

Text Analysis

CARL W. ROBERTS

Abstract

Even once words have been counted, or their themes and semantics quantitatively rendered as networks or grammars, it remains unclear what they reveal. Are the texts windows into historical facts that the analyst cannot experience in person, or are they windows into their authors' perspectives? A choice is needed here, because authors' perspectives may alter their renderings of "the facts" and, conversely, changes in an author's surroundings may prompt changes in her or his perspective. Next, is the researcher a novice who strives for fidelity to authors' perspectives, or is the researcher an expert whose perspective affords insights unknown to the authors? With contemporary growth in both world population and communication technologies, increasing contacts among peoples with disparate perspectives afford the social sciences an opportunity both to improve our understanding of these perspectives (or cultures) and to discontinue mining words for evidence consistent with theoretical perspectives of our own choosing. Modality analysis is a promising method for performing historical-comparative analyses of political cultures based on the volumes of texts only recently available to us.

INTRODUCTION

With the advent of microcomputers, databases of digital texts, and scholars motivated to explore new research horizons, the uninitiated is faced with the obvious question, "What do all these words mean?" Not to be facetious, but addressing this question requires first understanding what the question itself means. If one is asking what "the meaning" of words is, then please read no further. This essay is predicated on the assumption that you, the reader, accept that the same words can have multiple meanings (depending on the perspective that the reader brings to them). Nearly every multi-cultural person (e.g., someone who has learned to speak another language in another country) has learned that languages are NOT neutral media through which the "same ideas" can be transmitted from one person to another—a fact of which mono-cultural people are typically unaware. Instead, languages embody perspectives that grammatically link concepts in ways that make sense to those who use them in communication. And so our first question is

answered as follows: The meaning of “what words mean” differs with the perspective used in their interpretation.

Having passed that hurdle (and assuming only multi-cultural readers to still be in attendance), we progress to a second question prompted by my answer to the first: “Isn’t it pointless to study words, since they can have as many (potentially infinite?) meanings as there are perspectives for interpreting them?” This question is most often stated rhetorically by a radical relativist for whom its obvious answer is in the affirmative. In this essay, I shall outline three approaches that text analysts have offered as alternatives to radical relativism. First, one may choose a perspective (or theory), and apply it to the words under investigation. Second, one may consider words solely as windows into historical events. And finally, one may strive for empathetic understanding, or *Verstehen*, and explication of perspectives behind the words sources (i.e., speakers—when words are transcribed—or authors) use when making sense of the events they experience.

RESEARCHER VERSUS SOURCE EXPERTISE

Classical scientific training situates researchers as experts in the research process, because it is they who select theoretical perspectives that, they believe, will most likely explain the empirical data at hand. Yet the social sciences differ from other sciences in that they deal with data produced by people who themselves have explanations for their own verbal and nonverbal behaviors. Thus, an initial methodological question that must be addressed in any text analysis is whether researcher or texts’ sources are experts in explaining what the sources’ words mean. These alternative approaches to analyzing texts have been respectively labeled instrumental and representational (Shapiro, 1997).

Starting with *The General Inquirer* (Stone, Dunphy, Smith, & Ogilvie, 1966), text analysis software was initially tailored for instrumental text analysis. By simplifying the task of counting words (and grouping them into generic themes, or “meaning categories”), this software empowered researchers to quickly detect differences and/or trends in themes among documents. Yet troubles inevitably arose when researchers applied “contingency analyses” to their word-count data. For example, after noting a significant correlation between the frequencies with which journalists mention “economic” and “political” themes, a researcher (as a contingency-analysis expert) might interpret the association as indicative of journalists’ belief that politics influences the economy (or vice versa). The problem with such interpretations is known in the methodology literature as the ecological fallacy (i.e., the error of inferring [clause-level] micro-relations from associations among aggregated data [word-counts])—a fallacy unfortunately not yet eradicated

from contemporary analyses of themes in texts. Thematic text analyses (i.e., analyses of theme frequencies) are legitimate for investigating what sources' words are about, but not for making inferences about semantic relations among these words.

Starting in the mid-1970s, one finds the beginnings of database developments and software aides to assist users in encoding semantic relations among sources' words, later including programs for mapping networks of relations among words used within large blocks of texts (Carley, 1993; Kleinnijenhuis, de Ridder, & Rietberg, 1997; Markoff, Shapiro, & Weitman, 1974; Roberts, 1989). Yet even given clause-level information on these relations, words' meanings remain indeterminate. This is because words have a "dual life." On the one hand, they are expressions of a perspective (or grammar) in terms of which sources communicate. On the other hand, sources use words to convey their sensory experiences to others. The "play" in communications thus lies in the fact that sources' perspectives will alter renderings of their experiences, and vice versa. (This is, for instance, why one experiences atrocity less often from freedom fighters than from terrorists, and why charitable and self-interested perspectives may alternate as experiences of hard times come and go.) In brief, a source's words will always vary both in their fidelity to a particular perspective (i.e., in their grammaticality) and in their correspondence with sensory experience. Thus, armed with relationally encoded texts, the text analyst is faced with two decisions: are perspectives or experiences to be analyzed? And, are sources' or the researcher's perspectives to be studied?

REALITY OR PERSPECTIVE

When interested in learning about events experienced directly by texts' sources, but only vicariously accessible to others, researchers have typically sought out text populations authored by unbiased sources (e.g., "disinterested" journalists with few political or organizational constraints on their writing). As previously noted, this strategy is needed to minimize perspective-based "tainting" of the historical realities being investigated. Illustrations of this approach to analyzing texts can be found in Phil Schrodts studies of political events based on lead sentences in news service stories (e.g., Schrodts, 2006; also see Franzosi, 2010).

Alternatively, theorists may also apply their own perspectives to sources' words. Rational choice theorists could interpret texts as expressions of self-interest; functionalists might understand them as efforts toward social wellbeing; conflict theorists would view them as justifications of existing power-relations; and so on. Yet the danger here is an ethnocentric one, whereby researchers' theoretical perspectives may blind them to how their

sources view their own communications. Nonetheless, it might be argued that researchers' expertise affords them insights of which their sources may be unaware (cf. Namenwirth & Weber, 1987, p. 237).

In the early 1990s, French conventions theorists began suggesting—first in economics then later in other social sciences—that subjects' (even unconscious) motives should not be assumed, but should instead be considered a variable to be empirically studied (Boltanski & Thévenot, [1991] 2006). Indeed, with expanding global population pressure and increasing mediated and unmediated contacts among peoples with disparate perspectives and life experiences, one might argue that, by requiring researchers to make theoretical assumptions about their subjects' motivations, *classical scientific methodology may render large segments of the social sciences superfluous* in the twenty-first century. At issue is not if humans are self-interested, altruistic, or indoctrinated; at times they are motivated in all these ways. Social scientists' pressing (and hopefully "emerging") research questions must address the conditions under which people embrace, modify, and change why they behave as they do. We ignore such questions at risk of becoming irrelevant to everyone but ourselves.

CONTENT VERSUS FRAME

The vast majority of empirical research on author and audience perspectives is associated with the as-yet problematically operationalized "frame" concept. Recognizing that the number of frame types might explode in direct proportion to the number of frame-studies, Snow and Benford (1992) suggested the development of generic "master frames" under which more specific frames might be subsumed. Yet their audience of social movement scholars has continued to expand the number of movement-specific frame measures, with little progress toward a generic master-frame typology in sight (cf. Benford, 1997).

The absence in the social movement literature of consensus on a set of master frames likely stems, at least in part, from a focus on frames that correspond to the idiosyncratic types of content espoused by movements' leaders (e.g., human rights, intelligent design, and environmental protection). Yet it also, again in part, may have resulted from their instrumental approach to analyzing movement rhetoric. For example, Gamson (1992) classified such rhetoric as referencing either "an existing injustice," "our agency," or "others' adversity"—three frames applied by a researcher-as-expert, rather than in accordance with a researcher-as-novice's empathetic judgments regarding frames intended by the rhetoric's authors.

In contrast, communications scholars have attained tentative consensus on a set of generic "news frames" (i.e., on the types of frames they use in

analyzing news stories). Only modest variations have been proposed to Neuman, Just, and Crigler's (1992; also see Semetko & Valkenburg, 2000) finding that US media and audiences frame a variety of news issues as economic (cost vs benefit), conflict (us vs them), powerlessness (controllers vs controlled), human impact (hindered vs helped), or morality (right vs wrong). Nonetheless, as Scheufele and Tewksbury (2007, p. 13) point out, such categories may not correspond to authors' frames (i.e., their "modes of presentation") but may instead merely reflect the content about which authors are communicating.

Consider, for example, arguments such as, "We need strong markets to prevent social unrest," and "We must prevent social unrest to keep markets strong." The former argument's content is about economics but is framed in terms of conflict; the latter argument's content is about conflict but is framed in terms of economics. Moreover, note that each argument begins with a rhetorical assertion of a specific content's inevitability (the *need*, respectively, for "strong markets" and for "preventing social unrest")—an inevitability socially constructed for each argument's audience. If persuasive, each social construction would entail its audience's passive acceptance of or active support for the constructed inevitability's realization. However, whether or not an audience is persuaded by such arguments will depend on its adoption of the (respective, conflict or economic) perspective in terms of which the author's argument is framed. For example, the former argument would not engender support for strong markets, if its audience were to contest the premise (i.e., disagree) that social unrest should be prevented. Accordingly, frame identification involves more from the researcher than merely counting words that can be classified as economic, conflict, and so on. It requires recognition of the words' semantic use.

In sum, news frames are rhetorical strategies for the presentation of news content. Framing success occurs when an audience simultaneously accepts the reality of this content while leaving the frame itself uncontested. Yet still unexplained is why the same five news frames have been recurrent in studies of nationally mediated news content.

SOCIAL CONSTRUCTION OF NATIONAL DISCOURSE

Journalists for democracies' largest media organizations typically write for a national audience (namely, the nation's electorate or citizenry). Addressing such an audience calls for content relevant to citizens' shared "national reality"—a reality consisting of legally sanctioned rules-of-conduct that are *operative* but potentially under threat (thereby calling for *preemptive* action), and that citizens may *apply* or, in the case of politicians, *manipulate*. Consequently, journalists may frame their arguments economically

(operative application), as security-related (preemptive manipulation), politically (operative manipulation), and as welfare-related (preemptive application). Finally, journalists also frame arguments “culturally” (i.e., by framing reality in terms of citizens’ “national heritage” as God-fearing, tolerant, self-sufficient, etc.). These five frames correspond (albeit roughly) to Neuman *et al.*’s (1992) previously mentioned generic frames (respectively, economic, conflict, powerlessness, human impact, and morality). So why do these five news frames repeatedly occur in democracies’ national media? As just suggested, it is because journalists for these media must formulate their arguments in ways that relate citizens to their nation—(i) a nation that impacts its citizens via rules-of-conduct that may be followed or changed; (ii) a nation that ensures citizens’ existence but with an existence of its own that must be ensured; and (iii) a nation with a unique cultural heritage. Thus, when making their arguments relevant to a nation’s electorate, journalists will tend to justify (or frame) them as having *economic* consequences in accordance with the nation’s rules-of-conduct, as having *political* consequences that alter citizens’ rules-of-conduct, as ensuring the nation’s *security-related* existence, as ensuring citizens’ *welfare-related* existence, or as perpetuating the nation’s *cultural* character (Popping & Roberts, 2009).

Now let us turn to consideration of what is being framed in national discourse. A radical relativist’s argument would seem to apply here, given the infinite types of content that might be framed in these five ways. Yet in applying news frames journalists justify not the content but the “reality” of content relevant to their citizen-audiences. Moreover, the most common linguistic device for portraying a person’s reality is found with modal auxiliary verbs (e.g., can, must, and should). For example, the argument, “We *must* prevent social unrest to keep markets strong,” is not that “to keep markets strong” is why citizens prevent unrest; it is why citizens’ prevention of unrest is *inevitable* (and thus real). Of course, this inevitability corresponds not to a physical reality but to a socially constructed one (i.e., a reality contingent on audience-acceptance). This said, my point is to emphasize that it is not the content (we prevent social unrest) but the content’s inevitability that is framed here in economic terms.

Modal arguments afford journalists with four types of reality claims: inevitability, possibility, impossibility, and contingency. Why four? The answer can be found in any introductory text on modal logic. Most sentences can be negated in only one way. For example, the sole negation of “we prevent unrest” is “we do not prevent unrest.” Adding a modal auxiliary verb to such a sentence results in three forms of negation. For example, the three negations of “we *can* (i.e., *are able to*) prevent unrest” are “we *are not able to* (i.e., *cannot*) prevent unrest,” “we *are able not to* (i.e., *do not have to*) prevent unrest,” and “we *are not able not to* (i.e., *must*) prevent unrest.” (The

same relation exists between *should* [or *be morally obligated*] and *may* [or *be permitted*], because anyone who *is not morally obligated not* to do something is *permitted* to do it.) Briefly put, modal assertions in nationally mediated news are claims regarding possibility, impossibility, inevitability, and contingency (i.e., non-necessity) within citizens' social reality. When one considers these four forms of *reality claims* in conjunction with the five news frames journalists use in justifying them, the following semantic grammar emerges for the encoding of modal arguments in democratic nations' editorial news.

A modality analysis begins with representative samples of editorials among countries and/or over time. (Unlike typical news stories, editorials allow journalists license to engage their readers in modal argumentation.) All instances of the modal auxiliary verbs of interest (typically, various forms of *can*, *must*, *should*, and *may*) are then located and paired with the news frames used to explain why a citizen can, must, etc. be or do something. These news-frame-plus-reality-claim pairs (i.e., these *modal arguments*) comprise the researcher's units of analysis. Next, each modal argument is encoded in one of twenty (five news frames \times four reality claims) ways. Statistical modeling of national differences and trends in modal argumentation afford inferences about how nations' political cultures may differ, and how these cultures may have facilitated or been changed by historical events.

CONCLUSION

This essay is a call for historical-comparative analyses of political cultures via analyses of modal arguments in mediated national news. Although such news has become increasingly available in machine-readable form, it is not text-availability that should motivate this research. More than at any prior time in human history, cultural and national boundaries tend to coincide. This coincidence results, on the one hand, from national education systems that promote and refine the rhetorical strategies embodied in the world's regional languages. On the other hand, the recent global explosion in democratic national governance has increasingly freed national news media to cater to audiences of citizens for whom news from their country is of primary relevance relative to other countries' news. Modality analyses of national discourse afford social scientists a means to better understand how political cultures (i.e., generally accepted rhetorical strategies for influencing citizens' behavior and decision-making) may influence, or be influenced by democratic functioning. Although only in its infancy, the outline of a research agenda is beginning to take shape.

In a modality analysis of national discourse in Hungary's largest newspaper, Roberts, Popping, and Pan (2009) found evidence that during Hungary's

first 7 years as a democracy political argumentation matured from assertions of political reasons why possibilities are open to citizens (what people *can* do as citizens) to political reasons for citizens' inevitabilities (what people *must* do as citizens). Contrasting findings on established democracies show them to have relatively stable political cultures. For instance, consistently between 1965 and 1999 when addressing healthcare issues, Canadian editorialists' modal arguments were overwhelmingly framed in welfare-related terms, whereas US editorialists were equally as likely to frame their modal assertions in economic as in welfare-related terms (Roberts & Liu, 2014). Tentative conclusions from these and other modality analyses suggest not only that established democracies may have more stable political cultures than fledgling ones, but that political discourse among established democracies may differ in whether debates focus on *how* to implement a broadly accepted justification (e.g., citizens' welfare in Canada) versus *why* a specific implementation (e.g., mandatory healthcare insurance in the United States) is justified.

A functioning democracy is one in which responsive politicians implement laws in terms of their constituents' debates on the policy issues before them (i.e., in terms of their respective political cultures). Thus, among functioning democracies, those with distinct political cultures will likely have predictable differences in both their laws and the social consequences thereof. For example, one might hypothesize greater poverty and less environmental protection in countries having political cultures in which policies are evaluated in economic terms rather than in terms of their impacts on citizens' welfare. Laws may be formulated in more abstract terms (e.g., the promotion of health) in countries with broad consensus on a single frame, but in concrete terms (e.g., the delivery of services) in ones having a political culture characterized by multiple contested frames.

Yet by incorporating "political culture" as a variable in hypotheses, modality analysis exposes researchers to constraints and risks that would be avoided with a more classical scientific methodology (namely, one in which subjects' motivations are assumed but not measured). For instance, researchers may only claim empirical support that a nation's political culture was a driving force behind a trajectory of policy developments, if they provide evidence that this culture remained relatively unchanged as the developments took place. (References to a nation's "cultural change" are oxymoronic, of course, except when frame-use is analyzed during fundamental political transitions or over time periods covering multiple generations or centuries.) The risk to a modality analyst is of the sort that all theorists take when applying the scientific method, namely the risk of not finding evidence for their theories. Most notably, evidence might be found *contrary* to a functionalist's presumption that stable political cultures

exist or to a conflict theorist's contention that policy implementations drive public debates. On the other hand, what may be risk for a theorist is surely advantageous for the social sciences and their combined purpose in pursuing truth. Instead of digging in our theoretical heels, should we not be moving from rhetorical polemics to empirical tests on how humans frame their own decisions and behavior?

REFERENCES

- Benford, R. D. (1997). An insider's critique of the social movement framing perspective. *Sociological Inquiry*, 67(4), 409–430.
- Boltanski, L., & Thévenot, L. [1991] (2006). *On justification: Economies of worth*. Princeton, NJ: Princeton University Press.
- Carley, K. M. (1993). Coding choices for textual analysis: A comparison of content analysis and map analysis. In P. Marsden (Ed.), *Sociological methodology, 1993* (pp. 75–126). Oxford, England: Blackwell.
- Franzosi, R. (2010). *Quantitative narrative analysis*. Thousand Oaks, CA: Sage.
- Gamson, W. A. (1992). *Talking politics*. New York, NY: Cambridge University Press.
- Kleinnijenhuis, J., de Ridder, J. A., & Rietberg, E. M. (1997). Reasoning in economic discourse: An application of the network approach to the Dutch press. In C. W. Roberts (Ed.), *Text analysis for the social sciences: Methods for drawing statistical inferences from texts and transcripts* (pp. 191–207). Mahwah, NJ: Lawrence Erlbaum Associates.
- Markoff, J., Shapiro, G., & Weitman, S. (1974). Toward the integration of content analysis and general methodology. In D. R. Heise (Ed.), *Sociological methodology, 1975* (pp. 1–58). San Francisco: Jossey-Bass.
- Namenwirth, J. Z., & Weber, R. P. (1987). *Dynamics of culture*. Boston, MA: Allen & Unwin.
- Neuman, W. R., Just, M., & Crigler, A. (1992). *Common knowledge: News and the construction of political meaning*. Chicago, IL: University of Chicago Press.
- Popping, R., & Roberts, C. W. (2009). Coding issues in modality analysis. *Field Methods*, 21(3), 244–264.
- Roberts, C. W. (1989). Other than counting words: A linguistic approach to content analysis. *Social Forces*, 68(1), 147–177.
- Roberts, C. W., & Liu, H. (2013). On the cultural foundations for universal healthcare: Implications from late 20th century US and Canadian health-related discourse. *Journal of Communication*, 64(4), 764–784.
- Roberts, C. W., Popping, R., & Pan, Y. (2009). Modalities of democratic transformation: Forms of public discourse within Hungary's largest newspaper, 1990–1997. *International Sociology*, 24(4), 498–525.
- Scheufele, D. A., & Tewksbury, D. (2007). Framing, agenda setting, and priming: The evolution of three media effects models. *Journal of Communication*, 57(1), 9–20.
- Schrodt, P. A. (2006). Forecasting conflict in the Balkans using hidden Markov models. In R. Trappi (Ed.), *Programming for peace: Computer-aided methods for international conflict resolution and prevention* (pp. 161–184). Dordrecht, NL: Kluwer.

- Semetko, H. A., & Valkenburg, P. M. (2000). Framing European politics: A content analysis of press and television news. *Journal of Communication*, 50(2), 93–109.
- Shapiro, G. (1997). The future of coders: Human judgments in a world of sophisticated software. In C. W. Roberts (Ed.), *Text analysis for the social sciences: Methods for drawing statistical inferences from texts and transcripts* (pp. 225–238). Mahwah, NJ: Lawrence Erlbaum Associates.
- Snow, D. A., & Benford, R. D. (1992). Master frames and cycles of protest. In A. Morris & C. Mueller (Eds.), *Frontiers in social movement theory* (pp. 133–155). New Haven, CT: Yale University Press.
- Stone, P. J., Dunphy, D. C., Smith, M. S., & Ogilvie, D. M. (1966). *The General Inquirer: A computer approach to content analysis*. Cambridge, MA: MIT Press.

FURTHER READING

- van Benthem, J. (2010). *Modal logic for open minds*. Stanford, CA: CSLI Publications.
- Roberts, C. W. (1997). Semantic text analysis: On the structure of linguistic ambiguity in ordinary discourse. In C. W. Roberts (Ed.), *Text analysis for the social sciences: Methods for drawing statistical inferences from texts and transcripts* (pp. 55–77). Mahwah, NJ: Lawrence Erlbaum Associates.
- Roberts, C. W. (2008). *'The' fifth modality: On languages that shape our motivations and cultures*. Leiden, The Netherlands: Brill.
- Roberts, C. W., Zuell, C., Landmann, J., & Wang, Y. (2010). Modality analysis: A semantic grammar for imputations of intentionality in texts. *Quality & Quantity*, 44(2), 239–257.

CARL W. ROBERTS SHORT BIOGRAPHY

Carl W. Roberts is Emeritus Professor of Statistics and Sociology at Iowa State University. His research on world cultures is motivated by an interest in how language shapes perception and behavior. He has written extensively on linguistic structure in articles appearing in *Social Forces* and *Sociological Methodology*, as well as in his edited collection, *Text Analysis for the Social Sciences* (1997). His substantive work investigates cultural variations in social discourse based on sampled texts ranging from medieval documents to editorials in contemporary foreign and domestic newspapers. (Personal webpage: www.public.iastate.edu/~carlos/)

Free text analysis software: www.stat.iastate.edu/tca/

RELATED ESSAYS

Statistical Power Analysis (*Psychology*), Christopher L. Aberson
 To Flop Is Human: Inventing Better Scientific Approaches to Anticipating Failure (*Methods*), Robert Boruch and Alan Ruby

Hierarchical Models for Causal Effects (*Methods*), Avi Feller and Andrew Gelman

Regression Discontinuity Design (*Methods*), Marc Meredith and Evan Perkoski