output a result like the following indicating that the role pattern had matched one time in the text of the story:

```
(((?victim 0 485)
 ("was" 485 488)
 ("assassinated" 489 501)
 ("by" 502 504)
 (?criminal 504 2114)))
```

Tokens with "?" prefixes represent variables in the role patterns (i.e., they represent role names). Other tokens are strings. The numbers in the triplets represent the number of characters from the start of the story. Thus, 0 represents the beginning of the story and, in this case, 2114 represents the final character at the end of the story. The two numbers in each triplet correspond to the starting and ending points in the story in which the token of the patterns was found. Note that a string like "was" spans three characters in the story while a variable like ?victim potentially spans 485 characters.

Find-roles is a simple pattern matcher which cannot, by itself, determine which pieces of text from the story fill the role variables. The role variables, like ?victim and ?criminal, are assumed to be filled by noun phrases. Find-roles returns a range of positions within the text in which the actor that fills a role might occur. Find-actor-role-bindings combines the output of find-roles with the output of *find-noun-phrases* to create the actor-role bindings.

Find-noun-phrases

Given the text of a story find-noun-phrases first employs the Xerox part-of-speech tagger (Cutting et al., 1993) to tag the words of the story. The Xerox part-of-speech tagger uses the set of about fifty part-of-speech tags originally developed by linguists to tag the Brown Corpus (Francis and Kucera, 1982). The tags include specializations of standard dictionary-type parts of speech labels, like "noun," "verb," "adjective," "adverb," etc. For the use of the find-noun-phrases routine, certain of the tags have been designated as, initiators, tags which might begin a noun phrase (e.g., determiners like "the" often begin a noun phrase), other tags are designated as possible constituents of a noun phrase (e.g., adjectives are often included in noun phrases) and other tags are designated as possible "stand-alone" noun phrases (e.g., nouns and pronouns can stand alone as noun phrases). After tagging the text of the story, find-noun-phrases attempts to group together initiators, constituents, and stand-alones into noun phrases. Thus, after processing a sentence like "Bill went to the movies" would become (noun-phrase "Bill") "went to" (noun-phrase "the movies"). Find-noun-phrases is a slightly modified version of a noun phrase extractor written by Prof. Haase.

Find-actor-role-bindings, find-roles and find-noun-phrases

To construct a set of actor-role bindings *find-actor-role-bindings* combines the results of *find-roles* with the output of *find-noun-phrases*. Each role variable in a role pattern occurs, by definition, either at the beginning of the pattern (in which case it is immediately followed by a string in the pattern), at the end (in which case it is immediately preceded by a string in the pattern) or within the role pattern (in which case it is immediately preceded and followed by string in the role pattern). For each role variable matched by *find-roles*, *find-actor-role-bindings* looks through the result of *find-noun-phrases* to see if a noun phrase can be found that abuts a string in the role pattern which either precedes or follows the variable in the role pattern. So, to fill in the variables in the role pattern

(?victim was assassinated by ?criminal)

find-actor-role-bindings would look for a noun phrase which ends just before the "was" token starts (for the ?victim variable) and would look for a noun phrase which starts just after the "by" token.

The output of *find-actor-role-bindings* is, in fact, a bit more complex than simply a list of actor-role pairs. The output is a list of structures which record the actor-role pairs, the start and end positions (in the text of the story) for the actor-role pairs, and two other slots (coreferences and actor-groups) which are filled in and used by later steps of the analysis (namely the routines *coreference-actors*, *construct-high-level-weighted-actor-role-pairs*, and *identify-point-of-view*). The structures output by the *find-actor-role-bindings* are called *extended-bindings*. *Extended-bindings* have four slots:

- (1) a noun-phrase slot which is filled by a list of words and their start and end positions in the text; this slot represents the actor;
- (2) a role variable slot which names the role against which the actor was matched by *find-actor-role-bindings*;

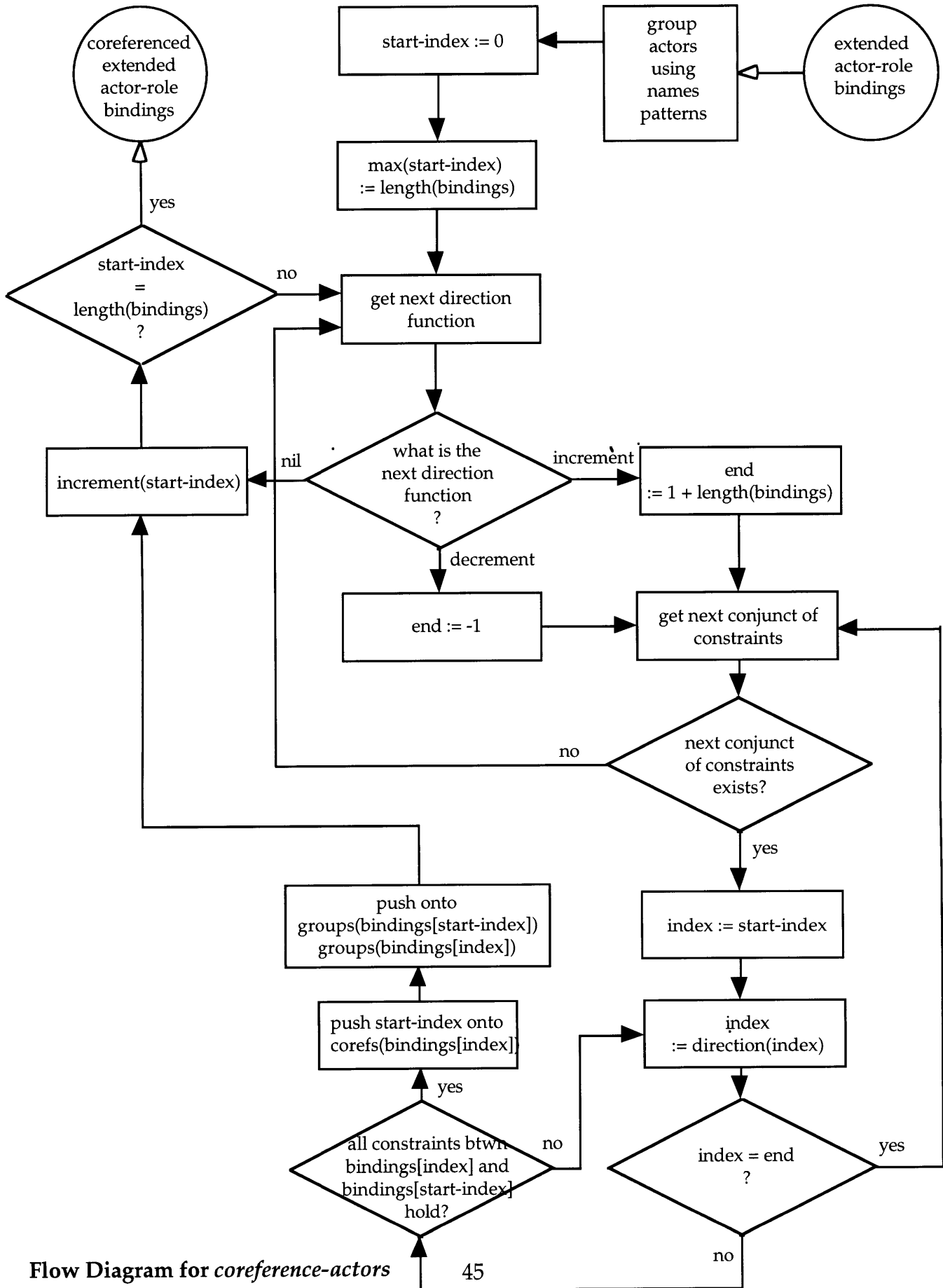
- (3) a coreferents slot which, at this point in the analysis of the story is empty; and,
- (4) an actor-groups slot which is also empty but which will later be filled by a list of groups (social, political, economic) that the actor might belong to; thus, for example, President Cristiani might be grouped together with other government officials.

Coreference-actors

The function *coreference-actors* starts with the output of *find-actor-role-bindings* and begins by matching a set of actor-grouping patterns against the noun phrases identified as actors by *find-actor-role-bindings*. For example, Maj. Rodriguez can be categorized as member of the Army by noting that his title "Maj." is a rank that only Army officers have. Here is the definition of the actor *army-officer* used in the *SpinDoctor* system:

```
(defactor army-officer
  :member-of  army
  :names      ((gen)
                (colonel)
                (col)
                (major)
                (maj)
                (lieutenant)
                (lieut)
                (officer)
                (officers))
  :pronouns   (*he* *she*))
```

The lists in the "names" slot of the definition are the patterns that are used to group actors. To determine if a given noun phrase is an instance of an actor (or group of actors as actors can be groups like the Army or the government) each of the words in the names slot is matched against the words in the noun phrase; if all of a given list of words in a name have correlates in the noun phrase, then the noun phrase is categorized as being in that group of actors and the "groups" slot of the extended-binding in which the noun



Flow Diagram for coreference-actors

phrase is found is filled with the name of the actor/actor group matched.

Actors are organized into larger groups of actors, thus. "army-officer" is defined to be a "member-of" the "army." (See Appendix A for a listing of the current database of actors used by SpinDoctor; the definitions include statements of "member-of" relations between the actors.) Analyze-story uses the output of coreference-bindings to identify how the most inclusive groups of actors are represented in a text, i.e., to identify which members of which groups of actors are assigned which roles.

After grouping the actors using the words contained in the noun phrases, coreference-actors then applies a set of constraints in order to try to resolve pronoun references and in order to partition all of the actors into mutually exclusive sets. It thereby identifies that, for example, a reference to "Maj. Rodriguez" and a reference to "Rodriguez" both represent the same actor and, furthermore, that the references to "Rodriguez" and "General Bustillo" both are instances which represent the more inclusive actor named the "Army."

The following constraints are used by coreference-actors:

- (1) the "member-of" relations which detail which groups of actors include other groups; e.g., "army-officers" is contained in the more inclusive actor group "army;"
- (2) the positions of the actors in the text of a story; e.g., the coreference procedure is written to first look upwards, toward the beginning of a story, to find a referent for a pronoun;
- (3) the role(s) assigned to an actor; e.g., if "Maj. Rodriguez" is determined to play the role of ?source in an article, then it is expected that other noun phrases to unknown actors (e.g., "the spokesman" or the "interviewee") which are also identified as ?sources would be more likely to be matched to "Maj. Rodriguez" than unknown actors which play other roles (e.g., the role of ?victim).
- (4) the pronouns associated with certain known actors; e.g., president-cristiani is defined in the database of SpinDoctor to have

a list of appropriate pronouns which include "he" and "him" but not "she" or "it;"

(5) the words in the noun phrases which constitute the actors; e.g., "Maj. Rene Rodriguez Hurtado" matches the actor "Maj. Rene Rodriguez" better than it matches "Gen. Bustillo."

Coreference-actors has been written to use all of the above constraints to coreferences actors. However, if a coreferent for an actor cannot be found, then the constraints are dropped one-by-one until only a subset of the constraints remain and either (a) a coreferent is found, or (b) no coreferent is found even with the smallest allowable set of constraints.

The control structure of coreference-actors has been designed to apply constraint (2) by first searching through all of the other actors which are positioned above the current actor in its search for a coreferent for the current actor before it searches through all of the actors positioned below the current actor. The extended-bindings input to coreference-actors are sorted by position in the story text. I.e., a binding which contains in its actor slot a noun phrase found at the start of the story will be in the sorted list of bindings before a binding which contains in its actor slot a noun phrase found at the end of the story. In the flow diagram of coreference-actors shown in Figure 3, the direction of the search can be seen to be controlled by a function called the "direction function." The "direction function" is initially set to "decrement" (i.e., to look through the actors positioned above the current one). If looking above the current actor yields no possible coreferent, then the "direction function" is set to "increment" to allow a search below the current actor to proceed.

Coreference-actors applies the remaining constraints in a slightly different manner depending upon whether the current actor is or is not a pronoun. If the current actor is a pronoun, constraint (5) is not applied, e.g., an earlier occurrence of "he" is considered a poor coreferent for the pronoun "he" even though the words of the noun phrases match exactly. For both pronouns and non-pronouns, non-pronoun coreferents are sought. A further specialization exists for the pronouns "I" and "we." "I" and "we" are

assumed to refer to actors which fill the role of ?source. Consequently, in the case of "I" and "we" pronouns constraint (3) is dropped in favor of a constraint which favors coreferents which fill the role of ?source (i.e., one might say that constraint (3a) is the role of the actor is assumed to be ?source).

The program uses the following "derefinement operators" to slowly relax the constraints if no coreferent can be found for an actor. The derefinement operators are lists of conjuncts of constraints. If the first conjunct of constraints in a list cannot be fulfilled, then the coreference-actors picks the next conjunct in the list and tries again, until the end of the list. Here, the conjuncts are represented by lists of numbers which refer to the number constraints above:

derefinement operator for non-pronouns:
((3 1 5) (1 5) (3 5) (3 1) (5) (1) (3))

derefinement operator for "I" and "we" pronouns:
((3a 1) (3a) (1))

derefinement operator for other pronouns:
((3 1) (3) (1))

The procedure coreference-actors propagates the relationships that it finds between actors and so, in the end, actors may be found to have several groups and several coreferents. The effect of coreference-actors is to fill in the "coreferents" and "actor-groups" slots of the extended-binding structures output by the find-actor-role-bindings routine.

Construct-high-level-weighted-actor-role-pairs

Given the extended-binding structures output by coreference-actors, the routine construct-high-level-weighted-actor-role-pairs further abstracts the actors into their most inclusive groups and then counts the number of times actors from each of the most inclusive actor-groups play each role matched in the text of the story under analysis. Analyze-story matches these weighted actor-

role pairs against the set of point of view definitions it has in its database to determine the point of view most likely represented by the story.

Each list of groups in an extended-binding structure is a list of actors. So, for example, a binding which contains the noun phrase "President Cristiani" in its actor slot might also contain the list of actors (president-cristiani government-official government) in its actor-groups slot. Construct-high-level-weighted-actor-role-pairs finds the most inclusive group in each of the actor-groups slots; it uses the "member-of" declarations in the actor definitions to determine the most inclusive group. In this example, president-cristiani is a member-of government-officials which is a member-of the government and so government is the most inclusive group in this case. Right now the system does not support multiple member-of declarations for an actor. Obviously this would be a worthwhile and interesting extension.

After construct-high-level-weighted-actor-role-pairs has determined the most inclusive group for each binding it then goes through and counts the number of times each most inclusive actor group is bound to a particular role. For instance, if "President Cristiani" appeared in the role of ?source once, then construct-high-level-weighted-actor-role-pairs would create a weighted actor-role pair like this (assuming that no other government-official appeared as a ?source in the story):

(government ?source 1)

The output of construct-high-level-weighted-actor-role-pairs is a list of weighted actor-role pairs.

Identify-point-of-view

The routine identify-point-of-view performs the final stage of analysis. Given a list of weighted actor-role pairs, identify-point-of-view compares them with each of the points of view defined in the database to determine which point of view matches best to the

actor-role pairs found in the story. Here, for example, are the four points of view that were defined for the set of analysis shown in Appendix C:

```
(defpov government
  ((government ?source)
   (government ?military)
   (guerrillas ?terrorist)
   (guerrillas ?criminal)))
```

```
(defpov guerrillas
  ((government ?criminal)
   (guerrillas ?source)
   (guerrillas ?military)))
```

```
(defpov democratic-convergence
  ((democratic-convergence ?source)
   (democratic-convergence ?victim)))
```

```
(defpov catholic-church
  ((catholic-church ?source)
   (catholic-church ?victim)))
```

The weighted actor-role pairs that were created after the analysis of the story MUC-15 (which is shown at the top of the section describing the routine analyze-story) were these:

```
((guerrillas ?source 1)
 (guerrillas ?military 2)
 (government ?victim 1)
 (unknown ?enemy 1))
```

The point of view that identify-point-of-view matched the weighted actor-role pairs to was the guerrillas point of view.

Identify-point-of-view takes each of the weighted actor-role pairs and matches them against each of the points of view. If one of the actor-role pairs matches a pair in the point of view, then the weight associated with the pair is added to the score for the point of view. The point of view which has the largest score -- after the weighted actor-role pairs have been matched against all of the

points of view -- is output as the point of view which matches the story. If two or more points of view have the same score then they are all output. Thus, at its most discriminating identify-point-of-view returns one point of view, while at its least discriminating, i.e., when it cannot determine the point of view for the story, identify-point-of-view outputs the list of all of the points of view defined in the database.

Chapter 4: An Extended Example of Actor-Role Analysis

The purpose of this chapter is to work through an actor-role analysis of a specific subset of a corpus of texts -- news stories about El Salvador -- in sufficient detail to express some of the most crucial knowledge necessary to distinguish a handful of differing points of view. It is then described how this knowledge can then be coded into SpinDoctor's database. With a database of actors, roles, and points of view, SpinDoctor is capable of analyzing a corpus of news stories automatically. Results describing SpinDoctor's performance are included in this chapter. Representations of actors, roles, and points of view used by SpinDoctor can be found in Appendices A, B, and C.

It is often possible, upon reading the first sentence or two of a news story, to determine the storyteller's point of view. For example, consider the following lead sentences from two different news reports about two different events:

1. Clandestine, 30 Mar 89 (Radio Venceremos) -- A report from San Miguel Department states that FMLN antiaircraft units ambushed a Salvadoran Air Force helicopter in Arenales canton at 1700 GMT on 29 March, as it was traveling from Usulután to San Miguel department. Our unit's effective antiaircraft fire hit and damaged the helicopter. [MUC3-0096]
2. San Salvador, 9 Jan 90 (DPA) -- The Salvadoran Army today prevented the occupation of cities in the eastern part of El Salvador, waging strong clashes between midnight and dawn, according to reports by military sources. [MUC3-0006]

One does not need to know about Salvadoran news agencies nor much about the past civil war in El Salvador to notice clues which give away the storytellers' points of view: In story 1 the author notes the actions of the FMLN's antiaircraft units and then refers to the antiaircraft units as "ours." Clearly the use of the possessive shows that the first story is told from the point of view of the

FMLN. In story 2 the point of view is less clear, but can be identified nonetheless. First of all, "military sources" are said to be the main source of the story. Secondly, and more subtly, it was said that the Army "prevented the occupation of cities" instead of, for example, "prevented the liberation." Together, these two clues mark story 2 as one told from the Salvadoran Army's point of view.

One does not need to know much about El Salvador to distinguish the two points of view represented in stories 1 and 2. Nevertheless, one needs to know *something*, if only enough to know that the FMLN and the Salvadoran Army are not different names for the same entity. It will be claimed that at least some of that "something" which one needs to know can be coded in the actor, role, and points of view patterns used by the SpinDoctor system. Before detailing what those patterns are for news stories about El Salvador, the corpus analyzed will be described and a very brief history of El Salvador presented.

The Corpus

The corpus is a set of 1300 news stories from Central and South America (El Salvador, Guatemala, Honduras, Argentina, Bolivia, Chile, Columbia, Ecuador, and Peru). These news stories are transcriptions of radio and television broadcasts that were translated into English by the Foreign Broadcast Information Service (FBIS). The FBIS is an entity of the government which largely works for the US intelligence agencies. The corpus was chosen for two reasons:

(1) It contains stories from a great number of widely differing points of view (or at least the FBIS's representation of those points of view). For example, broadcasts from the El Salvadoran rebel's illegal radio station, Radio Venceremos are included along with broadcasts from the official press agency of the Salvadoran armed forces, Coprefa.

(2) The corpus was originally chosen and used in the Third Message Understanding Conference (MUC3, 1991) sponsored by the Defense Advanced Research Projects Agency (DARPA). MUC3 (and its successors MUC4 and MUC5) was fielded by DARPA to test the leading artificial intelligence natural language processing (NLP) programs against one another. I chose the corpus, not with the intent to be a late entrant in the MUC3 competition, but rather to make it easier for the interested reader to see what sorts of information an actor-role analysis gives that no other existing NLP programs provide.

A further discussion of this issue will be presented in the Related Work chapter of this thesis. For the moment, it will suffice to say that the MUC3 entrants were instructed to program their computers to find all instances of “terrorist events” in the MUC3 corpus of news stories (or “messages” as they are termed at the MUC gatherings).

So, while the purpose of a MUC3 entrant was to count “terrorist events” my purpose is rather different. I am interested in “terrorist” and “terrorism” as pejorative terms indicative of roles which are assigned to certain parties of a conflict by other parties (most often official, governmental sources). Furthermore, an actor-role analysis of various points of view would naturally lead one to represent DARPA’s point of view and the point of view DARPA shared with the participants in the MUC3 competition. A representation of this point of view could be built around the ground rules of the competition which dictate, for example, that non-governmental organizations are normally to be assigned the role of “terrorist” and governmental parties, or members of the official armed sources of the military are only to be assigned the role of “terrorist” in exceptional cases (Rules 1 and 2 of the MUC3 competition; MUC3, 1991: p A1-A2). These ground rules seem to quite explicitly align the DARPA officials and the MUC3 entrants with the official US policy on, for example, El Salvador, and this point of view (or even *bias*) can of course be represented, at least partially, with the actor-role methodology. Representing the same

sort of bias with the methodologies and technologies of the MUC3 participants is likely impossible and, even if not impossible, highly unlikely to occur due to the politics of the participant's research funding. So, the use of the MUC3 corpus in this thesis is simply to make it clear to the reader that almost all of the other technologies being developed for NLP (including those being developed by MUC participants) do not address the same concerns as the ones addressed by SpinDoctor, namely, the representation and recognition of point of view and ideology.

An Extremely Brief, Contemporary History of El Salvador

Even though the MUC3 corpus I chose to focus on only one country, El Salvador, to provide an extended example for this thesis.

At the time during which the MUC3 corpus was collected, in the late-eighties/early-nineties there was an ongoing civil war in El Salvador. El Salvador is one of the smallest and most densely populated regions in Central America. As a result, land ownership has long been an issue of contention. In 1881 and 1882 the El Salvadoran government abolished the peasantry's traditional communal forms of land ownership. This action essentially destroyed the communes -- known as *ejidos* -- left many of the peasantry homeless, and crushed the indigenous people's cultures and languages when their communal lands were subsequently seized. Those who seized the lands became a small minority of families who dominate the political and economic life of El Salvador. In 1932 inadequate living and working conditions sparked a rural uprising lead by, among others, Augustin Farabundo Marti. The landowners called upon the military to put down the uprising. The military responded by massacring 30,000 peasants. This unbridled violence both established a dictatorial rule of the country by the military and dissuaded all serious opposition by the peasantry until 1972.

In 1972 the mayor of San Salvador, Jose Napoleon Duarte, was elected president. His party, the Christian Democratic Party (PDC)

opposed the entrenched, military-supported party, the National Conciliation Party (PCN). The electoral defeat was insufferable for the military which, through fraudulent means gave General Carlos Humberto Romero the presidency. Many trace the violent civil conflict of the 1980s to this event.

A group of young reformist-minded military officers overthrew General Romero in October of 1979. This civilian-military junta began to implement some forms of land reform, but the junta government was almost immediately brought under the control of the conservatives in the Army. In 1980 five insurgent organizations (the FAPL, FAL, ERP, FARN, and FARPL) organized themselves under the umbrella of the FMLN (Farabundo Marti National Liberation Front) and intensified their fight against the armed forces. At the same time the Left's political umbrella organization, the FDR (Revolutionary Democratic Front) was formed, and the two organizations functioned together. The "free" elections (sponsored by the Reagan administration) of 1982 excluded the FMLN-FDR, as did the "free" elections of 1984 and 1989.

In 1981 Jose Napoleon Duarte was serving as the president of the military-civilian junta government. At this time, and frequently throughout the 1980s, the civilian rural population was raped, tortured, mutilated, and murdered in genocidal sweeps of the countryside by the Salvadoran armed forces. In 1981 alone there were 13,229 political killings and 1,679 arrests and disappearances.³ The victims included both high profile, non-government, persons, e.g., the 1980 assassination of Monsignor Oscar Romero, the Archbishop of San Salvador, and the 1981 torture and murder of the leadership of the FDR), but many of the victims were large groups or even entire villages of peasants in the countryside (e.g., the infamous rape, torture, and massacre of the inhabitants of El

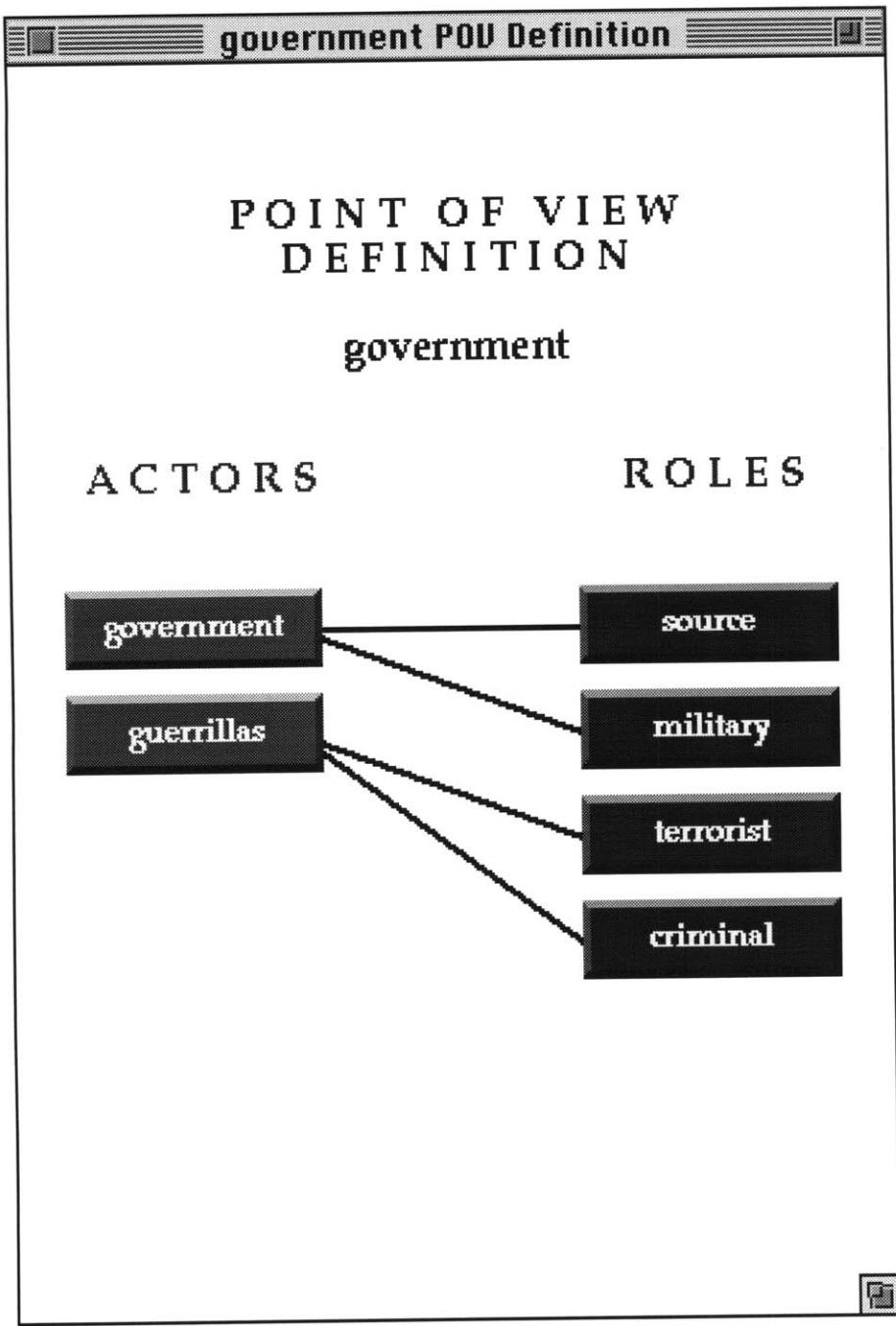
³Centro Universitario de Documentacion e Informacion (CUDI). 'La violacion de los Derechos Humanos en El Salvador', *Estudios Centroamericanos* (ECA), May/June 1982, pp. 543-556; Socorro Jurido, 'Cuadros de violaciones a los Derechos Humanos en El Salvador durante 1981', *Estudios Centroamericanos* (ECA), Dec1981, pp. 1194-1197; Washington Office on Latin America, 'Sintesis del Informe de la Comision Interamericana de Derechos Humanos de la OEA sobre la Violacion de los Derechos Humanos en El Salvador', Washington, DC, Dec 1981.

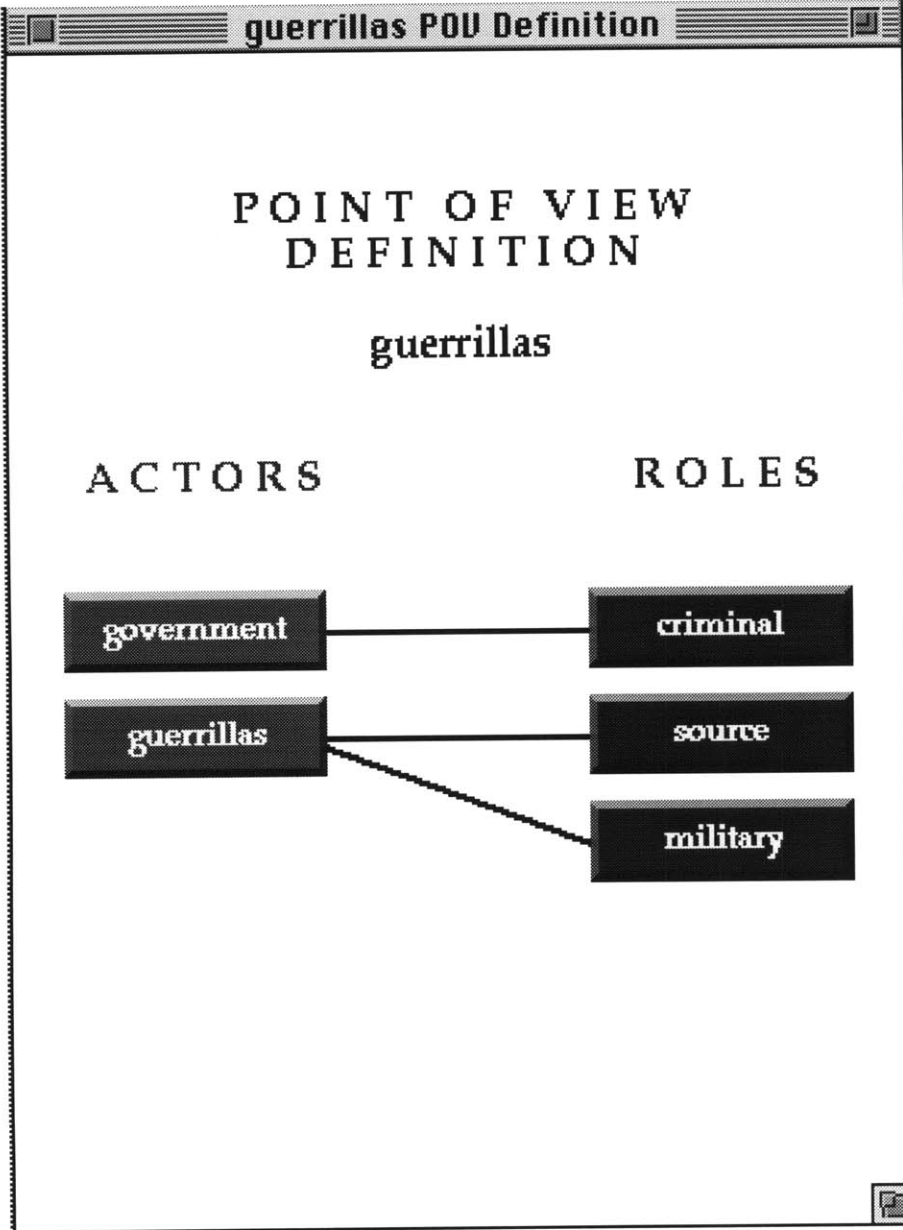
Mozote). Throughout the 1980s the Reagan and Bush administrations sent tens of billions of dollars of support and military training and aid to the Salvadoran government and its armed forces.

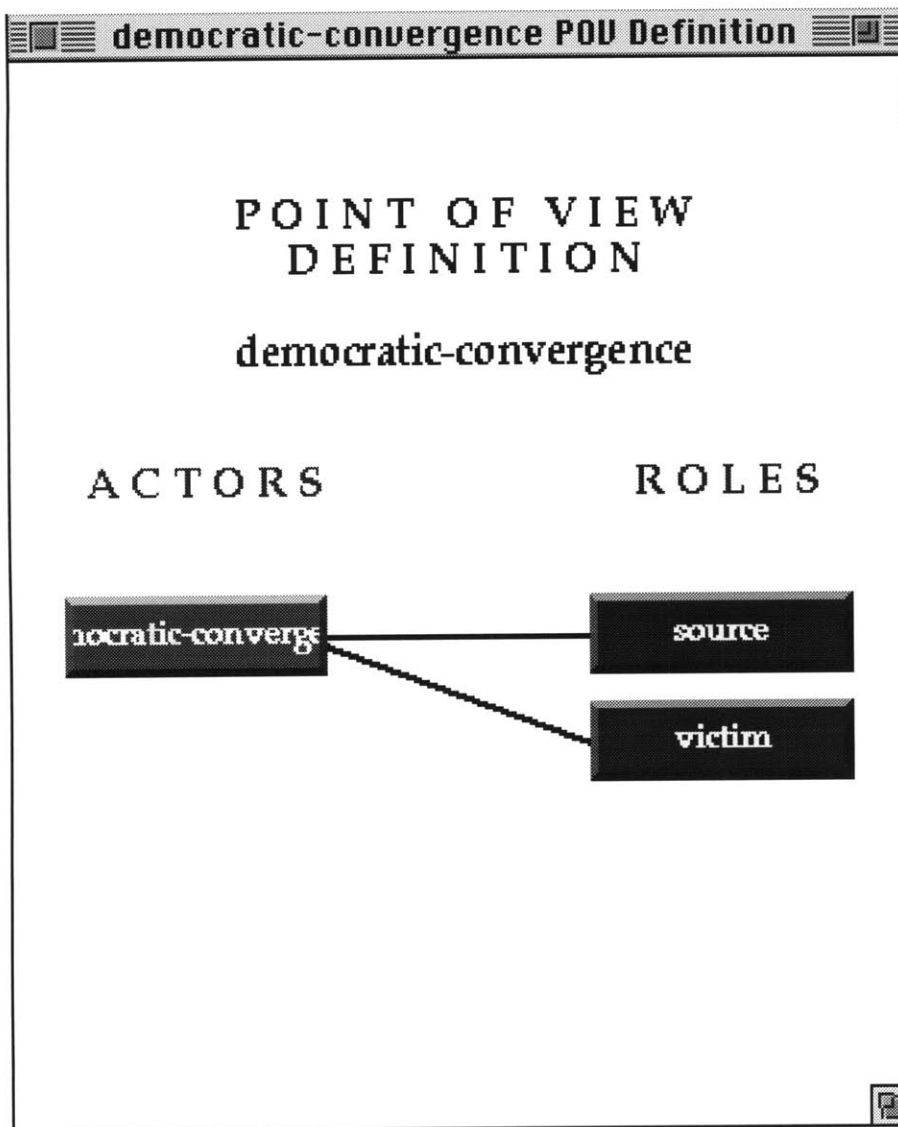
It is worthwhile to note that the ground rules of the MUC3 and MUC4 competitions effectively did not allow the large-scale massacres committed by the Salvadoran armed forces to be labeled as “terrorist events.” Rule 6 of the competition ground rules states that all events over two months old are not to be counted (MUC3, 1991: p. A2). Rule 7 states that incidents are only to be recorded if they are “specific” (MUC3, 1991: p. A2). Routinely it would take weeks for the atrocities of the Salvadoran military to surface and months thereafter for the “specifics” to be collected, e.g., the names of the peasants who died. Thus, even though the FMLN’s radio station, Radio Venceremos, comments on the massacre committed at El Mozote, it cannot be counted under the rules of the MUC competitions both because it is not recent enough and not enough “specifics” were included with the commentary.

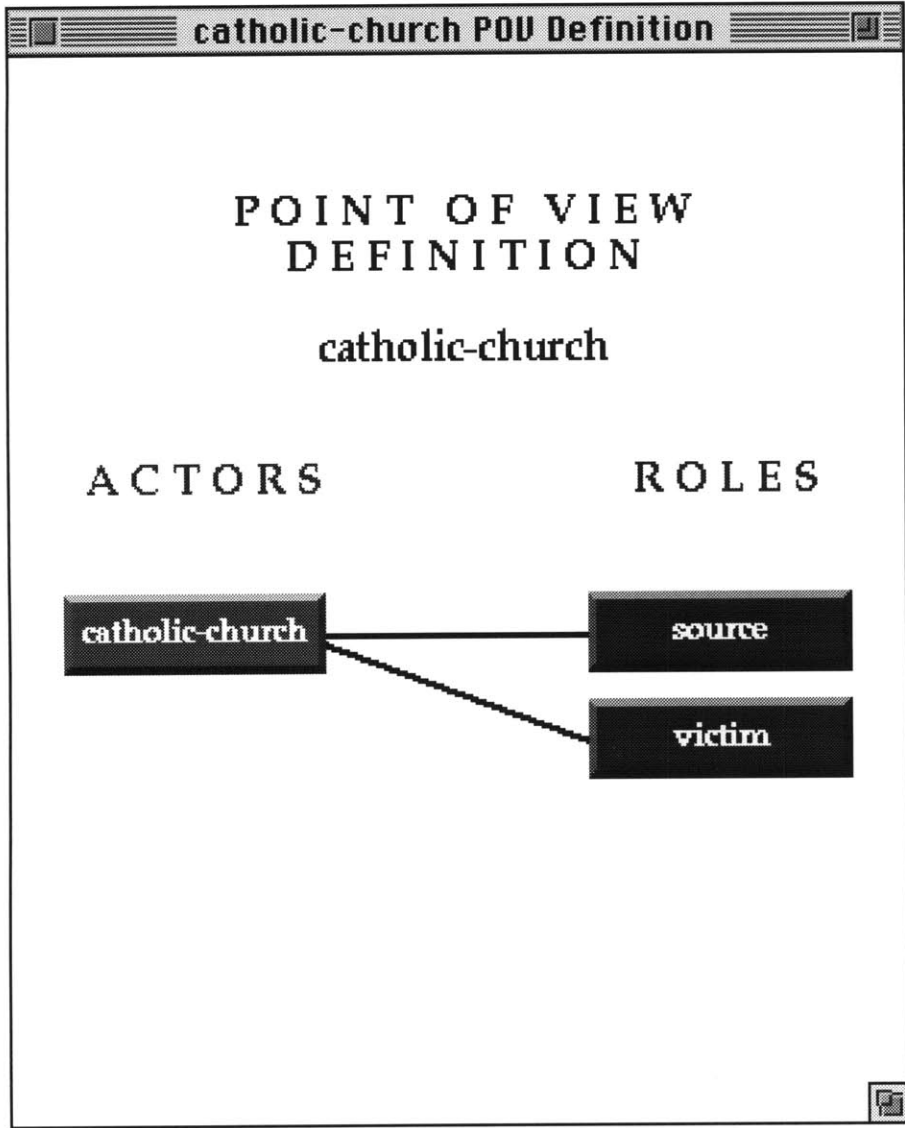
Representations of Salvadoran Actors, Roles, and Points of View

As one might imagine, the news stories analyzed were characterized by highly polarized points of view two of which dominated: the government/military point of view and the guerrilla (FMLN-FDR) point of view. Mediating the polemics were, at least, two parties: the Catholic Church and the Democratic Convergence (a center-left coalition of political parties not allied with the Salvadoran government). The next four figures depict the four points of view (government, guerrilla, Catholic Church, and Democratic Convergence) as I chose to represent them in the SpinDoctor database. (The figures illustrate a part of the graphical interface that a user can use to explore SpinDoctor’s databases.)









Probably due to the very violent nature of the news, a means to discriminate between these four points of view became immediately apparent. The government's point of view could be detected by looking for news articles in which the government was the source of the story, the guerrillas were characterized as criminals and terrorists, and, in which, the violent actions of the government and military were characterized as legitimate military actions (rather than criminal actions). The stories which represented the guerrilla's point of view often had the guerrillas as the source of the story and characterized the guerrilla's violent actions as legitimate military actions while assigning the role of criminal to the Salvadoran government and armed forces. In other words, the government, in these news stories, called the guerrillas terrorists and criminals while the guerrillas reciprocated by calling the government criminals. Displays of complementary epideictic rhetoric of this sort were the means by which the points of view could be readily discerned.

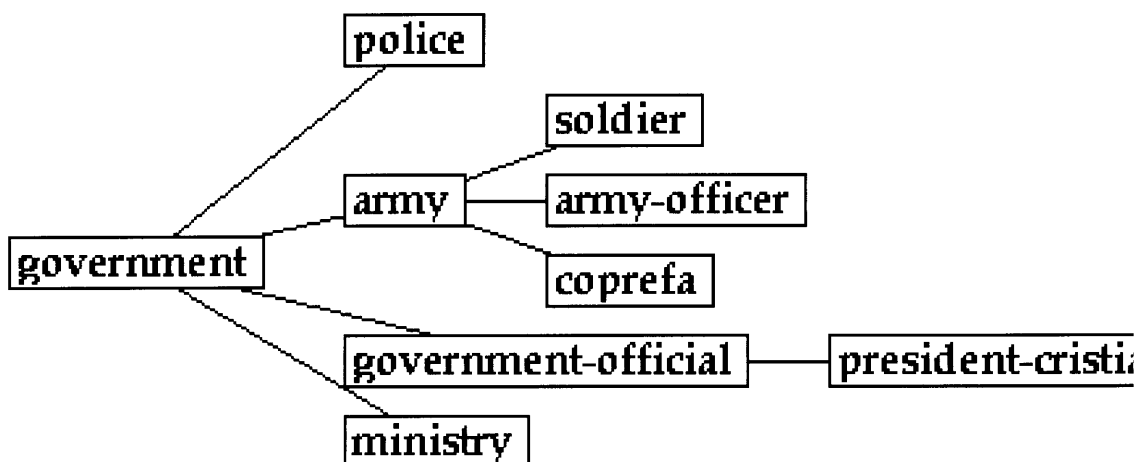
The two other points of view represented are probably also largely shaped by the violence of the civil war. The Catholic Church and the Democrat Convergence both suffered at the hands of the government and armed forces. Priests and leading politicians in the Democratic Convergence (a center-left organization) were murdered. Neither the Catholic Church nor the Democratic convergence had any military presence. It is not surprising then that neither of these parties engaged in the sort of name calling that characterizes the stories representing the government and the guerrillas. The Catholic Church and the Democrat Convergence were probably both too frightened to accuse the perpetrators of the crimes. Stories representing the Catholic Church tended to avoid blaming anyone for the deaths of its members (e.g., the Jesuits and their maids who were murdered by the Salvadoran armed forces in 1989). The Democratic Convergence also tended to not to cast anyone in the role of criminal when discussing the death of one of its members. Instead, these stories can be recognized by, for example, the use of the

Catholic Church as the source of a story and the assignment of one or more members of the Catholic Church as victim(s).

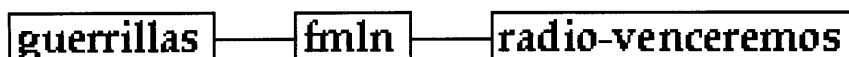
The roles of criminal, terrorist, source, military, and victim were key to this actor-role analysis. The interested reader can find in Appendix B the collections of verbs, adjectives, and phrases which I found to be indicative of these roles. For example, verbs like “murder,” “kill,” and “kidnap” were indicative of criminals while “capture,” “seize,” and “repel [a force]” were characteristic of military actions. I.e., the guerrillas might say that the armed forces murdered a number of people in a village, while the armed forces would say that they had seized the village.

The groups of actors identified for this analysis included the four most inclusive groups each with its own point of view (i.e., government, guerrillas, Democratic Convergence, and Catholic Church). Furthermore, each of these most inclusive groups are composed of a number of smaller, less inclusive groups. For example, the government includes the police, the army, the army’s officers and soldiers, government officials, ministries, and President Cristiani himself. Actor groups are declared in the “member-of” slots of the actor definitions which can be seen in Appendix A. SpinDoctor’s graphical interface also includes a way of browsing the actor groups. Actor groups are diagrammed as trees. The most inclusive group (the “root” of the tree) is positioned on the left-hand side of the diagram. The least inclusive groups (i.e., the “leaves” of the tree) are listed on the right-hand edge of the diagram and the intermediately inclusive groups are positioned in between the root and the leaves. Actor group diagrams for the government (and its subgroups), the guerrillas, the Democratic Convergence, and the Catholic Church can be seen in the following figure.

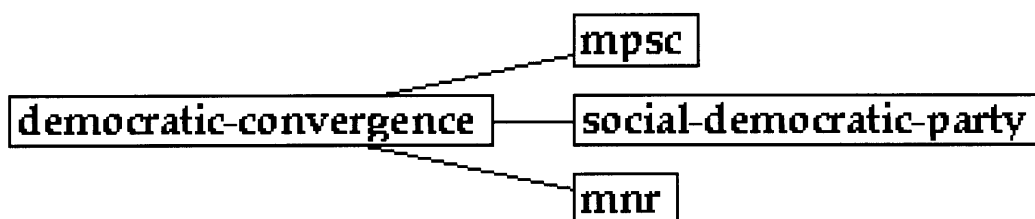
Actor groups included by government



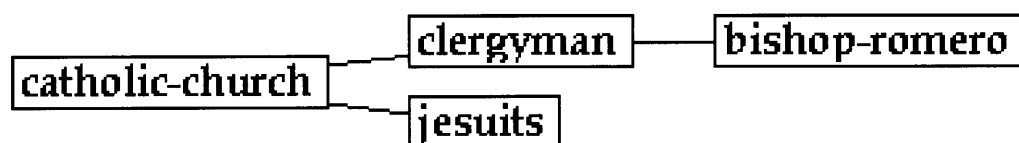
Actor groups included by guerrillas



Actor groups included by democratic-convergence



Actor groups included by catholic-church



Method

To build up the actor, role, and point of view representations the first twenty-five stories about El Salvador in the MUC3 corpus were scrutinized. Characteristic textual patterns were noted and then coded into the actor and role definitions. The point of view definitions were then defined by trying to articulate the strongest ideological differences that seemed to be represented in the stories and then by sketching out the sorts of actor-role assignment regularities that were discussed in the last section (e.g., that the guerrillas call the government criminals and vice versa). The database shown in Appendices A, B, and C took about two days to create.

Results

After coding the databases using the first twenty-five stories about El Salvador, SpinDoctor was then run on the first one hundred stories in the MUC3 corpus. Tables 1, 2, 3, and 4 summarize the results of that test. The columns labeled right contain the number of times in which SpinDoctor was able to pick which one of the four points of view was represented by a given story. The columns labeled wrong contain the number of times in which SpinDoctor picked a subset of the points of view (often a subset of one point of view) for the story and none of the elements of the subset were actually represented in the story. This was the case, for example, when the story was written from the point of view of the El Salvadoran government and SpinDoctor incorrectly deduced that the story either represented the Catholic Church or the Democratic Convergence. The columns labeled vague count the number of times in which SpinDoctor selected a superset of the points of view actually represented in the story. In such a case, SpinDoctor was not actually wrong, the stories did indeed represent a subset of the points of view selected by SpinDoctor. However, this was the case when, for example, SpinDoctor reported that a story represented both the Catholic Church and the

government when, indeed, perhaps it only represented the government's points of view.

Table 1: The Training Set: Twenty-five stories about El Salvador

| | right | wrong | vague | totals |
|-------------------|-------|-------|-------|--------|
| number of stories | 19 | 2 | 4 | 25 |
| percentage | 76 | 8 | 16 | 100 |

Table 2: Test Set #1: Seventeen unseen stories about El Salvador

| | right | wrong | vague | totals |
|-------------------|-------|-------|-------|--------|
| number of stories | 6 | 1 | 10 | 17 |
| percentage | 35 | 6 | 59 | 100 |

Table 3: Test Set #2: Fifty-eight unseen stories about other countries

| | right | wrong | vague | totals |
|-------------------|-------|-------|-------|--------|
| number of stories | 17 | 0 | 41 | 58 |
| percentage | 29 | 0 | 71 | 100 |

Table 4: Totals: One hundred stories seen and unseen

| | right | wrong | vague | totals |
|-------------------|-------|-------|-------|--------|
| number of stories | 42 | 3 | 55 | 100 |
| percentage | 42 | 3 | 55 | 100 |

As one can see in the Table 1, unsurprisingly, SpinDoctor did best with the subcorpus of stories that were used to create the databases (i.e., the first twenty-five stories about El Salvador). What is more interesting is to see is Table 2 which shows how well SpinDoctor did on stories about El Salvador which were not a part of the twenty-five used to create the databases of actors, roles, and points of view. In this test set of 17 stories, SpinDoctor was able to correctly determine the point of view represented by the news story 35% of the time, was wrong only once (6%), and was overly vague

in its assessment for the other 59% of the stories. This test result indicates that SpinDoctor, in conjunction with some other information retrieval technologies capable of narrowing down the topic of a story, could provide a user with a, at least partially effective, news filter to detect point of view. It is important to note that SpinDoctor was often overly vague, but rarely completely wrong in its assessment of point of view.

The three cases in which SpinDoctor was wrong are of interest. In two of the cases a particular source, from one group discussed a subject in a manner that would have been amenable to another group (E.g., in one case, the government repeatedly described a group of murdered priests as victims, i.e., the government repeated an actor-role combination characteristic of the Catholic Church's point of view). These two cases suggest that the sources of stories should be counted more heavily in the identification process. This could be trivially implemented by simply weighting the (<actor> ?source) pairs more heavily in the identify-point-of-view procedure of SpinDoctor's analysis routine.

The third case in which SpinDoctor was wrong, noted in Table 2, was not due to an underweighting of the sources, but rather due to a missing point of view. It just so happened that the first 25 stories about El Salvador are all from 1990 when Cristiani of the ARENA political party was already in power. However, the 17 stories represented by Table 2 were all from 1989 when the PDC was in power and Duarte was president. Consequently, in 1989 the ARENA party was a separate entity from the government, while in 1990 it was not. So, the database was built for the time when ARENA's point of view could be represented as equivalent to the government's. However, when SpinDoctor attempted to process stories from 1989 this assumption about ARENA was not valid and therefore caused it to pick the wrong point of view for one of the stories.

Perhaps what is the most surprising of the test results is that SpinDoctor behaved reasonable well with news stories which were neither from nor about El Salvador, but were rather descriptive of events in other countries in Central and South America. Table 3

shows these test results. With these stories, SpinDoctor was never wrong, overly vague 71% of the time, and guessed the point of view correctly for 29% of the stories. To a certain degree many of the other countries represented in the MUC3 corpus had, like El Salvador, serious internal, civil disputes. Consequently, some of the parties in these disputes are called “the government” and other parties are called “the guerrillas.” These structural similarities in the news discourses about various civil disputes in different countries allowed SpinDoctor to sometimes correctly detect the point of view represented by a news story about something that had no direct relation to the politics of El Salvador.

Table 4 shows the totals for all 100 hundred stories analyzed: 42 right, 3 wrong, and 55 overly vague.

Chapter 5: Related Work

Structuralism is 'anti-humanist', which means not that its devotees rob children of their sweets but that they reject the myth that meaning begins and ends in the individual's 'experience.' For the humanist tradition, meaning is something that I create, or that we create together; but how could we create meaning unless the rules which govern it were already in there? However far back we push, however much we hunt for the origin of meaning, we will always find a structure already in place. (Eagleton, 1983: 113)

The type of work exhibited in this thesis could be loosely termed (post)structuralist/constructivist.⁴ Some of the tools employed to construct the actor-role methodology are directly borrowed from the structuralist literatures on narrative (e.g., Greimas's thematic roles). The model of ideology that the actor-role representation has been directly compared to (Jameson's use of the semiotic square) is from the literature of post-structuralism. The most direct influence from media studies has been the analyses of the news done by van Dijk (van Dijk, 1988a; van Dijk, 1988b; van Dijk, 1991) which is informed by the (post)structuralist discourses about narrative theory. The most direct influence from the field of natural language processing (NLP) has been the work of Haase (e.g., Haase, 1994) which is the only work I know of in artificial intelligence (AI) which pays attention to what is known in the post-structuralist literature as "text" rather than, as most AI research on NLP does, exclusively focus on "plans" and "goals," i.e., teleology:

'Teleology', thinking of life, language and history in terms of its orientation to a telos or end, is a way of ordering and ranking meanings in a hierarchy of significance, creating a pecking order among them in the light of an ultimate purpose. But any such theory of history or language as a simple linear evolution

⁴ See Hawkes (1977) or Jameson (1979) for detailed descriptions of this genre of work. Or, see van Dijk (1988) for a shorter genealogy.

misses the web-like complexity of signs [...] It is that web-like complexity, indeed, which post-structuralism designates by the word 'text'. (Eagleton, 1983: 132)

It should also be acknowledged that the intention of this thesis is to create a theory of ideology and point of view and package it in a methodology not just as an intellectual exercise. The tools and methods developed from this work are intended to have pedagogical value at some point in their development: to teach others (and myself) to read the news with a critical eye. Ackermann's (e.g., Ackermann, 1991) politically-engaged, constructivist-style of education has been one of the main inspirations for some of my own work in the field of computers and education (e.g., Sack, 1993a) and has shaped the direction of this thesis as well.

Obviously the work of this thesis is multidisciplinary in nature and simply terming it (post)structuralist/constructivist may serve to alert the reader of some of my more obvious influences. However, the work can be positioned more exactly if the discussion is restricted to one of the disciplines that informs the work.

It may be possible to argue that the actor-role representation of ideology is a better model of ideology than others (e.g., Jameson's). However, the general concept of actor-role analysis is well known in cultural and media studies, even if my particular development and formalization of a kind of actor-role analysis is not. What is undeniably unique to this work is its representation of actor-role analysis in a form that is practical for the near-future recipient/producer of electronic news; namely, its encoding as a piece of software. Thus, perhaps the most appropriate set of work to compare the thesis work to is previous research in computational news analysis.

Other Work in Computational News Analysis

Computational news analysis is neither a professional field, nor an academic specialty. Rather, what this term is intended to

connote is that work that has been done in the fields of artificial intelligence and information retrieval to build computer programs to search, sort, and “understand” news stories. It will be impossible to mention more than a tiny fraction of the work accomplished to facilitate and automate these tasks. To position my own work I will need to caricature other, existing work, somewhat.

Work in computational news analysis comes in roughly two flavors: one flavor of work takes words seriously, the other flavor does not.

Those Who Take Words Seriously

Thus, for example, there is a long history in statistical style analysis in the humanities. One of the key studies produced by this subdiscipline is an analysis to determine whether certain of the *Federalist Papers* were authored by Hamilton or Madison. Mosteller and Wallace (1963) determined that most of the unsigned papers were authored by Madison by comparing the frequencies with which Madison and Hamilton employed a handful of prepositions, conjuncts, and articles. They found, for example, that Hamilton used the word “by” infrequently while Madison used it frequently. Using these frequency counts Mosteller and Wallace assigned the author of the papers to be Madison.

Within the social sciences there is a methodology analogous to the statistical methods employed in the humanities; the analogous method is known simply as “content analysis.” Many content analysts trace their modern origins to a set of quantitative newspaper analyses performed in the late-nineteenth century to show, for example, how religious, scientific, and literary matters had been replaced by gossip, sports, and scandal in leading New York dailies (Seed, 1893; Krippendorff, 1980). By the 1930s the giant in this field was Lasswell who began by studying propaganda of the First World War (e.g., Lasswell, 1927). Lasswell, his colleagues and students had a huge influence on many social scientists in almost all of the social science disciplines. Again, like

the humanities studies of style, the techniques of content analysis are largely based upon statistical techniques. The frequency and occurrence of particular words and phrases was tabulated in studies by anthropologists, folklorists, psychologists, sociologists and others of a huge range of texts with special attention for news texts.

The results of the Stanford *Symbol Studies* took a variety of forms. With Yule's K statistic, analyses were made of symbol variety to identify those papers and time periods in which there is a concentration of fewer symbols. Reports analyzing frequency patterns of internationalism, democracy, freedom, and doctrine were separately issued. Attitudes toward symbols were scored as being favorable, unfavorable, or neutral (Stone et al., 1966: 56).

What distinguishes the computational content analysts from the style analysts was their attempts to compile large dictionaries of grouped words and phrases wherein groups of words and phrases were assigned some larger theoretical significance. For example, Lasswell and his colleagues at the Hoover Institute at Stanford spent years compiling a dictionary of terms to analyze news and editorials and dictionary entries were made for categories like "internationalism," "democracy," "freedom," and "doctrine."

By the 1950s content analysts had discovered a use for computers in their field. At Harvard the General Inquirer system was programmed in the early 1960s and made available to researchers from many areas of the humanities and the social sciences (Stone, Dunphy, Smith, and Ogilvie, 1966) With the horsepower provided by the computer, the content analysts of the General Inquirer project were able to begin to pay attention, not only just to the co-occurrence of terms, but also, at times, also to the grammatical relations between terms. However, grouping terms into sets like "those which represent the concept of democracy" and counting their occurrence in texts continued to be the major occupation of the analysts.

The fruits of this type of computational content analysis can be seen in contemporary information retrieval software. Researchers like Gerard Salton were involved in building theories and technologies for the content analysts of the 1960s. Salton is now one of the leaders in the field of information retrieval (e.g., Salton, 1971; Salton, 1994; TREC, 1993). Salton and most of his colleagues in information retrieval represent texts as “vectors”:

The Smart system is a sophisticated text retrieval tool, developed over the past 30 years, that is based on the vector space model of retrieval. In the vector space model, all information items -- stored texts as well as information queries -- are represented by sets, or vectors, of terms. A term is typically associated with the text under consideration. In principle, the terms might be chosen from a *controlled vocabulary list* [Emphasis added] or a thesaurus, but because of the difficulties of constructing such controlled vocabularies for unrestricted topic areas, it is convenient to derive the terms directly from the texts under consideration. Collectively, the terms assigned to a particular text represent text content. (Salton et al., 1994: 1421).

Contemporary techniques of information retrieval are not simply keyword based, like the 1960s technologies were. Rather, today’s information retrieval technologies use morphological analyzers to find the roots of words, employ the convention of “stop words” (i.e., words like “the” and “and” are considered to be too numerous in all types of texts to serve as good query patterns and so they are “stopped” or rather taken out of queries posed by users), and have access to thesauri to find synonyms and word variants. However, even the newest information retrieval systems are fundamentally constructed around word frequencies and co-occurrences.

Salton’s reference to “controlled vocabulary list[s]” is, of course, an allusion to the sorts of dictionaries of groups of terms that Lasswell and his colleagues and students built. By trying to provide tools for “unrestricted topic areas” information retrieval

researchers have lost their interest in building such dictionaries preferring instead to replace them with statistical relationships between sets of words in the text corpuses they are working with, or with general reference thesauri, like Rogets Thesaurus or George Miller's Wordnet (Miller et al., 1993). Whereas Lasswell and his colleagues were attempting to do critical readings of propaganda and news, contemporary information retrieval specialists prefer to see all texts simply as "information." One of the points of this thesis is to point out that seeing all texts as "information" is folly. It matters very much who the source of a text was and how people are represented in the text.

Ironically, if contemporary information retrieval were to again take up the critical stance of the earlier content analysts, the products of the field would be quite close to SpinDoctor. In the dictionaries of Lasswell et al. certain terms were recognized as important and others were marginalized, just as in this thesis actors and roles are privileged and other aspects of the texts are practically ignored. The difference, however, between the hypothetical mix of technologies of content analysts and contemporary information retrieval and SpinDoctor would be the amount of sentence and discourse structure retained. In SpinDoctor there are noun phrases, coreference relationships between noun phrases, nouns, and pronouns, and positional constraints on the order of matched (role) phrases to, for example, determine which noun phrase represents the subject and which the object of a sentence. In information retrieval the "terms" used are typically just words or short sequences of words and the relationships recognized between words are restricted to a finite number of intervening words, or, even more, simply just to co-occurrence between two terms in a single text or set of texts.

Haase

As will be discussed in the next section, those who do not take words seriously, do not because, usually, they are more interested in complicated structures which can related words or terms

together to the exclusion of the terms themselves. However, there is some work that finds a balanced medium between information retrieval's refusal of complex relationships between words and an overinterest in abstract relationships to the exclusion of words. Haase's (1994) current work on analogical representation of texts employs a set of weak methods (in the artificial intelligence sense of weak, meaning generally applicable as opposed to problem specific) to group words into phrases and subjects, objects, actions, adverbial, and adjectival phrases together. After each text is parsed with the weak methods words and the grammatical relations between words are used to compare it to all previously parsed texts. The database of texts grows and similarly structured and worded texts are hashed together into the database. Haase's system uses a more sophisticated set of relationships between terms than the SpinDoctor employs. Consequently, one could imagine building a more sophisticated SpinDoctor system, not by programming the actor, role, and point of view patterns, but rather by "training" it: pointing out, naming and thus privileging certain constellations of words and grammatical relationships found by Haase's system.

Those who do not take words seriously

Unlike Haase's system which retains all of the grammatical and lexical information contained in the original texts, there exist a large number of artificial intelligence (AI) natural language processing (NLP) systems which effectively coerce a text into a small set of templates and in the rewrites of the text required to accomplish the coercion most of the original words and syntax are lost (or rather are not retained by the systems). Various kinds of epistemological investments have been made in the sorts of templates that these AI NLP systems use. The templates used have been called frames (Minsky, 1974), scripts, plans, and goals (Schank and Abelson, 1977), TOPs and MOPs (Schank, 1981), XPs (Schank, 1984), Taus (Dyer, 1983), and schemas (by a large number of people in cognitive science and psychology). Systems which used names like these for

their templates were often called “story understanding systems.” These various names are still used in some cases, but most of the more recent NLP work in AI tends to be less concerned with cognitive modeling and more concerned with building tools which will work on large corpora of texts. Consequently, many contemporary researchers tend to stick with the terms “form” or “template” (e.g., Jacobs and Rau, 1993).

Thus, to put it in the blandest of terms, contemporary AI NLP is concerned with building fast and accurate form-fillers, while information retrieval is interested in cross indexing texts by word frequency. Both of these fields seem to have capitulated to the position of seeing texts, not as strategic entities constructed by certain groups of people to influence, persuade, or entertain other groups of people, but rather as some kind of homogeneous material known as “information” (as it is defined in, for example, information theory). Unlike his later progeny, Lasswell was, for example, interested in propaganda and thus also ideology. Some of the “template-fillers” intellectual ancestors were also interested in ideology and point of view.

It is neither hard nor surprising to find that certain early members of the “story understanding” and the “content analysis” communities overlapped. For example, Robert Abelson worked both with one of Lasswell’s important colleagues in the 1950s (Ithiel de Sola Pool) and also later with Roger Schank, one of the proponents of “story understanding work.” Abelson’s work of the early 1960s is of most interest because he effectively straddled the concerns of the two communities. He was attempting to build computer programs which could do ideologically motivated readings of stories (e.g., Abelson and Carroll, 1964).

Unfortunately as Abelson’s work became to resemble more mainstream AI story understanding work, he, like other AI researchers came to place too much emphasis on goals, plans and actions of individuals, and not enough on names and the representation of group identities. In short, his work, and the work of his students (e.g., Carbonell, 1978) came to confuse ideology with teleology. Carbonell’s work, for example, attempted to represent

ideologies as trees of goals. Thus, a “conservative” ideology was represented as a descriptions of a set of plans and goals that Russia could take to achieve domination of the world, and a set of plans and goals that the United States could take to prevent Russia from doing so. The main inferencing performed by Carbonell’s system, POLITICS, was the set of inferences it constructed to counter-plan. It is obvious, fifteen years after the fact, that Carbonell’s representation of ideology was ridiculously naive. The representation had nothing to say about representation, it only described actions and counter actions, goals, plans, and counter plans. Furthermore, since it took counter-planning to be the canonical action of the groups represented, the work remained limited to binary oppositions.

What got “lost in translation” from Abelson’s early context analysis-like interest in ideology to his later support for representations of ideology as teleology was two things

(1) Since “story understanding systems” purportedly calculated more complex relationships between words than the correlations used by context analysts, the texts input to these systems were often toy texts and so an interest in large corpora was lost. Thus, for example, POLITICS (Carbonell’s system) was only able to “understand” a small set of very short news stories.

(2) What became of central interest in the “story understanding” literature was the templates that were filled by the story understanders rather than the way in which the stories were expressed. Obviously, when one does not pay attention to how something is stated the issues that concern rhetoricians and dialecticians are effectively marginalized. Thus, in the story understanding literature ideologically loaded texts, e.g., racists tracts or mass media propaganda, like that studied by Lasswell and his associates, were effectively parsed into the same sorts of representations as excerpts from instructional manuals for power tools.

My critique of the MUC3 systems, in the previous chapter, illustrates how those following the template-filling tradition continue to be blind to the issues of ideology and point of view. Nevertheless, the template-filling programs effectively perform the same sort of analysis as SpinDoctor's role matching in which it finds the subjects and objects of various verbs, adverbial, and adjectival phrases. However, the information that is most closely scrutinized by SpinDoctor, i.e., the information about which roles are assigned to whom, is attached no privileged significance by the template-filling programs.

It seems, in summary, worthwhile to underline what is perhaps obvious, but which few in contemporary computational news analysis seems to take heed of: computational technologies for news analysis embody in their very construction certain types of reading strategies. Like all readers, these systems ignore some parts of the texts analyzed and privilege others. Both those researchers in information retrieval (IR) and AI NLP seem to have adopted the point of view that texts are just information as opposed to entertainment, or arguments, or elements of discourse, in general. Consequently, both IR and AI NLP systems remain blind to ideological concerns that are embedded in all types of text. The purpose of this thesis simply to point out how those working in computational news analysis do not need to continue to turn a blind eye to these issues.

Chapter 6: Conclusions

This thesis has presented a theory and a piece of software to illustrate the theory. The theory, actor-role analysis, is a theory of how ideology and point of view can be represented so that news texts can be categorized according to ideological distinctions. The program, SpinDoctor, uses a computational form of the theory and can consequently automatically categorize texts. SpinDoctor has been tested on a small corpus of texts and the results of this test were reported.

Throughout the thesis the work described here is compared against work done by researchers in various fields to address similar issues. For example, it is claimed that actor-role analysis is a more nuanced model of ideology than the model proposed by Jameson (1981) which uses semiotic squares. SpinDoctor represents a kind of computational news analysis program that has few correlates in the literatures of information retrieval and artificial intelligence natural language processing. Although actors and roles can be found using a variety of contemporary computational text processing systems, none of these systems have been built to take special note of particular actor-role combinations. It is argued that actor-role combinations are noteworthy and, specifically, that they represent ideological distinctions.

It is possible that the SpinDoctor system could function as a set of filters for electronic news that would allow a user to define and then search for various points of view as they are represented in large corpora of news stories. Such a technology could certainly be used as a means of electronic censorship, but it could also be used by the curious to search out news stories presented from underrepresented points of view.

References

- Abelson, R.P. (1963) "Computer Simulation of 'Hot' Cognition", in S. Tomkins and S. Messick (editors) *Computer Simulation of Personality* (New York: Wiley).
- Abelson, R.P. (1973) "The structure of belief systems" in R. Schank and K. Colby (editors) *Computer Models of Thought and Language* (San Francisco, CA: W.H.Freeman)
- Abelson, R.P. and Carroll, J.D. (1965) "Computer Simulation of Individual Belief Systems", *American Behavior Scientist*, 8, 24-30.
- Ackermann, E. (1991) The Agency Model of Transactions: Toward an Understanding of Children's Theory of Control, in *Constructivism*, I. Harel and S. Papert (eds.), Ablex Publishing Corp.: Norwood, NJ.
- Ang, I. (1990) *Desperately Seeking the Audience*, Routledge: New York.
- Antworth, Evan L. (1990) *PC-KIMMO: a two-level processor for morphological analysis*. Occasional Publications in Academic Computing No. 16. Dallas, TX: Summer Institute of Linguistics. ISBN 0-88312-639-7, 273 pages, paperbound.
- Antworth, Evan L. (1991) *ENGLEX: An English lexicon for PC-KIMMO*. Unpublished user's manual. Summer Institute of Linguistics: Dallas, TX.
- Bourdieu, P. (1991) *Language and Symbolic Power*, Harvard University Press: Cambridge, MA.

Carbonell, J. (1979) *Subjective Understanding: Computer Models of Belief Systems*. Ph.D. Thesis. New Haven, CT: Yale University Computer Science Department, Technical Report 150.

Cutting, Kupiec, Pedersen, and Sibun (1993) A Practical Part-of-Speech Tagger, *Proceedings of the Annual Conference of the Association for Computational Linguistics*,

DeMarcken (1991) Parsing the LOB Corpus, in *Proceedings of the Annual Conference of the Association for Computational Linguistics*, pp. 243-251.

Dworkan, A. (1981) *Pornography: Men Possessing Women*, Putnam: New York.

Dyer, M. (1983) *In-Depth Understanding: A Computer Model of Integrated Processing for Narrative and Comprehension*, MIT Press: Cambridge, MA.

Eagleton, T. (1983) *Literary theory : an introduction* (Minneapolis: University of Minnesota Press).

Eagleton (1991) *Ideology: An Introduction*, Verso: New York.

Francis, W.N. and Kucera, F. (1982) *Frequency Analysis of English Usage*. Houghton Mifflin.

Gans, H. (1980) *Deciding What's News* (New York: Vintage Books).

Greimas, A.J. (1987) *On Meaning: Selected Writings in Semiotic Theory* (trans. P.J. Perron & F.H. Collins) (Minneapolis: University of Minnesota Press).

Haase, K. (1994) *Using Grammatical Relations for Document Retrieval and Analysis*, Working Paper, Machine Understanding Group, MIT Media Laboratory: Cambridge, MA.

- Haraway, D. (1989) *Primate Visions*, Routledge, New York.
- Hawkes, T. (1977) *Structuralism and Semiotics*, Berkeley, CA: University of California Press.
- Herman, E. and Chomsky, N. (1988) *Manufacturing Consent: The Political Economy of the Mass Media*, Pantheon, New York.
- Hobbs, J.R., Appelt, D., Typson, M., Bear, J., Israel, D. (1992) SRI International: Description of the FASTUS System Used in MUC-4, in *Proceedings of the Fourth Message Understanding Conference*, Morghan Kaufmann: San Mateo, CA.
- Jacobs, P. and Rau, L. (1993) Innovations in Text Interpretation, *Artificial Intelligence* , 63(1-2), 143-192.
- Jameson, F. (1977) *The Prison-House of Language: A Critical Account of Structuralism*, Princeton, NJ: Princeton University Press.
- Jameson, F. (1982) *The Political Unconscious* (Bloomington, IN: Indiana University Press).
- Jameson, F. (1987) Foreword, in Greimas, A.J., *On Meaning: Selected Writings in Semiotic Theory* (trans.P.J. Perron & F.H. Collins) (Minneapolis: University of Minnesota Press).
- Koskenniemi, Kimmo. 1983. *Two-level morphology: a general computational model for word-form recognition and production*. Publication No. 11. University of Helsinki: Department of General Linguistics.
- Krippendorff, K. (1980) *Content Analysis: An Introduction to its Methodology*, Sage: Beverly Hills, CA.

Lakoff, G. (1991) "Metaphor and War: The Metaphor System Used to Justify the War in the Gulf", *Journal of Urban and Cultural Studies*, Volume 2, Number 1, pp. 59-72.

Lasswell, H. (1927) *Propaganda Technique in the World War*, Knopf: New York.

Miller, G. (1990) Special Issue, Wordnet: An on-line lexical database. *International Journal of Lexicography*, 3(4).

Minsky, M. (1986) *The Society of Mind* (New York: Simon and Schuster).

Mosteller, F. and Wallace, D.L. (1963) Inference in an authorship problem, in *Journal of American Statistical Association* 58: 275-309.

MUC-3 (1991) *Proceedings of the Third Message Understanding Conference*, San Mateo, CA: Morgan Kaufmann Publishers, Inc.

MUC-4 (1992) *Proceedings of the Fourth Message Understanding Conference*, San Mateo, CA: Morgan Kaufmann Publishers, Inc.

Propp, V. (1968) *Morphology of the Folktale* (trans. L. Scott; ed. L.A. Wagner) (Austin, London: University of Texas Press).

Sack, W. (1993) Coding News and Popular Culture, *Presented at the IJCAI Workshop on Models of Teaching and Models of Learning*, Chambéry, France.

Sack, W. (1994a) Future News: Constructing the Audience Constructing the News, in *Proceedings of WRITE94*, Vancouver, BC.

Sack, W. (1994b) Indexing Multimedia by Ideology, in *Proceedings of the AAI Workshop on Indexing and Reuse of Multimedia*, Seattle, WA.

Sack, W. and Davis, M. (1994) IDIC: Assembling Videos from Story Plans and Content Annotations, in *Proceedings of the IEEE International Conference on Multimedia Computing and Systems*, Boston, MA.

Sack, W. Soloway, E. and Weingrad, P. (1993) Re:writing Cartesian Student Models, in the *Journal of Artificial Intelligence in Education*, special issue on Cognitive Models of Students, volume 3, number 4, Assoc. for the Advancement of Computing in Education, Charlottesville, VA.

Salton, G. (1971) *The SMART Retrieval System: Experiments in Automatic Document Processing*, Prentice-Hall, Inc.: Englewood Cliffs, NJ.

Salton, G., Allan, J., Buckley, C., Singhal, A. (1994) Automatic Analysis, Theme Generation, and Summarization of Machine-Readable Texts, in *Science*, Vol. 264, 1421-1426.

Saussure, F. (1966) *Course in General Linguistics* (eds. C. Bally, A. Sechehaye, A. Riedlinger; trans. W. Baskin) (New York: McGraw-Hill Book Co.).

Schank, R. (1982) *Dynamic Memory: A theory of reminding and learning in computers and people* (Cambridge: Cambridge University Press).

Schank, R. (1986) *Explanation Patterns: Understanding Mechanically and Creatively*, Lawrence Erlbaum Assoc.: Hillsdale, NJ.

Schank, R. & Abelson, R. (1977) *Scripts, Plans, Goals and Understanding: An Inquiry into Human Knowledge Structures* (Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers).

Steele, G. (1990) *Common Lisp: The Language, second edition*. Digital Press.

Stone, P., Dunphy, D., Smith, M., and Ogilvie, D. (1966) *The General Inquirer: A Computer Approach to Content Analysis*, MIT Press: Cambridge, MA.

Terdiman, R. (1985) *Discourse/Counter-Discourse: The Theory and Practice of Symbolic Resistance in Nineteenth-Century France*. Ithaca, NY: Cornell University Press.

Thompson, J.B. (1984) *Studies in the Theory of Ideology*. Berkeley, CA: University of California Press.

TREC (1993) *The First Text Retrieval Conference*, US Department of Commerce, National Institute of Standards and Technology: Gaithersburg, MD.

van Dijk, T. (1988a) *News as Discourse*. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

van Dijk, T. (1988b) *News Analysis: Case Studies of International and National News in the Press*. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

van Dijk, T. (1991) *Racism and the Press* (New York, London: Routledge).

Appendix A: Actor Patterns

(in-package :cl-user)

```
;;;=====
;;;
;;;
;;; ACTOR-PATTERNS.LISP
;;;
;;; Warren Sack
;;; 19 July 1994
;;;
;;;=====
```

```
(defactor government
  :names ((government))
  :pronouns (*it* *they*))
```

```
(defactor ministry
  :member-of government
  :names ((ministry))
  :pronouns (*it* *they*))
```

```
(defactor government-official
  :member-of government
  :names ((vice president)
          (presidential)
          (vice president)
          (secretary general)
          (attorney general)
          (minister))
  :pronouns (*he* *she*))
```

```
(defactor president-cristiani
  :member-of government-official
  :names ((cristiani)
          (alfredo cristiani burkard))
  :pronouns (*he*))
```

```
(defactor army
  :member-of government
  :names ((battalion)
          (garrison)
          (military)
          (infantry)
          (brigade)
          (national guard)
          (national guardsmen)
          (army))
  :pronouns (*it* *they*))
```

```
(defactor coprefa
```

```

:member-of army
:names ((coprefa)
        (armed forces press committee))
:pronouns (*it* *they*))

(defactor army-officer
:member-of army
:names ((gen)
        (colonel)
        (col)
        (major)
        (maj)
        (lieutenant)
        (lieut)
        (officer)
        (officers))
:pronouns (*he* *she*))

(defactor soldier
:member-of army
:names ((soldier)
        (national guardsman))
:pronouns (*he* *she*))

(defactor police
:member-of government
:names ((police)
        (policeman)
        (policemen)
        (prison guards)
        (prison guard))
:pronouns (*it* *they*))

(defactor arena
:names ((arena)
        (national republican alliance))
:pronouns (*it* *they*))

(defactor guerrillas
:names ((guerrilla)
        (guerrillas))
:pronouns (*it* *they*))

(defactor fmln
:member-of guerrillas
:names ((fmln)
        (farabundo marti national liberation front))
:pronouns (*it* *they*))

(defactor radio-venceremos
:member-of fmln
:names ((venceremos))
:pronouns (*it* *they*))

```

```

(defactor democratic-convergence
  :names ((democratic convergence))
  :pronouns (*it* *they*))

(defactor mnr
  :member-of democratic-convergence
  :names ((mnr)
           (national revolutionary movement))
  :pronouns (*it* *they*))

(defactor social-democratic-party
  :member-of democratic-convergence
  :names ((social democratic)
           (social democrat))
  :pronouns (*it* *they*))

(defactor mpssc
  :member-of democratic-convergence
  :names ((mpssc)
           (social christian people movement))
  :pronouns (*it* *they*))

(defactor civilians
  :names ((civilians))
  :pronouns (*it* *they*))

(defactor salvadorans
  :member-of civilians
  :names ((salvadorans))
  :pronouns (*it* *they*))

(defactor catholic-church
  :names ((the church)
           (the catholic church)
           (catholic))
  :pronouns (*it* *they*))

(defactor jesuits
  :member-of catholic-church
  :names ((jesuits)
           (jesuit))
  :pronouns (*it* *they*))

(defactor clergyman
  :member-of catholic-church
  :names ((pope)
           (cardinal)
           (bishop)
           (monsignor)
           (msgr)
           (priest)
           (priests)
           (deacon)
           (clergy))

```



```
:pronouns (*he*))  
  
(defactor bishop-romero  
:member-of clergyman  
:names ((monsignor romero)  
        (bishop romero)  
        (msgsr romero)  
        (oscar arnulfo romero))  
:pronouns (*he*))
```

Appendix B: Role Patterns

(in-package :cl-user)

```
;;;=====
;;;
;;; ROLE-PATTERNS.LISP
;;;
;;; Warren Sack
;;; 19 July 1994
;;;
;;;=====
```

```
;;; source role
(defrole source [ ?source ] [text] ) ;;; 1
(defrole source ?source has reported ) ;;; 1
(defrole source according to ?source ) ;;; 1
(defrole source ?source added that ) ;;; 1
(defrole source ?source reported that ) ;;; 1
(defrole source according to statements made by ?source ) ;;; 2
(defrole source ?source said ) ;;; 2
(defrole source [ ?source ] \ " ) ;;; 2
(defrole source ?source denied ) ;;; 2
(defrole source \ ( ?source \ ) -- [ ?_ ] [text] ) ;;; 4
(defrole source \ ( ?source \ ) -- [text] ) ;;; 5
(defrole source according to reports by ?source ) ;;; 6
(defrole source ?source stated that ) ;;; 13
(defrole source ?source has expressed ) ;;; 13
(defrole source ?source added that ) ;;; 13
(defrole source ?source urged ) ;;; 13
(defrole source ?source has announced that ) ;;; 14
(defrole source ?source announced ) ;;; 14
(defrole source \ , ?source said ) ;;; 14
(defrole source official sources ) ;;; 18
(defrole source [communique] [ ?source ] [text] ) ;;; 20
(defrole source [signed] ?source ) ;;; 20
(defrole source [communique] [ ?source ] ) ;;; 25
(defrole source [signed] ?source ) ;;; 25
(defrole source ?source reported that ) ;;; 26
(defrole source \ " ?source pointed out ) ;;; 28
(defrole source \ " ?source said ) ;;; 28
(defrole source ?indirect-source told ?source that ) ;;; 31
(defrole source ?source demanded that ) ;;; 33
(defrole source \ " ?source states ) ;;; 33
(defrole source ?source issued a communique ) ;;; 33
(defrole source ?source condemned ) ;;; 33
(defrole source ?source emphasized that ) ;;; 35
(defrole source ?source added ) ;;; 35
(defrole source ?source recalled that ) ;;; 35
(defrole source ?source stressed that ) ;;; 35
(defrole source ?source condemned ) ;;; 36
```

(defrole source ?source regrets and condemns) ;;; 36
 (defrole source ?source said today that) ;;; 37
 (defrole source ?source stressed that) ;;; 37
 (defrole source ?source also denied that) ;;; 39
 (defrole source ?source noted that) ;;; 38
 (defrole source ?source explained that) ;;; 39
 (defrole source \, ?source said) ;;; 39
 (defrole source ?source indicated that) ;;; 39

;;; indirect-source role
 (defrole indirect-source ?indirect-source is admitting) ;;; 4
 (defrole indirect-source ?indirect-source acknowledges) ;;; 4
 (defrole indirect-source ?indirect-source supports) ;;; 4
 (defrole indirect-source ?indirect-source lashes out at) ;;; 4
 (defrole indirect-source ?indirect-source has described) ;;; 27
 (defrole indirect-source ?indirect-source told ?source that) ;;; 31
 (defrole indirect-source ?indirect-source told ?source) ;;; 36
 (defrole indirect-source ?indirect-source shouted) ;;; 34
 (defrole indirect-source ?indirect-source implicated ?criminal) ;;; 38

;;; terrorist role
 (defrole terrorist terrorists of ?terrorist) ;;; 1
 (defrole terrorist terrorists) ;;; 1
 (defrole terrorist subversives) ;;; 1
 (defrole terrorist extremists) ;;; 1
 (defrole terrorist ?terrorist terrorist criminals) ;;; 2
 (defrole terrorist ?terrorist terrorists) ;;; 27
 (defrole terrorist ?terrorist are terrorists) ;;; 35
 (defrole terrorist the subversives) ;;; 35
 (defrole terrorist the terrorists) ;;; 35
 (defrole terrorist the seditious) ;;; 35
 (defrole terrorist the rebels) ;;; 35
 (defrole terrorist ?terrorist is responsible for terrorist attacks) ;;; 37
 (defrole terrorist ?terrorist has carried out bomb attacks) ;;; 37
 (defrole terrorist a terrorist)

;;; criminal role
 (defrole criminal kidnapped by ?criminal) ;;; 1
 (defrole criminal genocide carried out by ?criminal) ;;; 4
 (defrole criminal ?criminal murdered ?victim) ;;; 4
 (defrole criminal ?criminal commit genocide) ;;; 4
 (defrole criminal ?criminal participated in the crime) ;;; 14
 (defrole criminal ?criminal is responsible for this crime) ;;; 20
 (defrole criminal ?criminal accused of committing the massacre of ?victim) ;;; 22
 (defrole criminal ?judge decides to punish ?criminal) ;;; 22
 (defrole criminal ?criminal were responsible for the massacre) ;;; 22
 (defrole criminal assassination by ?criminal) ;;; 22
 (defrole criminal ?victim was abducted by ?criminal) ;;; 25
 (defrole criminal the culprits) ;;; 28
 (defrole criminal ?criminal must be punished) ;;; 28
 (defrole criminal the masterminds) ;;; 28
 (defrole criminal ?criminal massacred ?victim) ;;; 28
 (defrole criminal ?criminal killed ?victim) ;;; 28
 (defrole criminal ?criminal will be punished) ;;; 28

(defrole criminal ?criminal were the ones who murdered ?victim) ;; 29
 (defrole criminal the murderer was ?criminal) ;; 29
 (defrole criminal ?criminal murdered ?victim) ;; 29
 (defrole criminal ?criminal's crimes) ;; 29
 (defrole criminal ?criminal that murdered ?victim) ;; 29
 (defrole criminal ?criminal will be punished) ;; 32
 (defrole criminal there will be punishment for ?criminal) ;; 32
 (defrole criminal ?judge blame ?criminal for this assassination) ;; 32
 (defrole criminal ?criminal was involved in the massacre of ?victim) ;; 34
 (defrole criminal ?criminal was going to continue killing ?victim) ;; 34
 (defrole criminal the assassin of ?victim was ?criminal) ;; 34
 (defrole criminal ?criminal killed ?victim) ;; 34
 (defrole criminal the crime perpetrated by ?criminal) ;; 33
 (defrole criminal ?criminal responsible for this assassination) ;; 33
 (defrole criminal ?criminal participated in the crime) ;; 34
 (defrole criminal ?criminal \, a gang of assassins \,) ;; 34
 (defrole criminal ?victim was kidnapped by ?criminal) ;; 36
 (defrole criminal ?criminal is detestable) ;; 35
 (defrole criminal ?criminal killed ?victim) ;; 37
 (defrole criminal ?criminal has killed ?victim) ;; 37
 (defrole criminal ?indirect-source implicated ?criminal) ;; 38
 (defrole criminal ?criminal staged this assassination) ;; 38
 (defrole criminal ?criminal does go unpunished) ;; 38
 (defrole criminal the arrest of ?criminal) ;; 39

;; judge-role

(defrole judge ?judge decides to punish ?criminal) ;; 22
 (defrole judge ?judge blame ?criminal for this assassination) ;; 32

;; victim role

(defrole victim ?victim was killed) ;; 1
 (defrole victim ?victim were killed) ;; 1
 (defrole victim ?criminal murdered ?victim) ;; 4
 (defrole victim murder of ?victim) ;; 22
 (defrole victim ?criminal accused of committing the massacre of ?victim) ;; 22
 (defrole victim the disappearance of ?victim) ;; 25
 (defrole victim the kidnapping of ?victim) ;; 25
 (defrole victim ?victim was abducted by ?criminal) ;; 25
 (defrole victim ?victim disappeared) ;; 26
 (defrole victim ?victim was kidnapped) ;; 26
 (defrole victim ?criminal massacred ?victim) ;; 28
 (defrole victim ?criminal killed ?victim) ;; 28
 (defrole victim ?criminal were the ones who murdered ?victim) ;; 29
 (defrole victim ?victim 's murder) ;; 29
 (defrole victim ?criminal murdered ?victim) ;; 29
 (defrole victim ?criminal that murdered ?victim) ;; 29
 (defrole victim ?victim \, who was kidnapped \,) ;; 31
 (defrole victim ?victim was found dead) ;; 31
 (defrole victim ?victim's body was found) ;; 31
 (defrole victim the assassination of ?victim) ;; 32
 (defrole victim the assassination of ?victim) ;; 33
 (defrole victim ?victim were kidnapped) ;; 33
 (defrole victim ?victim bodies were found) ;; 33
 (defrole victim ?criminal was involved in the massacre of ?victim) ;; 34

(defrole victim ?criminal was going to continue killing ?victim) ;;; 34
(defrole victim the assassin of ?victim was ?criminal) ;;; 34
(defrole victim ?criminal killed ?victim) ;;; 34
(defrole victim ?victim was kidnapped) ;;; 36
(defrole victim the kidnapping and assassination of ?victim) ;;; 36
(defrole victim the body of ?victim was found) ;;; 36
(defrole victim ?victim was kidnapped by ?criminal) ;;; 36
(defrole victim ?criminal killed ?victim) ;;; 37
(defrole victim ?criminal has killed ?victim) ;;; 37
(defrole victim ?victim tried to escape) ;;; 37
(defrole victim ?victim was found dead) ;;; 38
(defrole victim the murder of ?victim) ;;; 39
(defrole victim ?victim 's assassination) ;;; 39
(defrole victim the assassination of ?victim) ;;; 39

;;; military role

(defrole military ?military repell attack) ;;; 1
(defrole military ?military comb the area) ;;; 1
(defrole military ?military will obey the orders) ;;; 2
(defrole military ?military gives orders) ;;; 2
(defrole military the ?military offensive) ;;; 4
(defrole military ?military prevented the occupation) ;;; 6
(defrole military ?military repelled the incursion) ;;; 6
(defrole military ?military bombed positions) ;;; 6
(defrole military ?military is carrying out operations) ;;; 6
(defrole military ?military has increased its surveillance) ;;; 6
(defrole military ?military gained control) ;;; 15
(defrole military ?military seized ?_ from ?enemy) ;;; 15
(defrole military ?military withdrew) ;;; 15
(defrole military ?military counteroffensive) ;;; 27
(defrole military ?military started its offensive) ;;; 29
(defrole military ?military invasion of ?geographic-area) ;;; 29

;;; geographic-area role

(defrole geographic-area ?military invasion of ?geographic-area) ;;; 29

;;; enemy role

(defrole enemy enemy) ;;; 5

Appendix C: Point of View Patterns

(in-package :cl-user)

```
;;;=====
;;;
;;; POV-PATTERNS.LISP
;;;
;;; Warren Sack
;;; 19 July 1994
;;;
;;;=====
```

```
(defpov government
  ((government source)
   (government military)
   (guerrillas terrorist)
   (guerrillas criminal)))
```

```
(defpov guerrillas
  ((government criminal)
   (guerrillas source)
   (guerrillas military)))
```

```
(defpov democratic-convergence
  ((democratic-convergence source)
   (democratic-convergence victim)))
```

```
(defpov catholic-church
  ((catholic-church source)
   (catholic-church victim)))
```

Appendix D: Ten News Stories from the MUC-3 Corpus

This appendix contains the text of the first 10 stories in the MUC-3 corpus of news stories from El Salvador, Guatemala, Honduras, Argentina, Bolivia, Chile, Columbia, Ecuador, and Peru. These news stories are transcriptions of radio and television broadcasts and were translated into English by the Foreign Broadcast Information Service. SpinDoctor's analysis of these ten stories can be seen in Appendix E.

[MUC3-0001]

San Salvador, 3 jan 90 -- [report] [armed forces press committee, Coprefa] [text] the Arce battalion command has reported that about 50 peasants of various ages have been kidnapped by terrorists of the Farabundo Marti national liberation front [Fmln] in San Miguel department. According to that garrison, the mass kidnapping took place on 30 December in San Luis De La Reina. The source added that the terrorists forced the individuals, who were taken to an unknown location, out of their residences, presumably to incorporate them against their will into clandestine groups.

Meanwhile, three Subversives were killed and seven others were wounded during clashes yesterday in Usulután and Morazan departments. The atonal battalion reported that one extremist was killed and five others were wounded during a clash yesterday afternoon near La Esperanza farm, Santa Elena jurisdiction, Usulután department.

It was also reported that a soldier was wounded and taken to the military hospital in this capital.

The same military unit reported that there was another clash that resulted in one dead terrorist and the seizure of various kinds of war materiel near San Rafael farm in the same town.

In the country's eastern region, military detachment no.4 reported that a terrorist was killed and two others were wounded during a clash in La Ranera stream, San Carlos, Morazan department. An m-16 rifle, cartridge clips, and ammunition were seized there.

Meanwhile, the 3d infantry brigade reported that Ponce battalion units found the decomposed body of a subversive in La Finca hill, San Miguel. An m-16 rifle, five grenades, and material for the production of explosives were found in the same place. The brigade, which is headquartered in San Miguel, added that the seizure was made yesterday morning.

National guard units guarding the Las Canas bridge, which is on the northern trunk highway in Apopa, this morning repelled a terrorist attack that resulted in no casualties. The armed clash involved mortar and rifle fire and lasted 30 minutes. Members of that security group are combing the area to determine the final outcome of the fighting.

[MUC3-0002]

San Salvador, 4 jan 90 -- [text] according to statements made by a salvadoran officer today, despite the change in the leadership of the salvadoran air force, the air force will continue its normal activities.

The salvadoran air force maintains the same operational level it has had for the last 10 years, despite the removal of its top commander, Jaun Rafael Bustillos, this week.

Paratrooper battalion commander major Rene Rodriguez Hurtado said today that the air force will continue to obey the orders of the general staff and to maintain a similar operational level.

[Rodriguez] \"we know general Villamariona [new air force commander] enough to know that we will continue with our battalion's attacks on all the terrorist criminals' redoubts. We depend directly on the general staff, since it is the strategic head of the entire army. The general staff gives us our orders to launch our operations.\"

Major Rodriguez Hurtado today denied that salvadoran air force planes had entered nicaraguan territory as the Sandinist government said last week. Maj Rodriguez Hurtado said that salvadoran air force planes have never entered the neighboring Teritory of Nicaragua.

[Rodriguez] \"i can tell you that those were neither salvadoran nor Honduran planes. In my opinion, the Sandinists want to invent an excuse to continue helping the Fmln [Farabundo Marti national liberation front]. Ours and the Honduran planes are small planes. I do not think that Honduran planes entered nicaraguan territory. Those communists are trying to invent an excuse to continue helping the Fmln terrorist criminals.\"

Maj. Rodriguez made these statements during the presentation of a number of weapons confiscated by the paratroopers battalion on 4 January.

[Rodriguez] \"here we have 12 rifles, 11 Ar-15 rifles and 1 Ak-47 rifle; 4,000 cartridges, of which 2,000 are for m-16 machine guns and 2,000 for Ak-47 rifles; 36 cartridges of an unknown caliber-- military intelligence has already been informed of this discovery; and i (icon) radio of the type used by the leaders of the terrorist groups.\"

The weapons were found at El Nispero creek near the Sacamil neighborhood. Among the confiscated weapons were cartridges of unknown make. This is Apparently the first time the army has found that type of cartridge. Maj Rodriguez Hurtado said that military intelligence will proceed to establish the origin of these cartridges and the weapons for which they are intended.

[MUC3-0003]

Santiago, 5 jan 90 -- [excerpts] the government junta yesterday approved a draft bill to \"modify law 18314, which defines terrorist activities and establishes penalties.\" The bill will restrict the application of the law to the most serious cases and will leave other crimes, that were formerly considered terrorist activity, for trial under common law.

The new bill accepts the possibility in some cases of a prisoner being released on bail or receiving a pardon or amnesty. The bill also establishes other benefits not contemplated by law 18314, including a new proceedings for speedier trials. [Passage omitted]

According to retired general Fernando Torres Silva, who is the army judge advocate general, the modifications establish lighter sentences, freedom on bail, and possible pardons or amnesties. These benefits were not contemplated under the former antiterrorist law.

As an example Torres cited that \"sentences of individuals tried as accomplices will be lower. They will receive lower sentences than that previously contemplated.\"

The new bill will also allow a great number of people who are being tried by the interior ministry to go out on bail.

According the general attorney's office, more than 1,000 prisoners might benefit from the modifications.

[MUC3-0004]

Clandestine, 8 jan 90 (radio Venceremos) -- [station commentary] [text] Goodbye [air force commander] general Rafael Bustillo, false hope for mass killers. Gen. Bustillo resigns or is made to resign as the air force chief. Because of this, the general is admitting that without the bombing on San Salvador's neighborhoods the Fmln [Farabundo Marti national liberation front] would have defeated the salvadoran armed forces. Bustillo's resignation is yet another sign of the failure of 10 years of the u.s. Counterinsurgency policy in El Salvador, as became clear with the offensive launched on 10 November.

For 10 years the military and their u.s. Sponsors based their trust in the air power to drown in blood the aspiration of the nation to achieve democracy and social justice. This first happened in the interior of the country, with the destruction of dozens of towns like Tenancingo, Berlin, Torola, San Fernando, and so many others that took place along with the genocide carried out by the infantry with the mass murders of civilians in El Mosoto, Sunpul, and Ocotapayo. Bustillo is leaving as so many generals and u.s. Ambassadors have. And like all of them his departure reveals the imperialist power's failure to consolidate its counterinsurgency program in El Salvador.

As all the others who have left, Bustillo also seeks excuses in the face of his failure, by accusing others. He accuses [former defense minister Eugenio] Vides casanova of being soft in fighting the people's plan. One would have to ask Bustillo if the air force, of which he represented its most encouraging pillar, was soft when it murdered 70,000 salvadorans, Monsignor Arnulfo Romero and the jesuit priests included, and even when it murdered patriot officers of that same army. Was the u.s. Pentagon soft when it gave \$3 billion, weapons, and bombs for the most cowardly act of genocide latin America can recall this century?

The u.s. Strategy, implemented through the top military command in the past 10 years with all its might did all it could so that the permanent rebellion in the rural areas reached the country's vital centers, as indeed happened as of the 10 November offensive. Bustillo states it clearly.

The bombing temporarily halted the defeat of the top command. Through an act of genocide, it managed to remove the disgruntled people from the neighborhoods where the guerrilla forces were. Bustillo acknowledges genocide is the sole way the armed forces can survive. Bustillo acknowledges the Fmln's strength and therefore urges the armed forces to be tougher and inflexible and rejects a political solution to the conflict. A political solution would express the consensus among the country's major sectors. Obviously the top command represents a decadent minority. Bustillo supports launching a fascist crusade against the people's wish for democracy. In this effort, he lashes out at the catholic church, all opposition sectors that report human rights violations, and all sectors that support democratizing the country. In conclusion, we say that Bustillo is not resigning of his own volition. The Fmln action eliminated Bustillo as it eliminated Garcia, Vides, and so many other military chiefs who left their posts without fulfilling the task commissioned by the united states: to stop the Fmln.

The Fmln offensive bared the genocidal, antipopular, and fascist nature of the Alfredo Cristiani regime to such an extent that it prompted a strong debate in the united states on cutting military aid. The jesuits' murders and the bombings so

shook the international community that it is pressuring for punishment of the officers involved in these incidents. However, the bush administration accepted this, since colonel Rene Emilio Ponce [chairman of the joints chiefs of staff] and his clique, even Cristiani, must shoulder this responsibility. We must not forget that in Costa Rica the salvadoran defense minister disclosed Cristiani authorized the bombings.

Bush needed a scapegoat to place it on the altar of the debates which will begin in the u.s. Congress on the fate of military aid. Bustillo is bush's scapegoat. However, Bustillo was already a useless element of the murderous machinery that remains active. Will the u.s. Public and congress be appeased with this foul tasting candy? Time and the spirit of the times are obviously running against the fascist program in El Salvador. While peoples of the world tear down dogmas, dictatorships, and walls, Bustillo, who leaves through the door of defeat, asks the armed forces to be tougher, commit more genocide, and build more walls of repression against the church, the people, and the nation as a whole. Goodbye, gen. Bustillo. We continue writing history without you. The decade we begin is of democracy, independence, and self-determination. That is why foreign intervention and the walls of intolerance have a short life span in this country, where the people grow in heroism and the struggle for social change is not deterred by bombings.

[MUC3-0005]

Clandestine, 8 jan 90 (radio Venceremos)-- [text] a war bulletin indicates that on 6 January at 1625, Fmln [Farabundo Marti national liberation front] troops clashed with the cavalry company in Finca Santa Elena, Santa Tecla, near San Salvador, killing three and wounding four enemy troops, including the patrol leader who was among those killed. Our troops seized an m-14 rifle from the enemy, 3,000 cartridges for a 7.62-mm rifle, five knapsacks, six grenades, and field equipment.

[MUC3-0006]

San Salvador, 9 jan 90 (Dpa) -- [text] the salvadoran army today prevented the occupation of cities in the eastern part of El Salvador, waging strong clashes between midnight and dawn, according to reports by military sources.

The attacks, according to the reports, took place in Usulután, Usulután department, 110 Km east of the capital, and in Santa Elena where as of 0700 clashes were still under way. Santa Elena is located 119 Km east of the capital, and both cities are coffee regions.

In Usulután, 6th brigade troops and rapid deployment battalions repelled the incursion with artillery fire. During the 2d week of December, there had been strong clashes on the outskirts of Usulután, in what the guerrillas called the \"battle of El Nisperal,\" in the canton that bears the same name, which cost the army 29 dead, although it has only admitted to 21 deaths.

In Santa Elena the clashes were also strong, but so far there have been no reports of casualties on either side.

From San Vicente, in the central part of the country, where the Chichontepec volcano is located, strong clashes were reported, and the 5th brigade bombed Rebel positions on the Colcano, where the Fmln has strong bastions.

The 1st infantry brigade is carrying out operations in northern San Salvador. And on the outskirts of the capital this morning powerful explosions, characterized as military actions to prevent Rebels from gathering, could be heard. The night before last, there were attempts to enter the capital from that direction.

The army has increased its surveillance to prevent mobilizations, and through careful searches vehicles are being controlled at the capital city's entrances and exits.

In other actions, the armed forces press committee reported at least six Fmln deaths in various parts of the country.

Last night there were fewer attacks on stores and electric towers in the capital. There were also threats of power rationing if these actions continued.

[MUC3-0007]

Santiago, 9 jan 90 -- [commentary] [Olga Klivadenko] [text] the deployment of u.s. Ships, including the aircraft carrier \"John f. Kennedy\" and the corvette \"Virginia\" along the colombian coast yesterday, seeking to intercept drug trafficking routes, could be the result of three policies:

First, that the united states needs to justify the invasion of panama under the excuse of its fight against drug trafficking.

Second, that both the deployment of ships and the invasion of Panama are part of a serious military plan to end the drug plague.

Third, that the united states must somehow justify the invasion of Panama and the need to recover its international image has provided a real opportunity to put an end to drug trafficking.

If the objective really is to eliminate drugs, then the battle is praiseworthy, But--like in Panama--the procedure is painful, overbearing, imperialistic, and humiliating. Instead of winning the latin american countries' appreciation and solidarity with this laudable cause, the united states is again provoking angry reactions and deep distrust. The side effects of this action could be very negative for the struggle against drug trafficking, in which all the countries of this continent should be cooperating. Alternatively, would it be better if they did not cooperate in this phase?

Could it be that what we have called imperialistic clumsiness, insensibility, and abuse, is simply a deliberate strategy with which, perhaps for well founded reasons and even at the risk of diplomatic conflicts, the united states seeks to isolate latin America on purpose and thus ensure the success of its operation? If this were the case then we must ask, why?

From its operations in Panama and Colombia, the united states is telling us that it is not concerned with the diplomatic aspect for the time being. The united states wants action and efficiency. The united states is losing its patience, and perhaps its confidence in the seriousness and capability of latin America to face this war.

If the action in Colombia goes further than a mere surveillance, the united states will be establishing a precedent in this battle against drug trafficking. This means that during a first phase the united states will carry out a military operation by itself. During a second phase it will discuss secondary subjects but with less urgency in the war against drug trafficking.

Next 15 February, bush will meet in Cartagena De Indias, Colombia, with his colombian, peruvian, and bolivian counterparts. Perhaps on that occasion they will talk about the other aspects of this war: the reorganization of the local economies to end the dependency on coca plantations, the recovery of drug addicts, and the consolidation of governments, etc.

During this current phase, however, the united states wishes to act alone. Who knows if it is because of latin America's lack of understanding? As we said a few days ago, the drug trafficking problem is creating new situations in international relations. The u.s. Proposal is probably something like: could a world power like the united states continue to ignore a serious problem like the production,

distribution, and consumption of drugs, merely to avoid harming latin american sensibilities, sensibilities of countries whose governments are always muddled up in useless diplomatic and bureaucratic procedures and fruitless negotiations? And all this while a large part of the developed world's population is becoming addicted to a deadly product?

The united states could also be asking itself: have the latin american governments been able to solve this serious problem by themselves? Can those governments stop drug trafficking when it is known that drugs and corruption go hand in hand; when in the past and perhaps in the present, high ranking government officials have been linked with this illegal trade and protecting the Medellin cartel chiefs? Why, fOr example, after all this hasN't the colombian government managed to extradite the main chiefs of that cartel to the united states?

Let us recall the horrible explosion of the colombian airliner, in which more than 100 people died; the bomb explosion in the secret police headquarters in Bogota; the murder of many journalists who were engaged in an heroic war against drug trafficking; and mainly, let us recall the ever-present difficulties that must be faced when the Medellin cartel chiefs must be punished and extradited.

[MUC3-0008]

Bogota, 9 jan 90 (Efe) -- [text] Ricardo Alfonso Castellar, mayor of Achi, in the northern department of bolivar, who was kidnapped on 5 January, apparently by army of national liberation (Eln) guerrillas, was found dead today, according to authorities.

Castellar was kidnapped on 5 January on the outskirts of Achi, about 850 Km north of Bogota, by a group of armed men, who forced him to accompany them to an Undisclosed location.

Police sources in Cartagena reported that Castellar's body showed signs of torture and several bullet wounds.

Castellar was kidnapped by Eln guerrillas while he was traveling in a boat down the Cauca river to the Tenche area, a region within his jurisdiction.

In Cartagena it was reported that Castellar faced a \"revolutionary trial\" by the Eln and that he was found guilty and executed.

Castellar is the second mayor that has been murdered in Colombia in the last 3 days.

On 5 January, Carlos Julio Torrado, mayor of Abrego in the northeastern department of Santander, was killed apparently by another Guerilla column, also belonging to the Eln.

Torrado's son, William; Gustavo Jacome Quintero, the departmental government secretary; and bodyguard Jairo Ortega, were also killed.

The group was traveling in a 4-wheel drive vehicle between Cucuta and the rural area known as Campanario when their vehicle was blown up by four explosive charges that detonated on the highway.

[MUC3-0009]

Medellin, 9 jan 90 -- [text] the national army was involved in new clashes within the past few hours with Farc [revolutionary armed forces of Colombia] and the Eln [army of national liberation] in the rural area of Yondo in Antioquia's middle Magdalena region.

Spokesmen for the 5th brigade reported 5 soldiers killed, several wounded, and at least 12 Rebel casualties.

The exodus of peasants from that region continued to increase today according to official reports. The peasants walked away from their properties to avoid becoming victims of the clashes.

[MUC3-0010]

Lima, 9 jan 90 (television Peruana) -- [text] president Alan Garcia has confirmed that the u.s. Troops that surrounded our diplomatic mission in Panama were withdrawn after a former minister surnamed Cordova left the embassy. The military man, who is involved in a murder trial, left the peruvian embassy as soon as he learned that the embassy considered him persona Non Grata.

[Begin recording] [Garcia] i think journalists have magnified the issue a bit. As far as we know, people sought protection....[changes thought] someone arrives at the door of the embassy and asks for refuge in view of a situation of force. The situation in Panama is abnormal; it has been totally invaded by 30,000 soldiers.

Some persons entered the embassy. At least one of them had been allegedly linked with the death of a panamanian minister, a mr Spadafora, who was murdered. The authorities of the embassy and of the foreign ministry stated that this person's presence was undesirable. We do not refuse to grant Refuge--a condition that is granted before granting political Asylum--to whomever feels harassed or mistreated. In a case like this, however, in which there was an alleged involvement in a murder which was committed years ago, this Person--i believe he is a man surnamed Cordova, apparently a Highranking panamanian guard Officer--was told that his presence was not wanted. He left the embassy of his own will.

There are other panamanian defense forces officers at the embassy, including mr Noriega's secretary, but their status should be established by Peru.

An incident was reported yesterday, as the u.s. Troops surrounded the embassy, deploying barriers and barbed wire.

There was an attempt to depict a dramatic and spectacular picture of a siege or perhaps of an invasion of the embassy.

Fortunately, the foreign ministry resorted to diplomatic channels and the troops were soon withdrawn. [End recording]

Appendix E: Ten of SpinDoctor's Actor-Role Analyses

This appendix contains summaries of the SpinDoctor system's analyses of the first ten stories in the MUC-3 corpus of news stories. All of the summaries are printed in a format which resembles this one:

```
*****
analysis of the story [MUC3-0015]
*****

-----
GROUP = RADIO-VENCEREMOS
actor: Venceremos           plays the role : ?source
      and has been coreferenced to
actor: Our people           plays the role : ?military
actor: our troops           plays the role : ?military

-----
GROUP = ARMY
actor: five national guard soldiers  plays the role: ?victim

-----
UNGROUPED
actor: enemy                 plays the role: ?enemy

-----
weighted high-level actor-role pairs:
(guerrillas ?source 1)
(guerrillas ?military 2)
(government ?victim 1)
(unknown ?enemy 1)
=====
POVs = (guerrillas)
=====
```

However, in the summaries which follow, the "actor:," "plays the role:," "and has been coreferenced to," and "weighted high-level actor-role plans," comments have been omitted. The reader who is having difficulty interpreting the following results should consult the chapter which describes the routines analyze-story, find-actor-role-bindings, coreference-actors, construct-high-level-weighted-actor-role-pairs, and identify-point-of-view.

These ten analyses are the first tenth of a test in which SpinDoctor was run on 100 stories. The version which was used on the 100 stories took approximately 10 to 30 seconds to analyze each

story. The system used the set of roles, actors, and points of view which can be found in Appendices A, B, and C. The text of the ten stories analyzed can be found in Appendix D.

Micronet 2050:MUC-3:Text:0001

GROUP = ARMY
the Arce battalion command = SOURCE =
that garrison = SOURCE
The atonal battalion = SOURCE
The source = SOURCE
The same military unit = SOURCE
military detachment no.4 = SOURCE
the 3d infantry brigade = SOURCE
The source = SOURCE

GROUP = FMLN
the Farabundo Marti national liberation front = TERRORIST =
terrorists = TERRORIST
the terrorists = TERRORIST
terrorists = CRIMINAL
terrorists = TERRORIST
subversives = TERRORIST
a terrorist = TERRORIST
three subversives = VICTIM
a terrorist = TERRORIST
a terrorist = TERRORIST
a terrorist = VICTIM
that one extremist = VICTIM

(GOVERNMENT SOURCE 8)
(GUERRILLAS CRIMINAL 1)
(GUERRILLAS TERRORIST 8)
(GUERRILLAS VICTIM 3)

=====
POVs = (GOVERNMENT)
=====

Micronet 2050:MUC-3:Text:0002

GROUP = GOVERNMENT
statements = SOURCE =
 a salvadoran officer today = SOURCE
 Paratrooper battalion commander major Rene Rodriguez
Hurtado = SOURCE
 Paratrooper battalion commander major Rene Rodriguez
Hurtado = SOURCE
 Major Rodriguez Hurtado today = SOURCE
 Rodriguez = SOURCE
 the Sandinist government = SOURCE
 Maj Rodriguez Hurtado = SOURCE
 Maj Rodriguez Hurtado = SOURCE
 Rodriguez = SOURCE
 Rodriguez = SOURCE

GROUP = FMLN
the terrorist = TERRORIST =
 the Fmln = TERRORIST

(GOVERNMENT SOURCE 11)
(GUERRILLAS TERRORIST 2)

=====
POVs = (GOVERNMENT)
=====

Micronet 2050:MUC-3:Text:0003

UNGROUPED
a terrorist = TERRORIST

(UNKNOWN TERRORIST 1)

=====
POVs = (GOVERNMENT GUERRILLAS DEMOCRATIC-
CONVERGENCE CATHOLIC-CHURCH)
=====

Micronet 2050:MUC-3:Text:0004

GROUP = RADIO-VENCEREMOS
Venceremos = SOURCE =
 station commentary = SOURCE
 station commentary = _

UNGROUPED
the general = INDIRECT-SOURCE =
 Bustillo = INDIRECT-SOURCE
 Bustillo = INDIRECT-SOURCE
 Bustillo = INDIRECT-SOURCE
 he = INDIRECT-SOURCE

GROUP = ARMY
the infantry = CRIMINAL =
 it = CRIMINAL
 patriot officers = VICTIM
 it = CRIMINAL
 it = CRIMINAL
 it = CRIMINAL
 patriot officers = VICTIM

GROUP = CIVILIANS
civilians = VICTIM =
 70,000 salvadorans = VICTIM
 70,000 salvadorans = VICTIM

(GUERRILLAS SOURCE 2)
(GUERRILLAS _ 1)
(UNKNOWN INDIRECT-SOURCE 5)
(GOVERNMENT CRIMINAL 5)
(GOVERNMENT VICTIM 2)
(CIVILIANS VICTIM 3)
=====

POVs = (GUERRILLAS)
=====

Micronet 2050:MUC-3:Text:0005

GROUP = RADIO-VENCEREMOS
Venceremos = SOURCE =
Our troops = MILITARY

UNGROUPED
enemy = ENEMY =
the enemy = ENEMY
enemy = ENEMY

UNGROUPED
an m-14 rifle = _

(GUERRILLAS SOURCE 1)
(GUERRILLAS MILITARY 1)
(UNKNOWN ENEMY 3)
(UNKNOWN _ 1)

=====
POVs = (GUERRILLAS)
=====

Micronet 2050:MUC-3:Text:0006

GROUP = ARMY
the salvadoran army today = MILITARY =
 The army = MILITARY
 The 1st infantry brigade = MILITARY
 rapid deployment battalions = MILITARY
 military sources = SOURCE

UNGROUPED
reports = SOURCE =
 the reports = SOURCE

(GOVERNMENT MILITARY 4)
(GOVERNMENT SOURCE 1)
(UNKNOWN SOURCE 2)

=====
POVs = (GOVERNMENT)
=====

Micronet 2050:MUC-3:Text:0007

UNGROUPED
Olga Klivadenko = SOURCE =
we = SOURCE

UNGROUPED
the invasion = MILITARY =
the invasion = MILITARY
the invasion = MILITARY
the invasion = MILITARY
the invasion = MILITARY
the invasion = MILITARY

UNGROUPED
panama = GEOGRAPHIC-AREA =
panama = GEOGRAPHIC-AREA
Panama = GEOGRAPHIC-AREA
Panama = GEOGRAPHIC-AREA
Panama = GEOGRAPHIC-AREA
Panama = GEOGRAPHIC-AREA

UNGROUPED
many journalists = VICTIM =
many journalists = VICTIM

UNGROUPED
the Medellin cartel chiefs = CRIMINAL

(UNKNOWN SOURCE 2)
(UNKNOWN MILITARY 6)
(UNKNOWN GEOGRAPHIC-AREA 6)
(UNKNOWN VICTIM 2)
(UNKNOWN CRIMINAL 1)

=====
POVs = (GOVERNMENT GUERRILLAS DEMOCRATIC-
CONVERGENCE CATHOLIC-CHURCH)
=====

Micronet 2050:MUC-3:Text:0008

UNGROUPED
authorities = SOURCE =
 Cartagena = SOURCE

UNGROUPED
Castellar = VICTIM =
 Castellar = VICTIM
 Castellar = VICTIM
 Castellar = VICTIM

GROUP = GUERRILLAS
Eln guerrillas = CRIMINAL =
 Eln guerrillas = CRIMINAL
 Eln guerrillas = CRIMINAL

(UNKNOWN SOURCE 2)
(UNKNOWN VICTIM 4)
(GUERRILLAS CRIMINAL 3)

=====
POVs = (GOVERNMENT)
=====

Micronet 2050:MUC-3:Text:0009

UNGROUPED
official reports = SOURCE

(UNKNOWN SOURCE 1)

=====
POVs = (GOVERNMENT GUERRILLAS DEMOCRATIC-
CONVERGENCE CATHOLIC-CHURCH)
=====

Micronet 2050:MUC-3:Text:0010

GROUP = MINISTRY
television Peruana = SOURCE =
the foreign ministry = SOURCE

UNGROUPED
an invasion = MILITARY =
an invasion = MILITARY

UNGROUPED
the embassy = GEOGRAPHIC-AREA =
the embassy = GEOGRAPHIC-AREA

(GOVERNMENT SOURCE 2)
(UNKNOWN MILITARY 2)
(UNKNOWN GEOGRAPHIC-AREA 2)

=====
POVs = (GOVERNMENT)
=====