

**Recent Titles in
Progress in Communication Sciences**

George A. Barnett, Series Editor

Progress in Communication Sciences, Volume 5
Brenda Dervin

Progress in Communication Sciences, Volume 6
Brenda Dervin

Progress in Communication Sciences, Volume 7
Brenda Dervin

Progress in Communication Sciences, Volume 8
Brenda Dervin

Progress in Communication Sciences, Volume 9
Brenda Dervin

Progress in Communication Sciences, Volume 10
Brenda Dervin

Progress in Communication Sciences, Volume 11
Brenda Dervin

Progress in Communication Sciences, Volume 12
William D. Richards and George Barnett

Progress in Communication Sciences, Volume 13
Franklin J. Boster and George Barnett

Progress in Communication Sciences, Volume 14: *Mutual Influence in
Interpersonal Communication*
Mark T. Palmer and George A. Barnett

Progress in Communication Sciences, Volume 15: *Advances in Telecommunications*
Harmeet Sawhney and George A. Barnett, editors

Theory, Method, and Practice in Computer Content Analysis
Mark D. West, editor

Applications of Computer Content Analysis

Edited by
Mark D. West

Progress in Communication Sciences, Volume 17
George A. Barnett, Series Editor



ABLEX PUBLISHING
Westport, Connecticut • London

Library of Congress Cataloging-in-Publication Data

Applications of computer content analysis / edited by Mark D. West.
p. cm. -- (Progress in communication sciences, ISSN 0163-5689 ; v. 17)
Includes bibliographical references and index.
ISBN 1-56750-504-X (alk. paper) -- ISBN 1-56750-505-8 (pbk. : alk. paper)
1. Content analysis (Communication)--Data processing. I. West, Mark D. II. Series.
P93.A67 2001
302.2--dc21 2001022157

British Library Cataloguing in Publication Data is available.

Copyright © 2001 by Mark D. West

All rights reserved. No portion of this book may be reproduced, by any process or technique, without the express written consent of the publisher.

Library of Congress Catalog Card Number: 2001022157
ISBN: 1-56750-504-X
1-56750-505-8 (pbk.)
ISSN: 0163-5689

First published in 2001

Ablex Publishing, 88 Post Road West, Westport, CT 06881
An imprint of Greenwood Publishing Group, Inc.
www.ablexbooks.com

Printed in the United States of America



The paper used in this book complies with the
Permanent Paper Standard issued by the National
Information Standards Organization (Z39.48-1984).

10 9 8 7 6 5 4 3 2 1

Contents

1. Applied Text Theory: Quantitative Analysis of Answers to Open-Ended Questions
Peter Ph. Mohler and Cornelia Zuell 1
2. Building WorldView(s) with Profiler+
Michael D. Young 17
3. Intersubjective Semantic Meanings Emergent in a Work Group: A Neural Network Content Analysis of Voice Mail
John C. Sherblom, N. L. Reinsch, Jr., and Raymond W. Beswick 33
4. Using the Computer to Identify Unknown Authors
James W. Tankard, Jr. 51
5. Using Neural Networks to Assess Corporate Image
Joseph G. T. Salisbury 65
6. Linking Gender Language in News about Presidential Candidates to Gender Gaps in Polls: A Time-Series Analysis of the 1996 Campaign
James A. Danowski and Rebecca Ann Lind 87
7. Semi-Automated Content Analysis of Pharmacist-Patient Interactions Using the Theme Machine Document-Clustering System
Bruce L. Lambert 103
8. Monitoring the Social Environment Using Computer Content Analysis of Online News Media Text: An Example in Natural Resources
David N. Bengston and David P. Fan 123

Linking Gender Language in News about Presidential Candidates to Gender Gaps in Polls: A Time-Series Analysis of the 1996 Campaign

James A. Danowski and Rebecca Ann Lind

INTRODUCTION

Chapter Goals

The goals of this chapter are twofold. Because this book attempts to illustrate different approaches to computerized content analysis, our primary goal is to show how one set of statistical content analysis tools can address communication research questions. As seen in other chapters, there are numerous methods for computer-based content analysis. We guide the reader on one journey through this wilderness, marking our trail along the way to make it easier for the reader to later retrace the steps, or to branch off onto different paths with more certainty as to where they may lead.

Our second goal is more substantive. We wish to determine whether changes in gender-encoded language in newspaper accounts of the 1996 U.S. presidential election campaign predict gender gaps in the electorate's candidate preferences as observed in tracking polls, the "horse race" yardstick of campaigns.

The concept of a gender gap in political campaigns has a long history. If one examines the vote breakdowns for candidates by gender of voter, one sees

that ever since the election of 1936 between Roosevelt and Landon, each presidential election has shown that candidate preferences differ by gender.

Our substantive focus on gender gaps in voter preferences for candidates contributes to the value of our illustration of a statistical content analysis. This is because the gender gap during the 1996 campaign received considerable media attention, which may have contributed to modifications by the candidates of their campaign messages to either capitalize on or attempt to reduce the gap. To the extent this happened, there could be sufficient variation in gender-related newspaper content to enable statistical content analysis to detect it, even though the language changes may be too subtle for manual content analysis to recognize. More generally, the analysis of full-text news stories is a good ground on which to examine the strengths and weaknesses of statistical content analysis. Such messages are voluminous, ubiquitous, and increasingly available from machine-readable sources.

Methodological and Theoretical Significance

Normally, the methodological aspects of a study are subordinate to the theoretical and conceptual aspects. In a book such as this, however, the reverse is the case. Accordingly, first we deal with the methodological significance of this research. There are several reasons why this research is a useful demonstration of some aspects of statistical content analysis. First, the content extracted from the newspaper and wire service coverage of a campaign creates a large corpus of text, which would be difficult to analyze manually. In this research we census rather than sample available campaign messages, resulting in a corpus of 50 megabytes of text. Second, we conduct the analysis over time. This demonstrates how content analysis can move beyond treatment of a body of text as an aggregate collection of messages, the approach common to traditional manual content analysis. We perform a time-series content analysis on 48 weekly subsets of the corpus. With each of the weekly-content files averaging a megabyte in size, it would be difficult to do an analysis of the full text of each file without the aid of computer software. Third, our substantive focus gives us a standard criterion variable that is not part of the content itself, the Gallup poll results. These "side data" allow us to examine the predictive validity of statistical content analysis variables.

In addition to the methodological significance of this study, there are several theoretical reasons why this research venture is significant. One is that critical, cultural, and media studies theories have reasoned that a focus on the natural language of communicants is essential (Habermas, 1979). Language is fundamental to processes of mediation of the modes of information production, reproduction, and distribution. There are few examples of social communication messages that do not contain words, even though graphics may dominate.

Second, in political communication research there has been insufficient attention to the language of news in the mix of other campaign variables. This is particularly significant because critical theories of media and language have been explicated to apply to the public sphere rather than to private spheres (Habermas, 1979). Political communication, as conceptualized and studied in the social sciences, occurs primarily in the public sphere—the sphere of interpersonal and mediated communication—and only secondarily within the interior spheres of individuals.

Third, even though the bulk of the electorate reports television as a primary source of campaign news (Zhao & Chaffee, 1995), we know from other research (Danielian & Reese, 1989) that print media often set the agenda for television, and hence it is important to study print media. Included in our conception of print media are the wire services, which provide mainly textual news content to all news media, newspapers and television included. Accordingly, our study analyzes wire service news and newspaper content.

Fourth, we assume that one of the main goals of candidates for elected office is to communicate messages to the voters. It is therefore important to study the extent to which these messages connect with women's and men's interests, to the extent they differ. The literature review later in this chapter supports the assumption that they do.

Fifth, the domain we have chosen within which to illustrate different approaches to statistical content analysis is one of gender. Traditionally, gender has been of interest as a demographic variable for mundane marketing and audience-rating purposes. More academically interesting, gender has been propelled into a more abstract domain by the emergence of feminist theories and study. At a minimum, these call for the study of gender in relation to other social processes.

Most of the research addressing gender and politics looks at gender-based perceptions of candidates or issues, and does not systematically investigate the extent to which candidates' messages or messages about the candidates and their campaigns are gendered. A recent exception is the work of Gidengil and Everitt (1999), which employed traditional content analysis. Because we do content analysis in a precise automated way, this research is important. It has further value because it looks for a systematic relationship between news story content and voters' opinions. This moves us beyond both the "mere exposure" and the "content only" focus of much prior political communication research.

Theory

Our theoretical perspective assumes that networks of individuals exchange messages that shape others' word-usage patterns. Some regularity in language patterns is due to the system-wide effects of symbolic control of the media and interpersonal communication by organizational agents (Danowski,

1993a), while some also results from microlevel sociolinguistic factors in the local interpersonal networks of actors (Danowski, 1986). (We assume some interactivity between the two independent variables.) More interlocking interpersonal networks foster more uniformity in language, fewer abstract words, and greater use of the present tense. The content of messages produced by media and expressed by actors in interpersonal situations is dynamically shaped by previous message content and by structural and processual factors emergent in interpersonal communication networks. Much language is reproduced, but over time there is sufficient innovation in language use to foster dynamic content processes. Zipf (1949) explicated these processes in terms of conservation-of-energy principles, describing them as following a "path of least resistance."

Our particular approach to message content is focused on the network of relationships among words, based on their co-occurrences (Danowski, 1982, 1993b). We map word co-occurrences in a sliding window of text, then partition the stream of word pairs into a network of words. In this way we map the natural language produced and reproduced in social networks, organizations, or institutions. Accordingly, our approach is consistent with critical theories and cultural studies, as well as social scientific approaches. Because we focus on patterns of relationships among words, this research is fundamentally qualitative. Nevertheless, we link the qualitative analysis of gender-encoded language patterns in text to quantitative gender gaps in voter preferences.

Political Gender Gap Research

This research examines a specific sort of gender gap. As we define the kind of gap we investigate, we should note, as does Conover (1996), that "gender gap" is a catch-all phrase for a number of different phenomena in and outside of politics (Wiris, 1986). In the political domain, some observers use the term to refer to the distribution of male and female candidates for public office. Here, our concern is with differences of opinion between voters, rather than between candidates, by gender. Furthermore, there are two kinds of voter-oriented gender gaps: electoral and partisan. Electoral gaps refer to the different voting choices of men and women, while the partisan gap refers to the different party choices of voters, with women being more likely than men to support Democratic candidates (Wiris, 1986).

Gender gaps in voter preferences for candidates may be based on a combination of factors. These include women's greater attention to certain issues (Conover, 1996; Gidengil, 1995; Montoya, 1996; Shapiro & Mahajan, 1986; Pratto, Stallworth, & Sidanius, 1997; Togeby, 1994), campaign tactics (Kern & Just, 1995; West, 1993), the greater likelihood of women's party affiliations being Democratic (Pratto, Stallworth, & Sidanius, 1997; Rozell & Wilcox, 1998; Wilcox, Hewitt, & Allsop, 1996), differences in socioeco-

nomie status (Clark & Clark, 1986), and socialization patterns (Trevor, 1999). The literature, however, reveals no candidate traits associated with gender gaps. Although informal discourse may point to a candidate's good looks, charm, or charisma and attribute differential gender responses to such variables, there is no evidence to support it (Newman & Sheth, 1984).

Our research questions concern links between gender-encoded news language and gender gaps in voter preferences for candidates. The 1996 U.S. presidential election campaign is the context in which we examine the following questions:

1. To what extent do news stories about the major candidates in the 1996 presidential election have gender-specific language?
2. Is there a relationship between news stories about candidates' gender references and the extent of male and female support for the candidates?

METHOD

The study uses a methodology of word n-gram statistical content analysis, analyzed using Danowski's WORDij tool set, and the WORDLINK, WORDGRAB, and LINKLIST programs (Danowski, 1993b). These word data are analyzed in conjunction with Gallup tracking polls that show the percentages of the respondents who prefer each candidate.

Network analysis is a method for representing the structure of relations among social units (Barnett, Danowski, & Richards, 1993). The social units, called "nodes," can be people, departments, organizations, nations, or message elements, such as words (Danowski, 1993b). Links are the attributes defining the relations among nodes.

In this research we conceptually define as the basic network link the co-occurrence of words within three word positions on either side of a word. Aggregated across news stories, the greater the co-occurrence the higher the link strength. Nodes are the words in the news stories. We analyze all words and do not exclude grammatical function words or any other words, unlike the practice in natural language database searching, in which "stop lists" of approximately 550 words are typically dropped. Also, unlike text processing for database searching, no stemming is performed to remove lexical variations. These kinds of data reductions in earlier research have been found to obscure important patterns in the text (Danowski, 1993b). Since our concern is not with the global context for the gender content, we do not need tools here for mapping the overall content network, although such methods have proven fruitful in previous research (Barnett & Danowski, 1992; Claffey, 1996; Danowski, 1982, 1988, 1993b; Danowski & Choi, 1999; Freeman, & Barnett, 1994; Jang & Barnett, 1994; Lind & Danowski, 1996; Rice & Danowski, 1993).

Corpus

To gather content for this study, we used a CD-ROM news source called NewsBank, which provides full texts of news stories from 500 newspapers in the United States. A description of NewsBank can be found at <http://www.newsbank.com>. This research investigates the universe of print media messages captured by NewsBank, not a sample. We searched the database and extracted the full texts of all stories containing the words "Clinton" or "Dole" starting with the week of the election, November 5, 1996 and working backward in time. We found that the beginning campaign-coverage boundary, judging from the frequency of stories, was the first week of December, 1995. There were a total of 1,756 stories available from within the content domain that contained either or both of the search terms.

We grouped news stories into weekly files, resulting in 48 files during the campaign period studied. The choice of a weekly interval for segmentation of the corpus, rather than some other larger or smaller time interval, was made because of the nature of the poll data we used as a dependent variable. Gallup polls during the campaign are typically conducted over a period of several days and are reported primarily on a week-by-week basis. Aligning the text segmentation with the poll frequency to the extent possible allows for greater internal validity. Figure 6.1 shows the total number of words per week from the news stories about the candidates. Along with word pairs and counts, individual words are output from the WORDLINK program.

The poll data were obtained directly from the Gallup organization. These data included a variety of variables used to break down the overall samples.

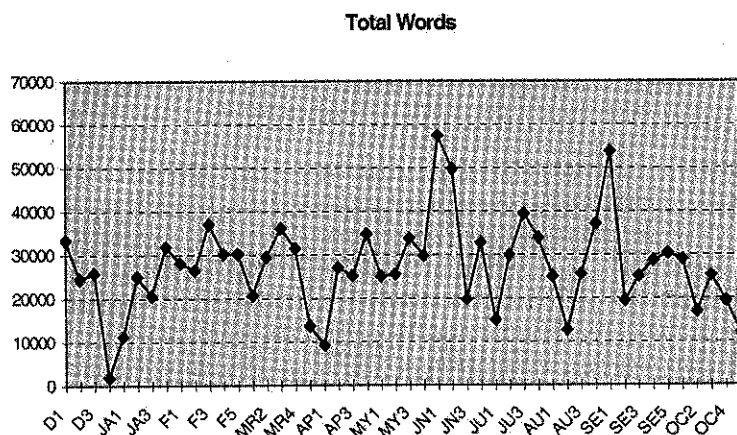


Figure 6.1 Total Words per Week in Candidate News Stories

For our purposes, the breakdown of respondent preferences for candidates by gender was the only portion of the poll data of interest in this research.

Analysis

Context in word networks can be viewed at node-centric and more global levels. For example, in our study the most global context would be revealed by the semantic features of the words in a network structure including all the word pairs extracted from a corpus. To begin getting acquainted with the data, we produced a global context of content, conducting a network analysis of all the word pairs from all the stories and represented these as one structure across the whole campaign. First, we extracted all word pairs occurring three words on either side of each word in the texts. Then we input these word pairs into three categorization techniques: network analysis, cluster analysis, and correspondence analysis. Groups of words represent different content themes.

We then constructed a list of gender words and isolated every instance in which a gender word appeared within three words of each of our focus words, "Clinton" and "Dole." To identify gender word strings we first generated a dataset of all pairs of words occurring within three word positions on either side of each word in the weekly content files. Next, to generate the gender word list to be used to select word pairs from the total pool output by the WORDLINK analysis of all text, we each identified as many such words as possible. After several cycles through this process we arrived at the final list. We operationalized gender-related words as including such male words as "him," "his," "he," "men," "Mr.," "man," "male," "males," "boy," "boys," "gentlemen," "gentleman," "masculine," "masculinity," "manly," "husband," "fathers," etc., and such female words as "hers," "her," "she," "women," "woman," "Mrs.," "girl," "girls," "lady," "ladies," "female," "females," "wife," "feminine," "femininity," "feminist," "feminists," "mothers," etc.

We then searched the data files for instances in which words we had classified as signifying gender and our focus words co-occurred. These data were then searched for all word strings of length three connecting to each word in the target word list, in this case containing the two words "Clinton" and "Dole." Strings of lengths greater than two involve indirect linkage to the focal word. Our definition and operationalization of indirect word links in this research includes two forms. One, a string of three words linking a gender word to a candidate may occur in a single story. Or two, it can occur through transitive association across stories. For example, word pair *a-b* occurs in one story, while *b-c* occurs in another, so *a-c* has a link by transitive association.

Word Triplets

To quantify indirect gender linkages to the candidates, we searched for all the word triplets where a gender word had an indirect link to the candidate name. The software works by accepting a target list of words, here our two candidate names and all the gender words, and then outputs all chains that are word triplets as well as the word pairs directly linking the candidate words to the gender words.

For each week we tabulated the number of male pairs and triplets and female pairs and triplets for each of the candidates. These counts we then used in a time-series analysis using the Gallup poll candidate preference data broken by gender. The Gallup data were conducted mostly on a biweekly basis from the second week of March through the second week of June, and then on a weekly basis from the third week of June through the first week of November. The data appear in Table 6.1 along with the SPSS commands used to conduct the analysis.

RESULTS

1. To what extent do news stories about the major candidates in the 1996 presidential election have gender-specific language?

We computed statistics across the time periods for the candidates. Dole had a total of 40 masculine direct word links and 121 indirect ones. Clinton had a total of 37 masculine direct word links and 17 indirect ones. So the two were nearly the same in masculine gender links over the course of the campaign. In feminine word links, Dole had a total of 9 direct links throughout the campaign and 83 indirect links. On the other hand, Clinton had 40 direct feminine word links and 81 indirect links. So, while the two candidates were similar on indirect links, Clinton had more than four times as many direct feminine word links than did Dole.

2. Is there a relationship between news stories about candidates' gender references and the extent of male and female support for the candidates?

To address this question we conducted a time-series analysis of the relationship between the direct and indirect masculine and feminine gender links for each candidate and the gaps in their support between males and females as seen in the Gallup poll data.

Time-Series Analysis

We used the ARIMA time-series model in SPSS. First we determined the parameters of the model, consistent with the procedures described by Ostrom (1990). A lag of 2 was found optimal, resulting in parameters of (2,1,0). After the lags were added to the model, the gender word pair and triplet counts

Table 6.1

Time-Series Data on One- and Two-Step Gender Links and Gallup Male/Female Candidate Preferences

```

data list/ time 1-2 week 5-7(A) MD1 14-15 WD1 19-20 MC1 24-25
          WC1 29-30 MD2 34-35 WD2 39-40 MC2 44-45 WC2 49-50
          galmend 54-55 galmenc 59-60 galwomd 64-65 galwomc 69-70
Variable Labels
time 'consecutive time period since first week of Dec.'
week 'month and week of data'
MD1 'number of direct links of masculine words to Dole'
WD1 'number of direct links of feminine words to Dole'
MC1 'number of direct links of masculine words to Clinton'
WC1 'number of direct links of feminine words to Clinton'
MD2 'number of two-step masculine word links to Dole'
WD2 'number of two-step feminine word links to Dole'
MC2 'number of two-step masculine word links to Clinton'
WC2 'number of two-step feminine word links to Clinton'
galmend 'Gallup percentage of men favoring Dole'
galmenc 'Gallup percentage of men favoring Clinton'
galwomd 'Gallup percentage of women favoring Dole'
galwomc 'Gallup percentage of women favoring Clinton'

begin data
15 MR2      1  0  2  1  3  1  0  2  37  42  31  52
16 MR3      1  0  2  0  6  0  0  1  40  39  32  52
17 MR4      1  1  1  1  8  2  6  1
18 MR5      1  0  1  3  1  1  4
19 AP1      1  0  1  1  0  1  3
20 AP2      1  1  1  0  3  3  2  2  40  43  31  54
21 AP3      1  0  1  1  3  2  3  2
22 AP4      0  0  1  1  5  4  2  3
23 MY1      2  1  2  0  0  0  6
24 MY2      2  0  2  1  1  2  1  2  35  44  29  50
25 MY3      3  0  1  1  2  1  6  2
26 MY4      1  0  1  0  6  1  7  3
27 JN1      1  0  2  3  10  11  1  8  35  45  32  53
28 JN2      2  1  2  1  11  5  3  5
29 JN3      0  0  0  1  7  1  3  1  38  41  28  57
30 JN4      1  0  1  2  3  4  4  3
31 JU1      1  1  1  1  0  0  0  1
32 JU2      2  0  1  1  9  4  10  2
33 JU3      2  0  1  3  4  4  7  1  36  43  30  56
34 JU4      1  0  1  2  6  4  7  5  37  45  32  56
35 AU1      1  0  1  3  4  4  5  0  34  47  27  56
36 AU2      1  0  1  0  1  1  0  1  43  48  32  53
37 AU3      2  1  2  2  2  4  2  3  42  47  39  52
38 AU4      1  0  1  2  4  4  8  5  42  45  33  57
39 SE1      1  1  1  2  0  3  8  3  38  50  31  59
40 SE2      1  0  1  0  2  0  3  0  41  48  31  59
41 SE3      1  0  1  0  2  1  4  3  38  47  30  59
42 SE4      2  0  1  2  0  3  4  0  40  47  32  57
43 SE5      1  0  1  0  3  2  3  2  42  47  32  56
44 OC1      2  1  2  2  1  5  0  5  40  49  32  58
45 OC2      0  1  0  0  4  0  0  0  40  49  32  59
46 OC3      0  0  0  1  8  1  5  0  40  47  31  58
47 OC4      2  0  0  1  2  3  7  2  39  46  32  55
48 NO1      0  0  0  1  0  1  4  0  39  46  34  54

end data
compute gapw=galwomc - galwomd
compute gapm=galmenc - galmend
arima gapw with mdl to wc2 / model = (2,1,0) (0,0,0)
arima gapm with mdl to wc2 / model = (2,1,0) (0,0,0)
    
```

Table 6.2
Time-Series Results: Significant Parameters

A2	MD1	WD1	MC1	WC1	MD2	WD2	MC2	WC2*
Gap for Women Polled	-	-	-	-	-	-	-	-
Gap for Men Polled	+024	-008	-	-	-	-	+026	-058

* MD1 means masculine words associated with Dole with a direct one-step link
 WD1 means feminine words associated with Dole with a direct one-step link
 Those codes ending in 2 mean an indirect, two-step link
 Clinton codes have "C" instead of "D"

were added. The results identified the significant relationships shown in Table 6.2.

Gender words associated with candidates have no effect on the Clinton-Dole gap for women. For men, however, masculine words directly associated to Dole widen the gap in Clinton's favor. When more feminine words are directly associated to Dole, the gap with Clinton closes. This suggests that if Dole's media strategy had frequently associated him with feminine words and avoided masculine words, he might have won more votes.

Clinton's two-step, or indirect, word associations are significantly correlated with the gap. Masculine word associations to Clinton are associated with an increasing gap in his favor. Feminine word associations are associated with a decreasing gap, indicating that Dole is doing better relative to Clinton in males' candidate preferences. These indirect relationships are suggestive of association mediated by something else, perhaps particular issues.

DISCUSSION

At first glance it may be surprising to find that men, not women, are significantly affected by gender words linked to candidates. Consider, however, the literature cited earlier indicating that women are more issue-oriented than men, that women are historically more liberal than men and are more likely to vote Democratic. These orientations may account for our finding that gender-word associations linked either directly or indirectly to candidates

have no relationship to women's voter-preference changes over time during the campaign.

In short, the effects of these other factors are perhaps strong enough so that the particular candidate-linked gender-encoded language of news stories is of little consequence. The more deeply-rooted orientations may override the effects of specific message content. Women are widely thought to be more verbally and linguistically developed than men. Perhaps this increased sophistication allows them to go beyond the mere presentation of surface-level signs and allows them to better access the latent content of the texts.

On the other hand, prior to this research, one might have thought that because women are more linguistically developed, they would pay more attention to specific words. This study tilts interpretation toward the former explanation. Yet, given that it is only one study of one election, there is heightened value for future research that explores in more direct ways women's processing of text compared to men's. For example, the "think along" method, which would have respondents comment during their reading of stories, could be useful. This suggests value in looking at information-processing strategies of the genders as they attend to and interpret news and other messages about and from the candidates.

One way to obtain better data on voter information processing about candidates is to conduct exit polls that ask those leaving the polling place to explain why they voted for the candidates they did. Voter News Service is the consortium that conducted exit polls for major media organizations during the 1990s. Unfortunately, it only asked closed-end questions of voters leaving the polls.

In its focus on the language of news stories, however, this study is at least partially congruent with the language-focused approaches promoted by the critical theory and cultural studies perspectives. Adding value to qualitative analysis by linking it to empirical analysis, this study exemplifies how to provide a ground for falsification that is beyond the researchers' interpretation, unlike what occurs in literary analysis or rhetorical analysis.

Furthermore, this study may demonstrate that studies of media effects that do not analyze textual content are likely to be less fruitful. Even categorical analyses of media content are probably limited in finding effects because they do not measure messages and audience responses with the same model, a word-network approach to both. Unfortunately, these are speculations on our part because we were not able to analyze audiences' open-ended comments about why they prefer or dislike candidates.

Another value of this study is that it adds textual analysis to political communication electoral research, which has largely ignored the content of campaign messages as the basis for systematic analysis. We advance political communication research in finding language features associated with opinion poll data.

Limitations

This study, unlike some prior studies, only used newspapers as a source of media content (Lind & Danowski, 1998; Lind & Danowski, 1999). In the earlier studies our chief source of media texts was Journal Graphics, a company that provided transcripts of television and radio news texts from CNN, ABC, and a variety of public affairs shows. Such transcripts of television and radio news were not available for this time period. In early 1996 another firm acquired Journal Graphics and discontinued provision of these transcripts.

Nevertheless, in this study we used a diverse, broad sample of U.S. newspapers, so we have ruled out the rival explanation for our findings that the gender-language effects were peculiar to a particular publication. Still, our use of newspapers alone as media texts may lead some readers to believe that the results are only generalizable to the domain of the press. Technically this is true, for there is no evidence available for electronic news texts. But in a more theoretically exploratory vein, we note that hypotheses about other media effects can be reasoned from the current evidence and the findings of past research. We know from other research that broadcast news writes scripts that follow the leads of the press. News producers choose pictures to illustrate what the words say, rather than pick the pictures first, then write the story scripts.

From a different direction, new factors are shaping media texts in both print and electronic media, the Internet. Here, textual content is fundamental to Web page format and content. Journalists are now seeking to beat one another with the story by posting it on the Internet. This way, regardless of the medium's institutional time cycles, all can participate on a virtually real-time medium. Textual dynamics increasingly shape news content in a multimedia manner. So, we can expect some generalizability of the findings to electronic media as well.

One main rival explanation for our findings on women being unaffected by genderized candidate content is that women may not use newspapers as a primary political-information medium. This may be an explanation for our finding that women's candidate preferences were not significantly associated with news language, while men's were.

Another limitation of this study is that there was no good direct measure of the news reading of the polled individuals used for candidate-preference measures. Screening respondents with high readership of national and local newspapers could be expected to result in larger effects than those we observed in this research. This is encouraging for further research on news language and public opinion.

We limited our analysis to generic gender words, those serving general syntactic connector roles in showing gender. The particular word strings—themes and issues—that these gender words were connected with were not

of concern in our research. This is because our main goal for this chapter has been to demonstrate the use of tools for statistical content analysis in the arena of news stories and public opinion. It was not to foreground the content domain of gender gaps in respondents' stated preferences for presidential candidates.

SUMMARY

This study has illustrated a node-centric method of statistical content analysis of text, coding the direct and two-step links of gender words to presidential candidates' names in the 1996 election. To demonstrate how statistical content analysis can be linked with side data, this study did a time-series analysis of the gender links to the trends in Gallup presidential-candidate preference polls, conducted weekly in the heat of the campaign.

Results showed that for the women's candidate-preference gap, there was no significant difference for any of the direct or indirect gender links to either candidate. In contrast, the candidate-preference gap for men was significantly and positively associated with direct masculine links to Dole, negatively associated with direct feminine links to Dole, positively associated to indirect masculine links to Clinton, and negatively associated to indirect feminine links to Clinton.

Thus, the gap between Dole and Clinton increased in Clinton's favor when there were more direct masculine words linked to Dole, while the reverse occurred for feminine directly linked words. The gap reduced in Dole's favor when he was linked directly with feminine words. For indirect associations, masculine word associations to Clinton were associated with an increasing gap in his favor. Feminine word associations were associated with a decreasing gap, indicating that Dole was doing better relative to Clinton in males' candidate preferences.

The type of statistical content analysis we used in this study was a node-centric string analysis focusing on linkages to focal words of interest, in this case to candidate names. We tied these strings in a time-series design to poll data indexing differences between two groups, men and women. This research illustrates how, through this form of statistical content analysis, we were able to investigate research questions that would be much more difficult to address with manual content analysis.

We have linked empirical qualitative analysis, the measurement of a type of word strings, to quantitative poll data. Our approach also demonstrates how statistical content analysis need not be conducted only at global levels that represent macrolevel structures of word associations. Although we have found such analyses useful in prior research, the questions addressed here called for a different level of analysis, a more microlevel linguistic one that we were able to conduct using computerized content analysis in a novel way.

REFERENCES

- Barnett, G., & Danowski, J. (1992). The structure of communication: A network analysis of the International Communication Association. *Human Communication Research, 19*(2), 264-285.
- Barnett, G., Danowski, J. & Richards, W. (1993). Communication networks and network analysis: A current assessment. In G. Barnett & W. Richards (Eds.), *Progress in communication sciences* (Vol. 12, pp 1-20). Norwood, NJ: Ablex.
- Claffey, G. (1996). *Customer feedback: Using content analysis to reduce uncertainty in a changing environment*. Paper presented at the annual meetings of the International Communication Association, Chicago.
- Clark, C., & Clark, J. (1986). Models of gender and political participation in the United States. *Women and Politics, 6*(2), 5-26.
- Conover, P. (1996). Feminists and the gender gap. In M. Githens, P. Norris, & J. Lovenduski (Eds.), *Different roles, different voices: Women and politics in the United States and Europe* (pp. 51-60). New York: HarperCollins.
- Danielian, L., & Reese, S. (1989). A close look at intermedia influences on agenda setting: The cocaine issues of 1986. In P. Shoemaker (Ed.), *Communication campaigns about drugs: Government, media, and the public* (pp. 47-66). Hillsdale, NJ: Erlbaum.
- Danowski, J. (1982). A network-based content analysis methodology for computer-mediated communication: An illustration with a computer bulletin board. *Communication Yearbook, 6*, 904-925.
- Danowski, J. (1986). Interpersonal network radiality and mass and non-mass media behaviors, in G. Gumpert & R. Cathcart (Eds.) *Inter-media* (3rd ed., pp. 168-175). New York: Oxford University Press.
- Danowski, J. (1988). Organizational infographics and automated auditing: Using computers to unobtrusively gather and analyze communication. In G. Goldhaber & G. Barnett (Eds.), *Handbook of organizational communication* (pp. 335-384). Norwood, NJ: Ablex.
- Danowski, J. (1993a). An emerging macro-level theory of organizational communication: Organizations as virtual reality management systems. In L. Thayer & G. Barnett (Eds.), *Emerging perspectives in organizational communication* (pp. 141-174). Norwood, NJ: Ablex.
- Danowski, J. (1993b). Network analysis of message content. In G. Barnett & W. Richards (Eds.), *Progress in communication sciences* (Vol. 12, pp. 197-222). Norwood, NJ: Ablex.
- Danowski, J., & Choi, J. H. (1999). Convergence in the information industries: Telecommunications, broadcasting, and data processing, 1981-1996. In H. Sawhney & G. A. Barnett (Eds.), *Progress in communication sciences* (XV, pp. 125-150). Norwood, NJ: Ablex.
- Freeman, C., & Barnett, G. (1994). An alternative approach to using interpretive theory to examine corporate messages and organizational culture. In L. Thayer & G. Barnett (Eds.), *Organizational communication: Emerging perspectives* (pp. 60-73). Norwood, NJ: Ablex.
- Gidengil, E. (1995). Economic man—social woman? The case of the gender gap in support for the Canada-United States Free Trade Agreement. *Comparative Political Studies, 28*, 384-408.
- Gidengil, E., & Everitt, J. (1999). Metaphors and misrepresentation: Gendered mediation in news coverage of the 1993 Canadian leaders' debates. *Harvard International Journal of Press-Politics, 4*(1), 48-65.
- Habermas, J. (1979). *Communication and the evolution of society*. Boston: Beacon Press.
- Jang, H., & Barnett, G. (1994). Cultural differences in organizational communication: A semantic network analysis. *Bulletin de méthodologie sociologique, 44*, 31-59.
- Kern, M., & Just, M. (1995). The focus group method, political advertising, campaign news, and the construction of candidate images. *Political Communication, 12*(2), 127-145.
- Lind, R., & Danowski, J. (1996). *A semantic network analysis of male and female Afro-American and Euro-American assessments of television news stories*. Paper presented at the annual meetings of the International Communication Association, Chicago.
- Lind, R., & Danowski, J. (1998). The representation of Arabs in U.S. electronic media. In Y. Kamalipour & T. Carilli T. (Eds.), *Cultural diversity and the U.S. media* (pp. 157-167). Albany, NY: SUNY Press.
- Lind, R., & Danowski, J. (1999). A computational, linguistic analysis of the representation of the homeless in electronic media. In E. Min (Ed.), *Reading the homeless: The media's image of homeless culture* (pp. 109-120). Westport, CT: Praeger.
- Montoya, L. (1996). Latino gender differences in public opinion: Results from the Latino National Political Survey. *Hispanic Journal of Behavioral Sciences, 18*(2), 255-276.
- Newman, B., & Sheth, J. (1984). The "gender gap" in voter attitudes and behavior: Some advertising implications. *Journal of Advertising, 13*(3), 4-16.
- Ostrom, C. (1990). *Time series analysis: Regression techniques* (2nd ed.). Newbury Park, CA: Sage.
- Pratto, F., Stallworth, L., & Sidanius, J. (1997). The gender gap: Differences in political attitudes and social dominance orientation. *British Journal of Social Psychology, 36*(1), 49-68.
- Rice, R., & Danowski, J. (1993). Is it really just like a fancy answering machine? Comparing semantic networks of different types of voice mail users. *Journal of Business Communication, 30*, 369-397.
- Rozell, M., & Wilcox, C. (1998). A GOP gender gap? Motivations, policy, and candidate choice. *Women and Politics, 19*(1), 91-106.
- Shapiro, R., & Mahajan, H. (1986). Gender differences in policy preferences: A summary of trends from the 1960s to the 1980s. *Public Opinion Quarterly, 50*(1), 42-61.
- Togebly, L. (1994). The gender gap in foreign policy attitudes. *Journal of Peace Research, 31*(4), 375-392.
- Trevor, M. (1999). Political socialization, party identification, and the gender gap. *Public Opinion Quarterly, 63*(1), 62-89.
- West, D. (1993). *Air wars: Television advertising in election campaigns, 1952-1992*. Washington, DC: Congressional Quarterly.

- Wilcox, C., Hewitt, L., & Allsop, D. (1996). The gender gap in attitudes toward the Gulf War: A cross-national perspective. *Journal of Peace Research*, 33(1), 67-82.
- Wirfs, D. (1986). Reinterpreting the gender gap. *Public Opinion Quarterly*, 50(3), 316-330.
- Zhao, X., & Chaffee, S. (1995). Campaign advertisements versus television news as sources of political issue information. *Public Opinion Quarterly*, 59(1), 41-65.
- Zipf, G. (1949). *Human behavior and the path of least resistance*. Cambridge, MA: Addison-Wesley.

Semi-Automated Content Analysis of Pharmacist-Patient Interactions Using the Theme Machine Document-Clustering System

Bruce L. Lambert

INTRODUCTION

This chapter describes an automated system for extracting thematic features from computer-readable text. The system, dubbed the "Theme Machine," is based on methods for document clustering that were originally developed to facilitate information retrieval (Hearst & Pedersen, 1996; Rasmussen, 1992; Salton, Allan, Buckley, & Singhal, 1994; SPSS Inc., 1997; Voorhees, 1986; Willett, 1988). Preliminary studies indicate that the methods may also be useful for content analysis. In order to encourage further experimentation with similar methods, this chapter offers a methodological introduction and an example application of the Theme Machine. Specifically, we show how the Theme Machine can be used to illuminate issues that arise in medication-counseling interactions between pharmacists and patients.

Pharmacist-Patient Interaction

This is the age of chronic illness (e.g., heart disease, cancer, hypertension, diabetes), and by far the most common treatment for chronic illness is drug therapy. Prescription and over-the-counter drugs, when used properly, can