

Using Open Source News Collections for Political Science Research

A Validation Study of the FBIS and SWB Global News Archives

Paul Testa, Scott Althaus, and Leslie Caughell

August 25, 2015

Abstract

The archives of the Foreign Broadcast Information Service/Open Source Center (FBIS/OSC) and the BBC's Summary of World Broadcasts (SWB) may be the world's leading source of high-quality English-language translations of news coverage originating in every major country in the world from the 1940s through today. Yet questions about the origins and validity of these data have limited their use in academic research. We combine in-depth interviews with extensive content analysis to help scholars realize the inherent potential of these massive and underused datasets. Our interviews with current and former staffers at every level of these organizations provide qualitative insights into how and why these data were collected, thereby revealing the kinds of selection biases scholars might encounter when working with these data. Through content analysis, we are able take these insights one step further and empirically assess the magnitude of these selection biases as well as important differences between FBIS/OSC and SWB. We find that while SWB and FBIS/OSC clearly differ from a random sample of all possible news stories, they do so in predictable ways that make these data suitable for a wide range of research. Combining these qualitative and quantitative findings of our research, we offer guidelines for using SWB and FBIS/OSC in scholarly research.

Word Count: 11,206

Paper prepared for presentation at the 2015 Annual Meeting of the American Political Science Association, San Francisco CA, September 3-6.

¹We would like to thank Kalev Leetaru and Sean Reynolds for their assistance with this project, as well as the Cline Center of Democracy for providing access to these data and extensive research support.

INTRODUCTION

Since the early 1940s, the U.S. and British governments have collaborated on a joint effort to monitor foreign news broadcasts around the globe. Employing teams of translators, editors, and engineers, they provided high-quality, on-demand, English-language translations of open source information from every major country in the world. Originally founded to track Axis radio broadcasts during World War Two, the scope and mission of these organizations soon expanded to include coverage of print, broadcast, and now online media. The American partner in this collaboration is the Central Intelligence Agency’s Foreign Broadcast Information Service, reconstituted in 2006 as part of the CIA’s Open Source Center (FBIS/OSC). The British partner is BBC Monitoring’s Summary of World Broadcasts (SWB). Although untold billions of dollars have been spent on these high-quality English translations of news coverage over the past eight decades, they are little-known to academics and rarely used in scholarly research. In 2007, the Cline Center for Democracy at the University of Illinois Urbana-Champaign began compiling a text-minable archive of these unparalleled data sets. Today, the Cline Center’s global news archive contains over 4 million news reports compiled by SWB between 1979 and 2012, over 3 million news reports compiled by FBIS/OSC between 1994 and 2004, and another 6 million pages of digitized microfilm and microfiche records covering SWB and FBIS/OSC news translations from the 1940s through the 1990s.

The geographic and historical breadth of these datasets make them useful for a wide range of research in the social sciences and humanities. The data are already being used in cross-national research on conflict, civil unrest, and political communication. For example, the Social, Political, Economic Event Database (SPEED) incorporates SWB and FBIS/OSC along with historical archives of major U.S. papers with ongoing search of English language news sources from around the globe to measure civil unrest occurring worldwide (Nardulli and Slana, 2013; Nardulli, Althaus and Hayes, 2015). A SPEED study analyzing the effects of natural disasters on civil unrest estimated that SWB contained nearly 13 times the coverage of *The New York Times* (Nardulli, Peyton and Bajjalieh, 2015). Scholars have also mined these data to uncover patterns in-group discourse during the Iraq conflict (Althaus and Leetaru, 2011) and coverage of economic scandals in foreign media

(Leetaru, 2008). Recently, Shelef (2015) used FBIS to show how nationalist discourses defining territorial homelands increase the probability of conflict. These few studies, however, only scratch the surface of what is possible with these data.

Realizing this potential requires us to understand just what is (and is not) contained in these data. However, the selection criteria that guide their construction remain classified. And since few (if any) broadcast news outlets around the world have retained a complete description of the entire population of their own news coverage from the 1940s to the present, it is exceedingly difficult to identify or even enumerate the population of global news content from which the SWB and FBIS/OSC records were drawn. We can be sure of one conclusion: The content gathered by SWB and FBIS/OSC do not represent a simple random sample of global news coverage. As will be described in detail below, this material was collected to serve the strategic interests of the U.S. and British governments. Non-random, however, need not mean unrepresentative. Scholarly use of SWB and FBIS/OSC data requires understanding how these organizations decided what content to monitor, translate, and distribute, as well as how that selection process in turn shapes the kinds of data that are available for academic research.

This paper addresses these uncertainties head on. First, we combine information from historical accounts of these organizations with in-depth interviews from current and former SWB and FBIS/OSC staffers to describe the general contours of how these organizations operate. We use these insights to develop clear expectations for how the media content contained in SWB and FBIS/OSC likely differs from the population of global media that these collections draw from, as well as how these two sources may differ from each other. Next, we test these expectations using a content analysis of articles from a geographically diverse selection of 12 English-language newspapers monitored by SWB and FBIS/OSC during June 1994 and June 2003. This design produced a total 1,844 articles that we compared to a constructed week for each time period and paper. The 11,027 articles in these constructed weeks form a random sample of the population of news items that SWB and FBIS/OSC could have drawn from. This comparison allows us for the first time to assess the size and significance of potential selection biases in SWB and FBIS/OSC data.

Our findings show that SWB and FBIS/OSC differ from the population of news content they

draw from, and from each other, in important but predictable ways. We find that front-page news, by-lined stories, and editorial content are represented FBIS/OSC at higher levels than in the population of news content while SWB's reliance on online newspaper content makes any similar conclusions difficult to draw. Over time, both SWB and FBIS/OSC have increasingly relied on stories obtained from online rather than print versions of newspapers. Stories picked up by these data sources are somewhat more likely to emphasize national perspectives on topics than international or local perspectives. SWB and FBIS/OSC emphasize news coverage about foreign relations, government, and elections more than the population of news coverage from which they draw, while giving relatively less emphasis to domestic or economic issues. Stories mentioning either the U.S. or U.K. are somewhat more likely to be included than stories without such references when they contain a clear evaluative tone, but are otherwise no more likely to be selected than other stories. FBIS/OSC contains more articles than SWB, and is also more likely to include the full text of articles, while SWB contains more excerpted stories with ambiguous authorship. We also find circumstantial evidence that a considerable amount of translated content may be shared between these two organizations, with as much as 25 percent of the content in SWB and 11 percent of the content in FBIS/OSC originating from the other source.

Our findings illustrate both the tremendous potential of these data, as well as some important limitations. We summarize these insights by offering some recommendations for scholars who wish to use these data in their own research. Although not random, the global news data captured by SWB and FBIS/OSC represent a broad sample of domestic media from around the globe and are suitable for a number of large scale analyses involving event coding and textual analysis. At the same time, researchers must recognize that some types of stories and articles are more likely to be represented in these data than others, particularly in earlier time periods. We begin by presenting the results from our in-depth interviews. After summarizing the key findings from this work, we lay out our strategy for assessing these claims through content analysis. Next we present the results from this analysis, and conclude with a discussion of implications for further research.

HOW DO SWB AND FBIS/OSC OPERATE?

OVERVIEW

The global news archives of SWB and FBIS/OSC present a unique set of opportunities and challenges for scholars. The sheer breadth of their historic and geographic coverage make them potentially valuable to a number of subjects in political science, from research on inter- and intra-state conflict to theories of protests and social movements and studies of comparative political communication. Despite the inherent promise of these data, studies that use either FBIS/OSC or SWB are few and far between.¹

At least three related factors account for the limited use of SWB and FBIS/OSC. First, even though portions of SWB and FBIS/OSC have been available to researchers through commercial services like Dialog and LexisNexis, few scholars seem aware of their existence. Second, those few who are familiar with SWB and FBIS/OSC often report concerns about the generalizability of the data, confusion over what these data actually contain, and uncertainty about what kinds of the questions they can conceivably be used to answer. Some appear to suppose that SWB and FBIS/OSC are simply a collection of Kremlin speeches and press releases from Beijing. Third, the computer interfaces through which these data sets could be queried allowed for limited keyword searches that produced results for one-at-a-time human reading that made it difficult to explore the full breadth and depth of the archived holdings. Only recently have these datasets been brought into a text-minable archive where data analytics could reveal the broader dimensions of these massive holdings.

Addressing these concerns requires knowing something about how these organizations operate and why they collect the news holdings that they do. The challenge, for researchers, is that most of these details are classified. In this section, we attempt to clarify the strategic interests that drive SWB and FBIS/OSC by combining existing research and public documents with in-depth

¹A brief survey of the top three general-interest journals in political science, found only 11 studies citing these sources in the *American Political Science Review*, six in the *American Journal of Political Science*, and five in the *Journal of Politics*. The vast majority of these studies, used SWB and FBIS/OSC as source for historical anecdotes and quotations. The few empirical studies using these data have used them primarily as source for events data (Johnson, Slater and McGowan, 1984; Enders and Sandler, 1993) , and occasionally for more detailed content analysis (Frost, 1989; Leetaru, 2008; Althaus and Leetaru, 2011).

interviews of current and former staffers of FBIS/OSC and BBC Monitoring.

During the summer of 2014, we interviewed staff members with experience at every level of SWB and FBIS/OSC, from monitoring to management. All of our interview subjects had long and varied careers in terms of general duties, specific regions, issues, and media covered. We interviewed three former members of FBIS/OSC, J. Ransom Clark, Douglas Naquin, Anthony Olcott. Clark joined FBIS in 1965 and also served in multiple positions in the Far East, South America, the Middle East, and Europe until his retirement in 1990. Naquin joined FBIS in 1979 as editor of the Soviet Union Daily Report. He held a number of positions in Asia, Latin America, and the Middle East, eventually becoming the director of FBIS and overseeing its transition to the Open Source Center in 2005-06. Olcott joined FBIS in 1990 as a Russia analyst and became senior analyst on the New Media team within the Emerging Media group before retiring in 2013. We also interviewed two members of BBC Monitoring, Jack Doughty and Rosie Wolfe. Doughty joined BBC Monitoring in 1963, serving as a Russian monitor, Assistant Chief Monitor and finally as Senior Assistant, Video Monitoring before retiring in 1990. Wolfe currently serves as the Head of Business Development and Customer Relations, Partnerships and Rights for BBC Monitoring.

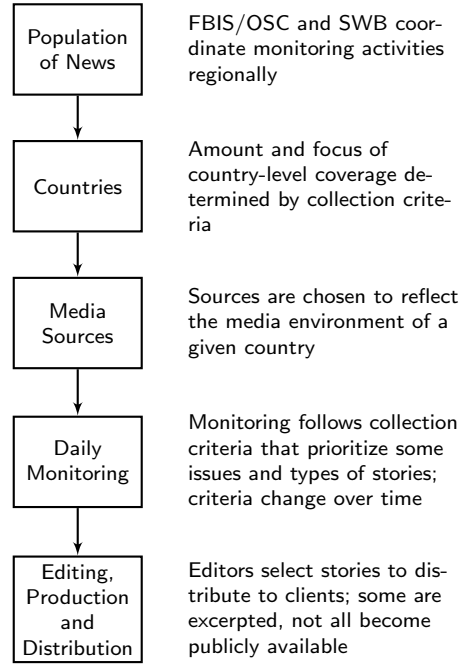
The breadth of experience contained in our interviews provides a unique and important perspective on the organization and operation of SWB and FBIS/OSC. Our goal is not to provide a definitive history of these organizations.² Nor do we claim to offer an exhaustive account of why and how the U.S. and British governments gather open source information, as this process continues to change over time.³ Instead, we wish provide a general overview of the way these organizations operate that highlights both the research opportunities presented by these sources as well as potential concerns that might affect how scholars should work with these data.

We begin by reviewing the multiple organizational processes that influence which news items eventually get selected for distribution. Figure 1 outlines five stages in the SWB and FBIS/OSC operations that are vital to understanding what these data are and how they can be used for scholarly research: 1) Institutional collaboration between SWB and FBIS/OSC; 2) The prioritization of

²See Roop (1969); Rotheray (2007); Shank (1991) for useful accounts.

³See Olcott (2012) and Bean (2011) for useful discussions of the value and challenges of open source information to the intelligence community.

Figure 1: Selection Process of FBIS/OSC and SWB



countries and topics; 3) Choosing media sources to monitor; 4) Daily monitoring; and 5) Editing, production, and distribution. For each stage, we summarize insights gained from our interviews and past research about SWB and FBIS/OSC. We focus particularly on how features of these stages may lead the information contained in SWB and FBIS/OSC to deviate from what one might find from a simple random sample of global news media.

INSTITUTIONAL COLLABORATION AND THE GLOBAL MEDIA ENVIRONMENT

The collaboration between SWB and FBIS/OSC dates back to World War II. The British began monitoring foreign broadcasts in 1939, and the Americans followed suit in 1941, with both sides pooling resources and sharing information. Although these efforts produced valuable information during the war, both organizations faced uncertain futures soon after the fighting ended. However,

the global rise of Communism and transition into Cold War eventually ensured that both countries would continue to place a premium on the insights that open source information could provide, particularly about closed societies (Roop, 1969; Rotheray, 2007).

The working partnership established during the war was formalized during a series of negotiations in 1947-48 and the basic tenets of this collaboration remain in effect today. As Clark notes “coverage responsibility (monitoring and translating) was a coordinated effort but how the material was used, presented, or disseminated rested with the individual organizations.”⁴ The two countries agreed to share resources and to divide the task of monitoring global media along geographic lines. The British focused on primarily on Europe and the Soviet Union. The U.S. provided coverage of East Asia and Latin America, and the two shared coverage of Africa, the Middle East, and other countries (Leetaru, 2010; Shank, 1991). Scholars should expect some overlap and duplication within the two sources reflecting this geographic division of labor, although our interviews suggest the boundaries of these divisions have grown more permeable over time.⁵

While SWB and FBIS/OSC provide a service of common concern for their respective countries, they serve different audiences and operate with different institutional constraints. FBIS/OSC is a part of the CIA, serving the broader intelligence community, as well as other government branches and agencies, such as the Department of State. Its budget is classified, but reporting by *Washington Post* put the 2013 budget for open source intelligence at \$614.8 million.⁶ Topical and geographic coverage is determined by the National Intelligence Priorities Framework, “a classified matrix of countries and topics ranked in terms of their relative importance to national interests.”⁷

SWB is produced by BBC Monitoring, which is a division of the British Broadcast Corporation. Like FBIS/OSC, it serves key customers within the government, particularly the Foreign and Commonwealth Office, the Ministry of Defence and the Cabinet Office and the Security and Intelligence Agencies, as well as BBC News and other commercial clients.⁸ Until recently, BBC

⁴Personal correspondence, J. Ransom Clark, July 1, 2014.

⁵Personal correspondence, Naquin, July 24, 2014.

⁶See “The Black Budget” *Washington Post* accessed online at <http://www.washingtonpost.com/wp-srv/special/national/black-budget/>

⁷Personal correspondence with Anthony Olcott, June 19, 2014.

⁸See BBC Monitoring Scheme, March 2013 available online at http://downloads.bbc.co.uk/bbctrust/assets/files/pdf/regulatory_framework/protocols/2013/bbc_monitoring_scheme.pdf

Monitoring was funded through grants from the government. Now it is funded through part of the BBC's general license fee with an annual budget in 2013 around \$31.6 million (£20.2 pounds).⁹ Like FBIS/OSC, BBC Monitoring's overall coverage is guided by collection protocols whose existence is public knowledge, but whose specific contents remain classified.¹⁰

Although SWB and FBIS/OSC share a common goal and may pool resources to some degree, they view their tasks differently. FBIS/OSC sees its mission as "provid[ing] open source material directly to US Government customers while also enabling the effective use of open sources in supporting the work of specific national security agencies and missions."¹¹ FBIS/OSC therefore is structured to provide both content and analysis. In contrast, BBC Monitoring makes it clear that it provides clients with open source information only, leaving analysis to the agencies that receive the information.¹²

Differences in the budget and missions of SWB and FBIS/OSC have implications for both the content and format of the information they provide. Consistent with its larger budget, FBIS/OSC tends to produce a higher volume of output, and several of our interviews suggested that FBIS/OSC is more likely to offer full-text translations.¹³ In contrast, SWB appears more likely to summarize or excerpt content, particularly when covering major events or breaking news.¹⁴

CHOOSING THE COUNTRIES AND TOPICS TO COVER

SWB and FBIS/OSC coordinate their coverage of global media according to one or more sets of classified guidelines. Our interviews and research nonetheless suggest that the implications of SWB and FBIS/OSC collection criteria are relatively straightforward. First, while SWB and FBIS/OSC provide global coverage of media from over 130 countries, the amount and depth of that coverage

⁹See BBC License Fee Settlement and Annual Report — Culture, Media and Sport Committee, accessed online at <http://www.publications.parliament.uk/pa/cm201012/cmselect/cmcumeds/454/45407.htm>

¹⁰These priorities are defined in Annex A, "Definition of 'Core Services' by Priority" and Annex B "List of Countries by Priority" of the BBC Monitoring Scheme available online at http://downloads.bbc.co.uk/bbctrust/assets/files/pdf/regulatory_framework/protocols/2013/bbc_monitoring_scheme.pdf. Both are listed as not for publication.

¹¹See OpenSource.gov FAQs, accessed online at https://static.opensource.gov/public/imageserver/webservices/opensource.gov/static_content/newHelp/faqs.htm

¹²Personal correspondence, Wolfe, August 6, 2014.

¹³Personal correspondence with Olcott, June 19, 2014.

¹⁴Personal correspondence, Doughty, August 5, 2014.

will vary to some extent according to the relative importance of that country in relation to the U.S. and U.K. at a particularly moment in time. Second, while SWB and FBIS/OSC monitor a broad swath of daily media, only a small subset of those stories will be deemed of sufficient interest to the U.S. and U.K. to merit translation and distribution. Third, while both sources contain coverage of events relating to and occurring in the U.S. and U.K. (Leetaru, 2010), neither organization directly monitors domestic sources within its own country.¹⁵

The geographic implications of SWB and FBIS/OSC's coverage priorities are relatively clear. During the Cold War, consistently high-volume coverage was given to the Soviet Union, China, and the areas these countries were active in, such as Eastern Europe and East Asia. Coverage outside these spheres of influence tended to be lower volume, more topically focused and more driven by local or global events that could create a temporary spike in interest such as a coup or hostage crisis. Over time, as the strategic priorities of the U.S. and U.K. changed, so too did coverage of SWB and FBIS/OSC, perhaps most notably with increased monitoring of the Middle East and a turn toward more globally-balanced coverage following the end of the Cold War.¹⁶

While researchers can have a reasonable understanding of which countries are likely to receive more monitoring attention, it is less clear what types of stories SWB and FBIS/OSC choose to select. Former FBIS/OSC Director Naquin describes "the basic service FBIS provided the entire US Government as a filtered and high-quality 'best of' compilation from the world's public information media (radio, TV, press, and press agency)."¹⁷ Similarly, Clark, suggests the organization's coverage extended beyond a simple bulletin of public affairs, to include "political, socioeconomic, technical, scientific, and other types of events that might shape a country's relationship with the United States, its neighbors, [and] the world community."¹⁸ Leetaru (2010) finds a similar breadth of coverage examining the topic tags FBIS/OSC began assigning stories in 1999. He found a broad mix of coverage about politics, government, the economy, and foreign affairs. Although the stories

¹⁵The reasons for this arrangement are complex, arising from both the longstanding "UKUSA" intelligence sharing agreements as well as the various laws in the U.S. and U.K. governing domestic surveillance. For more on the history of this relationship see Warner (1989); Lefebvre (2003); Dittmer (2015); Richelson (2015). While SWB does not monitor sources in the U.S. or U.K., FBIS/OSC does monitor sources with the U.K.

¹⁶Still, Leetaru (2010) finds that even after the fall of the Berlin Wall, reports from Russia and China still made up over 20 percent of the overall content in FBIS and SWB from 1994 to 2004.

¹⁷Personal correspondence, Naquin, July 24, 2014

¹⁸Personal correspondence, Clark, July 1, 2014

SWB and FBIS/OSC select are clearly not random, for example, there seem to be few stories about such popular topics as weather, sports, or the latest entertainment trends, the collection criteria seem to encompass a broad range of topics that could be representative of the general population of news items about public affairs. But without some baseline for comparison, it is difficult to say whether SWB and FBIS/OSC tend to cover some topics more than others. Such a baseline will be provided later in this paper.

SELECTING SPECIFIC MEDIA SOURCES

Given a set of geographic and topical priorities, SWB and FBIS/OSC identify a list of monitored sources and programs that will meet those collection criteria. Historically, both SWB and FBIS/OSC have expanded their monitoring efforts over time from an initial focus on radio broadcasts to later include print, television, and online news media.¹⁹

With regard to what sources and types of media get monitored within a specific country, Naquin describes the process as:

[L]argely a matter of evaluating the available media; determining the most important sources (based on reach, ownership, influence, etc.); establishing the pedigree of these sources (conservative, liberal, state-controlled, etc.); arranging collection/monitoring for these sources; and establishing customer requirements/priorities from these sources.

For many closed societies, the list of potential media sources is relatively short and clear. In more open media environments a number of political and social factors came into play. For example, in Chile, where different media sources were often aligned with specific parties, attention was paid to providing coverage from a politically diverse selection of pro- and anti-government sources.²⁰ Similarly, the source mix in different regions appears designed to reflected the way people in these regions actually consume media, leading to greater coverage of radio in Africa and of television in the Middle East.²¹ Outside of major official news services like China's Xinhua and Russia's ITAR-

¹⁹Leetaru (2010) finds that from 1994 to 2004 material from print sources made up about 50 percent of total coverage in FBIS and SWB with the remaining coverage split between radio and television.

²⁰Personal correspondence, Clark, July 1, 2014.

²¹Personal correspondence with Wolfe, August 6, 2014.

TASS, considerable effort is made to obtain coverage from primary sources, with wire services like Agence France Press and Reuters viewed as sources of last resort.²² Finally, SWB and FBIS/OSC monitor both English-language and vernacular sources, as these sources are often intended for different (e.g. foreign or domestic) audiences, and comparisons between coverage of the same event in different sources could yield important insights.²³

THE PROCESS OF DAILY MONITORING

Given the collection guidelines and source mix, the monitoring process is relatively straightforward. Within FBIS, monitors (often foreign nationals or people with linguistic and country specific knowledge), were assigned a list of predetermined sources and programs to cover. Some sources were covered live, while others were slated to be monitored within a day or other set time period after their initial broadcast (Shank, 1991) These monitors initially produced a “program summary” from which an editor (typically, a CIA-cleared intelligence officer who was also a U.S. citizen) would select a set of items for translation based on classified collection criteria.²⁴ SWB employed a similar process with monitors selected for their country-specific skills working closely with more senior staff to produce summaries, excerpts and translations for further distribution.²⁵ Monitors in both organizations received extensive training, and were subject to frequent review.

This process has changed over time both in response to current events and to secular trends. Both SWB and FBIS/OSC place an emphasis on “surgeability”, what the BBC Monitoring Scheme defines as “an ability to meet short term surges in demand for monitoring and output against Priority 1 countries, and to use reasonable endeavors to meet such surges against Priority 2 countries.”²⁶ While traditionally much of this monitoring was done in-house, technological and economic changes have led some of this work to be contracted out. For example, Wolfe describes how “in regions where news is ‘slow’ but customer interest is present, most of the monitoring is done by locally recruited

²²Personal correspondence, Naquin, July 24, 2014.

²³For example, Naquin recalls how during the Vietnam War, Hanoi Radio’s Vietnamese broadcasts tended to be far more vitriolic than their English-language programs.

²⁴Personal correspondence, Clark, July 1, 2014

²⁵Personal correspondence, Doughty, August 8, 2014

²⁶BBC Monitoring Scheme available online at http://downloads.bbc.co.uk/bbctrust/assets/files/pdf/regulatory_framework/protocols/2013/bbc_monitoring_scheme.pdf. Lists of “Priority 1” and “Priority 2” are unfortunately not made publicly available.

freelancers who are plugged into the news and media agenda well enough to be able to ‘switch’ on and off as required.”²⁷

Although staff in both organizations received extensive training, and were subject to frequent review, the monitoring process would almost certainly vary across monitors and editors, who might disagree about how the selection criteria applied to specific cases. This creates the potential for regional or local monitoring biases, but without access to classified details on organizational structure and staffing assignments, such biases would be difficult to identify. Given the scale and scope of these operations,²⁸ as well as the many layers of editing and frequent rotation of editors, we expect that selection biases at the level of individual monitors and editors should not bias the overall content of SWB and FBIS/OSC to any important extent.

EDITING, PRODUCTION AND DISTRIBUTION

Once complete, translated news items were reviewed by an editor and then passed on through the organization for further distribution. Over time, the frequency and nature of source publications and content types have changed. Certain items, like a speech by a major figure, would be “flashed” and tagged for expedited translation followed by immediate dissemination. The rest of the material was collected and published in daily reports organized by geographic regions.²⁹ Clients could request coverage of specific countries or topics. Early on, space constraints limited the number of stories that could be transmitted over teletype and published in weekly summaries. This appears to have created some incentive to give preference to shorter stories or excerpts when deciding which items to translate.³⁰ As technology has enabled more continuous publication and dissemination, it was not uncommon for the same story to appear multiple times in the feed, sometimes with corrections

²⁷Personal correspondence, Wolfe, August 6, 2014

²⁸BBC Monitoring employees about 370 people (BBC Monitoring, About Us, accessed online at <http://www.bbc.co.uk/monitoring/about-us>. Employment estimates for FBIS/OSC are more difficult to come by, however, news accounts during the transition from FBIS to OSC reported the organization had doubled the size of its staff (“CIA mines ‘rich’ content from blogs” *Washington Times*, April 18, 2006, accessed online at: <http://www.washingtontimes.com/news/2006/apr/18/20060418-110124-3694r/>).

²⁹The exact classifications have varied somewhat over time, but for much of FBIS’s history, the daily report was divided into eight parts: 1) Middle East, Africa, Near East and South Asia; 2) Sub-Saharan Africa and South Asia; 3) China; 4) Asia, Pacific and East Asia; 5) Latin America 6) Eastern Europe 7) Soviet Union and Central Eurasia; and 8) Western Europe. Similarly, BBC Monitoring’s SWB for the USSR, Eastern Europe, Far East, Middle East, Africa, Latin America & the Caribbean, and the rest of the world.

³⁰Personal correspondence, Olcott, June 19, 2014

or clarifications appended.

This final stage involves one additional and important instance of selection. Although SWB and FBIS/OSC monitor open sources of information, not all of what they produce is made available to the public. Sometimes, SWB and FBIS/OSC are unable to obtain the copyrights to distribute a given article. Additionally, there may be some risk that by disclosing a particular article or set of articles, FBIS/OSC or SWB might expose classified interests. As former Director of the CIA Michael Hayden put it: “The information is unclassified, our interest in it is not” (Hayden, 2008). Both Naquin and Wolfe estimated only about 70 percent of what SWB and FBIS/OSC produce becomes available to the general public through commercial services, and it is difficult to say what types of content are withheld from the public.³¹ SWB content can be accessed through LexisNexis under the source “BBC Monitoring” from 1979 to the present. Until recently, portions of FBIS’s content from 1995 to late 2013 were available to the public through the Department of Commerce’s World News Connection database. These portions continue to be available to institutional customers through Eastview Information Services (available at <http://wnc.eastview.com>), although this data stream is no longer updated by FBIS/OSC. To our knowledge, the Cline Center for Democracy at the University of Illinois is the only academic organization with a text-minable archive of SWB and FBIS/OSC news data. It is also the only academic organization with a text-minable archive of the complete microfilm records of both sources going back to the 1940s, although the Readex Corporation maintains a keyword-searchable archive of FBIS microfilm records from 1941 to 1996.³²

SUMMARY AND IMPLICATIONS

SWB and FBIS/OSC are no simple random sample of global news content. Researchers using these data need to know the consequences of their non-random collection processes. Our interviews in combination with available analysis of SWB and FBIS/OSC holdings confirm that these data sources are much more than an odd assortment of speeches and propaganda, as supposed by some. SWB

³¹Personal correspondences with Naquin, July 24, 2014 and Wolfe, August 6, 2014

³²Older Portions of FBIS remain accessible through subscription services like Readex: <http://www.readex.com/content/foreign-broadcast-information-service-fbis-daily-reports-1941-1996>. Both LexisNexis and ProQuest provide access to SWB.

and FBIS/OSC appear to cast a broad net in terms of the type of sources and stories they monitor. Yet this broad net is structured in ways that make the collections more suitable for some research questions than others. The findings from our interviews, summarized in Table 1, tell us both where to look and what to expect in assessing the relative representativeness of SWB and FBIS/OSC. For example, SWB's commitment to providing only open source information may make it more likely than FBIS/OSC to contain straight news articles as opposed to opinion and commentary. Although less likely to contain breaking news, opinion pieces and editorials, particularly those from government-controlled sources, may be more likely to contain information relevant to the intelligence interests of the U.S. government, and therefore more likely, to be included in the content of FBIS/OSC. The background information reviewed in this section presents testable, empirical questions about the selection process used to generate SWB and FBIS/OSC holdings. We now turn to a quantitative analysis of SWB and FBIS/OSC content that for the first time sheds light on the nature of selection biases in these data.

Table 1: Summarizing the differences between FBIS/OSC and SWB

Stage	FBIS	SWB	Implications
Institutional Structure	Operates within CIA	Part of the BBC	Duplicate coverage in SWB and FBIS/OSC; Potential differences in amount and type of coverage
Coverage Priorities	Geographic and topical priorities set by National Intelligence Priorities Framework	Geographic and topical priorities set by Monitoring Scheme	FBIS and SWB provide a non-random, but not-necessarily unrepresentative, sample of global news; Amount and type of coverage likely to vary by country, region, and time period
Source Mix	Determined by local political and media environment	Determined by local political and media environment	FBIS and SWB are intended to reflect the media actually consumed by residents of a country
Monitoring	Monitoring by people with linguistic and country specific knowledge. Selection by CIA cleared staffers using classified collection criteria. Potential outsourcing over time	Monitoring by people with linguistic and country specific knowledge; Potential outsourcing over time	Content potentially more representative at higher levels of geographic and temporal aggregation
Editing, Production, Distribution	Content disseminated to interested parties. Not all material made publicly available	Content disseminated to interested parties. Not all material made publicly available	Only about 70 percent of what's produced becomes publicly available

COMPARING SWB AND FBIS/OSC CONTENT TO THE POPULATION OF GLOBAL NEWS COVERAGE

OVERVIEW

The qualitative findings from the previous section suggest SWB and FBIS/OSC will differ from the population of global news content in important but predictable ways. Assessing the magnitude and significance of these potential biases requires some way of comparing content in SWB and FBIS/OSC to the population to possible content that these collections draw from. Such questions are naturally answered through content analysis ([Krippendorff, 2012](#); [Ridout, Fowler and Searles, 2012](#)). We first outline the research design of our content analysis study before turning to a discussion of results.

SELECTING SOURCES, TIME FRAMES AND STORIES

SWB and FBIS/OSC monitor thousands of print and broadcast outlets around the globe. Since there is no obvious population of translated news content with which to compare SWB and FBIS/OSC holdings, we focus our analysis narrowly on English-language daily newspapers for three main reasons. First, only print outlets have widely-available and authoritative records for defining a comparable set of news population data. While radio and television broadcasts make up a large portion of total output in these collections ([Leetaru, 2010](#)), in most cases there is simply no archive of transcript data for enumerating the population of broadcasts of even one station in any given country. In the few cases where such a population record might be possible to reconstruct, we deemed such an effort practically infeasible. In short, we conclude that there is no authoritative population data of broadcast news content against which the holdings of SWB and FBIS/OSC could be compared. But by obtaining the microfilm records of print newspapers, we can reconstruct the population of newspaper stories from which FBIS/OSC and SWB's monitors made their selections. Second, only English-language news sources could be used as a basis of comparison to SWB and FBIS/OSC holdings. SWB and FBIS/OSC data consist of both high-quality English translations as well as items originally published in English. Even if we could translate stories from non-English

newspapers, we could not be assured that our translations were as thorough and nuanced as those appearing in SWB and FBIS/OSC data. We therefore limit our analysis to English-language newspapers in order to remove the possibility that differences in the quality of our translations might bias any comparison to those produced by SWB and FBIS/OSC's trained monitors. Third, we are confident that the same selection criteria were used to select English-language news items as were used to prioritize non-English material for translation. English is often described as a *lingua franca* in global media (McCrum, 2011; Hachten and Scotton, 2015), and while English-language dailies are often directed to different audiences within a country, such as expatriates and foreigners (De Beer and Merrill, 2004), nothing from our interviews leads us to expect that SWB and FBIS/OSC would apply their overall collection criteria differently when monitoring these sources. If anything, our interviews suggest that an emphasis was placed on obtaining similar coverage across multiple sources to look for differences in how a topic may be presented to different audiences inside and outside of a given country.

From the subset of English-language newspapers monitored by FBIS/OSC and SWB, we selected 12 newspapers for inclusion in the study that had a sufficiently high volume of records in SWB and FBIS/OSC to permit a population comparison, and that maximized the geographic, political, and economic diversity of our sample: *Bangkok Post* (Thailand), *China Daily* (China), *South China Morning Post* (Hong Kong), *Business Day* (South Africa), *Daily Nation* (Kenya), *New Vision* (Uganda), *Jerusalem Post* (Israel), *The Dawn* (Pakistan), *Daily Telegraph* (United Kingdom), *Times of London* (United Kingdom), *The Independent* (United Kingdom), and *The Globe and Mail* (Canada).

For each newspaper, we examined all stories contained in SWB and FBIS/OSC for the months of June 1994 and June 2003. We chose these two time periods to capture any possible effects from the growing availability of newspapers on-line. Rather than attempting to analyze the entire population of stories from each of these newspapers in June 1994 and June 2003, we employed a common strategy in content analysis known as constructed-week sampling to make our task more manageable (Riffe, Aust and Lacy, 1993; Krippendorff, 2012). Constructed week-sampling has been shown to be more efficient than alternative strategies, such as random-day sampling, and avoids

potential biases related to news cycles found in consecutive-day sampling (Hester and Dougall, 2007; Riffe, Aust and Lacy, 1993). As the name suggests, we created a constructed week for the month of June by randomly selecting one of each of the seven days of the week for that month in 1994 and 2003. For each newspaper, we then attempted to obtain microfilm copies of that day's edition. We then coded all of the news and editorial content for that day.³³

During the collection phase of our study, we encountered two issues that forced us to alter our design an analysis. First, coverage of print sources in SWB did not extend back 1994.³⁴ Second, coverage of the print sources we selected in FBIS/OSC in 1994 was also more limited than we expected. Only four papers contained at least 50 items in June 1994 (*The Bangkok Post*, *China Daily*, *The South China Morning Post*, and *The Times of London*). Overall, this constructed-week sampling frame produced 11,027 stories to be coded and compared to 1,844 stories contained in the archives of SWB and FBIS/OSC. The results of our efforts are presented in Table 2.

Several factors warrant further discussion. First, because SWB included no print sources in 1994, we limited our primary analysis to results comparing the 1,525 stories from SWB and FBIS/OSC June 2003 to the 7,388 coded from our constructed-week sample for June 2003. We also analyzed FBIS/OSC content for 1994 and found the results of our analysis for 2003 are largely mirrored in 1994 data. In text, we make note of any major deviations between years for FBIS/OSC, and report a full set of comparisons for 1994, 2003, and combined samples in the appendix. Second, even in 2003, coverage in SWB is more limited than FBIS. Specifically, five papers that appear in FBIS are not included in SWB (*The Bangkok Post*, *China Daily*, *The Daily Telegraph*, *The Globe and Mail*, and *The Independent*). Third, for one paper, *The Times of London*, we were unable to obtain archival coverage for 2003. Fourth, in 2003, the Jewish holiday of Shavout fell on Friday, June 6. *The Jerusalem Post* did not publish an issue that day, and so we re-sampled the remaining Fridays in June, selecting June 13, 2013 for coding for that paper. Finally, some of the papers included in our sample did not produce publications every day of the week. Four publications, (*The Bangkok Post*, *Business Day*, *China Daily*, and *The Globe and Mail*) did not publish on Sundays. *Business*

³³We excluded stories explicitly located within sections like sports and arts and entertainment that our interviews suggested were unlikely to be monitored by SWB and FBIS/OSC.

³⁴SWB appears to have begun directly monitoring the online version of print sources in 1999. Text of from print sources prior to 1999 occasionally appears through SWB's coverage of wire services like Xinhua.

Day, *China Daily* and *The Jerusalem Post* did not publish on Saturdays, and *Business Day* also did not publish on Mondays.

DEVELOPING CATEGORIES FOR CODING

Our understanding of how SWB and FBIS/OSC operate leads us to expect that the content produced by these organizations will differ from the population of global media in certain predictable ways. To assess the magnitude of these differences and determine the representativeness of the data contained in FBIS/OSC and SWB, we developed a coding scheme to assess each story along multiple dimensions.

We considered a number of structural features of the content related to an article's location, type, origin, and completeness. First, each story was coded by its placement in the paper (front page, section front page, other page). During reliability coding, we created a fourth category, internet version, for stories in SWB and FBIS/OSC as it became clear that in 2003 both organizations increasingly drew content from online versions of these papers. Second, we coded stories by type (editorial, op-ed, news analysis outside the opinion page, and unlabeled straight news). Third, we coded each story's origin as falling into one of six categories: bylined story written by newspaper staff, unsigned story written by newspaper staff, wire service story, story written by another news outlet, guest author, and finally, a residual unable to tell/ambiguous origin category. Fourth, given that SWB and FBIS/OSC sometimes excerpt longer stories, we coded articles in SWB and FBIS/OSC as either full-text or excerpted.

Next, we examined the content of each article along a number of dimensions. First, using each article's headline, we coded the story for the presence of positive and negative evaluative tone, using linguistic indicators (verbs, adverbs, and adjectives). Stories could have either no clear evaluative tone, positive tone, negative tone, or both positive and negative tone. Second, drawing on the FBIS/OSC issue classification scheme (Leetaru, 2010) and the coding scheme for the Policy Agenda Project (Boydston, Bevan and Thomas III, 2014), we coded each article by the primary topic it addressed using six categories: government/elites, the economy, elections, other domestic issues, other foreign issues, and a residual miscellaneous category. Each story was coded as having

only one primary topic, and the categories were hierarchically ordered such that if a story addressed some aspect of the government and the economy it would be coded as the being primarily about government. Similarly, if it addressed an election along with some other domestic issue, it would be coded as an article about an election.³⁵ Third, we recognized that news coverage tends to be divided geographically and coded stories based on whether they had a primarily local, national or international focus. Finally, we coded a story as containing a mention of the U.S. or U.K. if it mentions the country, its leaders, its citizens, or companies based within these countries.

We extensively tested this coding scheme using human coders working with random samples of content from FBIS/OSC and SWB. Disagreements among coders were revisited and used to clarify a set of rules to guide the coding process. In the final reliability test, conducted on a random sample of 200 stories from 10 papers in 1994 and 2003, coders agreed between 88 percent and 99 percent of the time, with Krippendorff's alpha's ranging between a high of 0.95 for mentions of the U.S. and U.K. and a low 0.69 for the presence of both positive and negative evaluative tone.³⁶ The full set of reliability statistics is reported in the appendix.

EXPECTATIONS

Based on our understanding of how SWB and FBIS/OSC operate, our expectations for our content analysis are relatively straight forward. First, with regard to story placement, we know that newspapers tend to place more newsworthy or important items on the front page of the paper or section of the paper (Reisner, 1992). It seems possible then, that SWB and FBIS/OSC's selection criteria may magnify newspapers' perceptions of newsworthiness by oversampling front page content. Second, while straight news stories make up the bulk of newspaper content, we suspect that SWB and FBIS/OSC may be more interested in non-news pieces that may reflect the attitudes of government, particularly in more closed societies like China where the government closely controls the media. Third, it seems likely that SWB and FBIS/OSC will be more likely to focus on the original content provided by a paper and less likely to draw on newspaper content produced by

³⁵Further details, examples, and coding rules are provided in the appendix.

³⁶Our coders agreed close to 97 percent of the time on these cases and the relatively low alphas merely reflect the rarity of evaluative tone in our sample.

external sources, such as wire services, guest authors or other publications. Fourth, given the gap in budgets between FBIS/OSC, and SWB, as well as the broader focus of FBIS/OSC on mining open sources for intelligence, we suspect that FBIS/OSC will be less likely than SWB to excerpt or summarize content.

Our expectations with regard to content of monitored articles are similarly direct. First, we expect that SWB and FBIS/OSC may be more likely to include stories with evaluative tone, as they may provide insights into the general view of a society or government on an issue. Second, in terms of topical coverage, we know that SWB and FBIS/OSC place higher priority on some topics over others, however, the implications of these selection criteria for the representativeness of coverage remains an open question. We suspect SWB and FBIS/OSC monitor a broad range of topics, but were unsure how the distribution of topical coverage in these sources would compare to the population of global news. Third, while much of a newspaper's coverage may focus on local news, it seems likely that SWB and FBIS/OSC may be more interested in stories with national or international significance. Finally, we also expect that SWB and FBIS/OSC may be more likely to include stories that mention U.S. and British interests, since such coverage might be of particular interest to the government sponsors of SWB and FBIS/OSC.

Table 2: Newspapers and Story Counts in FBIS, SWB, and Constructed Week for 1994 and 2003

1994										
Paper	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Total	FBIS	SWB
	June 12	June 20	June 28	June 29	June 30	June 24	June 18			
<i>Bangkok Post</i> (Thailand)	55	158	156	186	202	153	106	1,016	61	–
<i>China Daily</i> (China)	62	46	36	39	60	43	43	329	128	–
<i>South China Morning Post</i> (Hong Kong)	127	160	200	237	199	177	184	1,284	73	–
<i>The Times of London</i> (UK)	151	106	159	164	170	140	120	1,010	57	–
2003										
	June 22	June 16	June 3	June 11	June 19	June 6/13	June 7	Total	FBIS	SWB
<i>Bangkok Post</i> (Thailand)	93	103	135	162	119	117	93	822	90	0
<i>Business Day</i> (South Africa)	–	–	167	154	138	158	–	617	54	26
<i>China Daily</i> (China)	–	62	69	78	93	66	–	399	285	0
<i>Daily Nation</i> (Kenya)	–	54	76	71	69	59	50	379	102	92
<i>Daily Telegraph</i> (UK)	125	78	84	100	195	111	142	745	78	0
<i>Jerusalem Post</i> (Israel)	70	72	66	63	74	103	–	448	59	14
<i>New Vision</i> (Uganda)	43	122	112	110	84	93	93	657	108	82
<i>South China Morning Post</i> (China)	69	148	184	152	162	173	176	1065	238	9
<i>The Dawn</i> (Pakistan)	148	130	149	146	152	145	140	1010	105	2
<i>The Globe and Mail</i> (Canada)	–	79	100	106	101	98	98	582	98	0
<i>The Independent</i> (UK)	90	73	92	98	112	84	115	664	83	0
Notes: The table presents daily story counts for our constructed-week sample and monthly totals for SWB and FBIS/OSC in June 1994 and June 2003. Entries with – denote days a newspaper did not publish an edition. SWB did not include print coverage in 1994										

RESULTS

In this section, we present the results from our content analysis. We begin by briefly comparing the amount of content in SWB and FBIS/OSC to estimates of the overall population from which they were drawn. Then, we compare the structural characteristics of the content in SWB and FBIS/OSC (story location, type, origin, and completeness) to our constructed-week sample. Next we consider differences in the content across these samples, looking first at story tone, topic and geographic focus. We discuss both aggregate comparisons within categories, as well as some marginal comparisons across, such as the relationship between tone and coverage of the U.S. and U.K. Finally, we present estimates of the degree of overlap between SWB and FBIS/OSC using a list common words spelled differently in the American and British English.

For the reasons discussed in the previous section, our analysis here pertains only to data from 2003. While we focus primarily on the general trends that arise from the data, we also use regression analyses to provide more formal tests of the patterns that emerge. The appendix provides a host of tables and statistical detail presenting the raw counts, proportions, and statistical analysis of the data across samples and years.

SWB,FBIS/OSC AND THE GLOBAL POPULATION OF NEWS

The archives of SWB and FBIS/OSC contain millions of stories from around the globe. To place the scope of this content in perspective, we use the total story counts from our constructed-week sample of June 2003 to produced monthly estimates of the total population of newspaper stories monitored in this group of English-language titles by SWB and FBIS/OSC. Table 3 presents the results of these calculations. On average, SWB contains just under 1 percent of the total stories produced, while FBIS/OSC captures about 4.4 percent of the total content. Coverage also varies across source, with Kenya’s *Daily Nation* the most covered outlet in our sample for SWB (6.1 percent) and *China Daily* in FBIS/OSC (17.9 percent). It is apparent that SWB and FBIS/OSC contain only small fractions of the total possible content. Yet, as we show below in our content analysis, even this thin slice remains representative of many aspects of global news content.

Table 3: Percentage of Population Content Found in SWB and FBIS/OSC for June 2003

Paper	Population		SWB		FBIS/OSC	
	N (C-W)	N (Pop. Est)	N	(% Total)	N	(% Total)
<i>Bangkok Post</i> (Thailand)	822	3288			90	2.7
<i>Business Day</i> (South Africa)	617	2468	26	1.1	54	2.2
<i>China Daily</i> (China)	399	1596			285	17.9
<i>Daily Nation</i> (Kenya)	379	1516	92	6.1	102	6.7
<i>Daily Telegraph</i> (UK)	745	2980			78	2.6
<i>Jerusalem Post</i> (Israel)	448	1792	14	0.8	59	3.3
<i>New Vision</i> (Uganda)	657	2628	82	3.1	108	4.1
<i>South China Morning Post</i> (China)	1065	4260	9	0.2	238	5.6
<i>The Dawn</i> (Pakistan)	1010	4040	2	0	105	2.6
<i>The Globe and Mail</i> (Canada)	582	2328			98	4.2
<i>The Independent</i> (UK)	664	2656			83	3.1
Total	7388	29552	225	0.8	1300	4.4

Notes: Population estimates obtained by multiplying constructed-week (C-W) totals by four. Percent totals for both SWB and and FBIS/OSC are calculated using the total population estimate for the sample

STORY LOCATION, TYPE, ORIGIN AND COMPLETENESS

Figure 1 presents the results from our analysis of story location, type, origin, and completeness for June 2003. The first column in each panel presents the relative distribution of each category for the constructed-week sample of population coverage, which we compare to the distributions in FBIS/OSC (center column) and SWB (right column). Turning first to the results for story location, presented in the upper-left panel, two trends emerge. First, in 2003, both SWB and FBIS/OSC obtained the bulk—and for SWB all—of their content from the online versions of these papers.³⁷ As we show in the appendix, in 1994, however, over a quarter of the material in FBIS/OSC came from the front page or section fronts of newspapers, compared to just six percent of the articles in our constructed week. The selection criteria employed by FBIS/OSC (and presumably SWB) appear to mirror and magnify newspaper’s decisions about the kinds of stories that warrant front-page coverage.

Next we examine the types of content (news, editorial, opinion, and analysis) contained in

³⁷About 3 percent of the articles (N=40) in FBIS/OSC for 2003 make reference to a specific location in the print paper. The rest (97 percent, N= 992) came from online versions of papers.

FBIS/OSC and SWB. The top-right panel in Figure 1 largely fits with our intuitions and expectations derived from our qualitative discussions. First, straight news stories are the most common type of content across all three samples. Editorials, opinion pieces and analysis, however, appear to be overrepresented in FBIS/OSC and, to a lesser extent, in SWB. We should also note that since our coding scheme required op-eds be located within the opinion or editorial section of a paper, no op-eds were coded in SWB, although about eight percent (N=17) were coded as being news-analysis.

Figure 1: Story Location, Type, Origin and Completeness

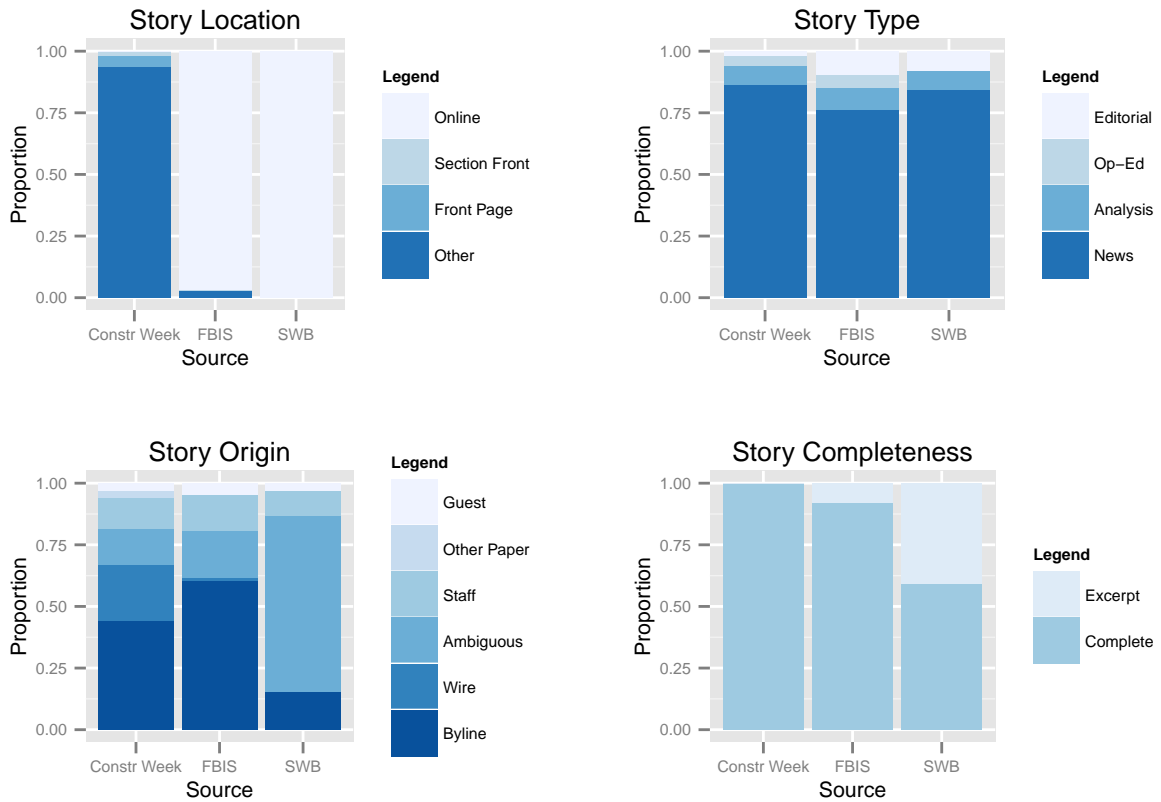


Table 4 presents a more formal test of this analysis, using a multinomial logistic regression to compare story type across the samples. The coefficients on the constant tell us about the frequency of other content types (editorial, op-eds, and analysis) in the constructed-week sample relative to news content, the referent (excluded) category for the dependent variable. That is,

editorial, op-eds, and news analysis all occur less frequently than straight news content in our constructed-week sample. The coefficients on SWB and FBIS/OSC tell us how these samples differ from the constructed week, for each category. Editorial content is far more common in FBIS/OSC and SWB, as are op-eds and news analysis in FBIS/OSC. Suggestions that SWB focused more on straight news than FBIS/OSC are also evident in Figure 1, although the magnitude of these differences is relatively small (76 percent of the stories in FBIS/OSC vs 84 percent of the content in SWB) and non-significant in multinomial logistic regressions.

With regard to the origin of content, we expected that SWB and FBIS/OSC would be more likely to draw on original content published by papers and less likely to include material produced by secondary sources such as wire services. These expectations are clearly supported for FBIS/OSC in the lower-left panel in Figure 1. While close to half of the content in our constructed-week sample was coded as coming from a newspaper’s writers or staff, nearly three-quarters of the content in FBIS/OSC came from such sources. For SWB, the story is more complicated, as more than 70 percent (N=160) of the content lacked information that would allow us to determine the story’s origin. This appears to be due in part to SWB’s greater tendency to excerpt and summarize content, revealed in the distributions presented in the lower right panel of Figure 1. In SWB, more than 40 percent of the content was excerpted or summarized and for close to 90 percent of these cases (N=81) we were unable to make final determination of a story’s origin (e.g. bylined staff, wire service, etc.). In contrast, FBIS/OSC provided the full-text for the majority of its content (92 percent, N=1,199).³⁸

³⁸This is true in FBIS/OSC for both 1994 and 2003. In fact, the percent of excerpted content is slightly higher in 2003 (7.7 percent, 101) than in 1994 (3.4 percent N=11)

Table 4: Differences in Type of Stories Contained in FBIS and SWB Relative to Constructed-Week Sample

	<i>Dependent variable:</i>		
	Editorial (1)	Op-Ed (2)	Analysis (3)
FBIS	1.789*** (0.129)	0.355** (0.140)	0.243** (0.107)
SWB	1.507*** (0.261)	-7.217 (12.403)	-0.033 (0.257)
(Constant)	-3.861*** (0.087)	-3.065*** (0.059)	-2.380*** (0.043)
Akaike Inf. Crit.	10,195.060	10,195.060	10,195.060

Note: *p<0.1; **p<0.05; ***p<0.01

The table presents the results of a multinomial logistic regression comparing content type in FBIS and SWB to the constructed-week sample with straight news content as the excluded category for the dependent variable.

STORY TONE, TOPIC, AND GEOGRAPHIC FOCUS

Next, we examine the content of the stories, comparing the differences in evaluative tone, primary topic, and geographic focus of SWB and FBIS/OSC to our constructed-week sample. The relative distributions from our content analysis are presented in Figure 2. We see from the first, top-left panel that stories containing evaluative tone are relatively rare in the constructed-week sample: only about nine percent (N=972) of the stories deemed to possess a somewhat positive or negative evaluative tone from their headlines or first sentences. In contrast, almost twice as many stories in FBIS/OSC (19 percent, N=253) and SWB (21 percent, N=49) possessed a clear evaluative tone, suggesting both organizations disproportionately pay attention to stories that take a clear side on an issue. Moreover, compared to the constructed week population data, there seems to be some evidence of a negativity bias in FBIS/OSC. Stories with a negative evaluative tone are over five times more common in FBIS/OSC than stories with positive evaluative tone, compared to the rate of 2.7 times more common in the constructed-week sample.³⁹

Given the way SWB and FBIS/OSC operate, we expected the topics contained in these data to systematically differ from the underlying population distribution. However, without a baseline for comparison, we had no way of assessing the magnitude and implications of these differences. The top-right panel of Figure 2 provides some crucial insights into this issue. We see from the first column that stories relating to economic issues are the most common in our constructed-week sample (33 percent, N=2,446), followed by coverage of domestic issues (32 percent, N=2,396), foreign/international news (15 percent, N=1,103), and news relating to the government and political elites (13 percent, N=968). Coverage of elections was relatively rare (1 percent, N=103) perhaps due in part to the time periods we selected. In contrast, SWB and FBIS/OSC provide greater coverage of events relating to foreign or international affairs (32 percent in FBIS/OSC, 35 percent

³⁹The differences in proportions of negative to positive tone between FBIS/OSC and the constructed week is significant at $p < 0.01$. SWB was also more likely to include content with negative evaluation tone, but the rates do not differ significantly from the constructed-week sample ($p = 0.61$). We also examined the types of stories that made such evaluations. Close to 72 percent (N=123) of the stories containing a positive evaluative tone in our constructed week sample were news stories. Similarly 75 percent (N=9) of the content with positive evaluative tone in SWB and 61 percent (N=25) of the content in FBIS/OSC, came from news articles. For negative evaluative tone, however, 50 percent or more of the content in FBIS/OSC (N=117) and SWB (N=19) came from editorials and op-eds, compared to just under 30 percent of the content (N=146) in our constructed week sample.

of the stories in SWB), and governments and political elites (15 percent in FBIS, 25 percent in SWB). They provide somewhat less coverage of economic events (21 percent in FBIS/OSC, 21 percent in SWB), and provide comparable coverage of elections and domestic news⁴⁰

Figure 2: Story Tone, Topic, and Geographic Focus

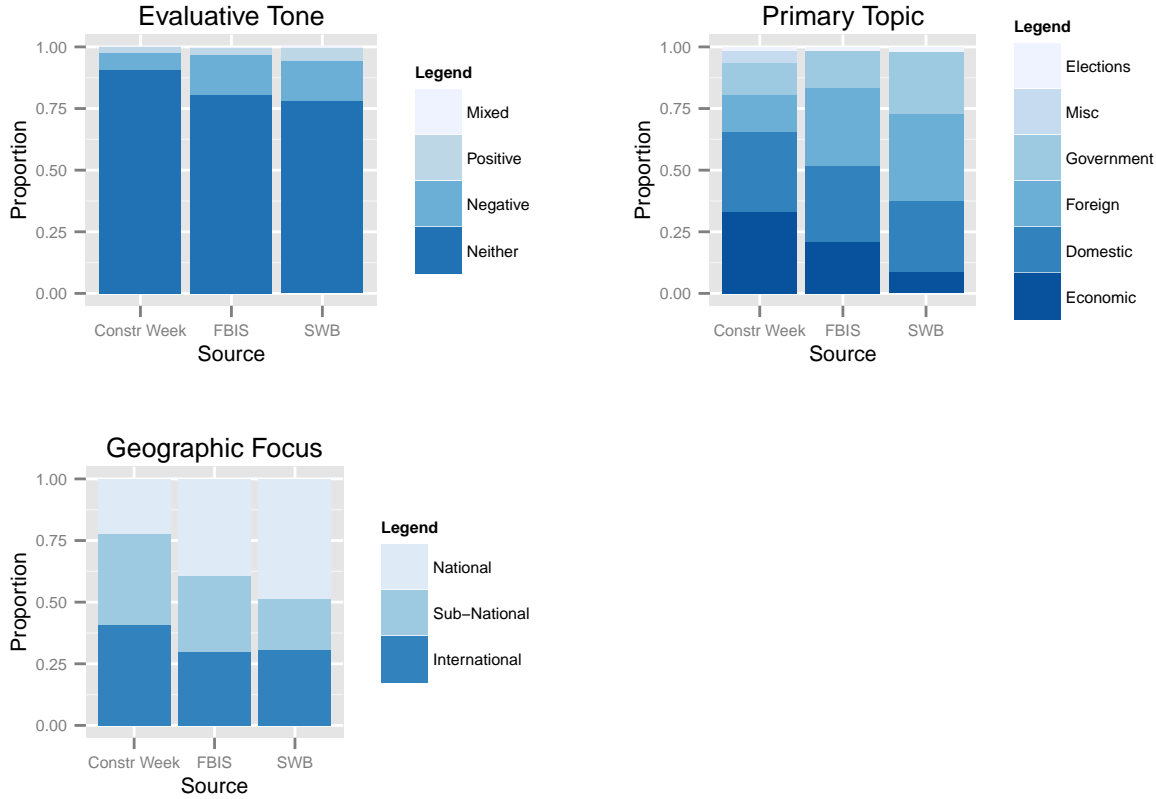


Table 5 provides a more formal assessment of these patterns of topical coverage, again using a multinomial logistic regression. In the models, economic coverage is the excluded, referent category for the dependent variable. The coefficients on the constant tell us how the coverage of other topical categories in our constructed-week sample differs from the baseline rate of economic coverage (i.e. the remaining topical categories occur less frequently than the economic coverage). The coefficients for SWB and FBIS/OSC tell us how each of these samples differ from the constructed week. We

⁴⁰All of the content in SWB and FBIS/OSC was coded as falling in one of the five primary topical categories, while five percent of the content (N=372) in our constructed-week sample fell outside of these primary topic categories and was coded as “miscellaneous.”

see that relative to the constructed week, SWB is significantly more likely to feature news related to governments, elections, foreign and domestic events. With the exception of election coverage (a relatively rare event in our sample) the differences are also significant, if smaller, for FBIS. Overall than, economic coverage appears to be underrepresented in FBIS/OSC and SWB, while coverage of foreign and international events, as well as of government elites, appears to be somewhat overrepresented, particularly in SWB.

The lower-left panel of Figure 2 examines the geographic focus of coverage in our three samples. Within the constructed-week sample, the majority of coverage focuses on events occurring within a country, either at the national level (22 percent, N=1,638) or sub-national/local level (37 percent, 2,730) with the remaining coverage focused on events occurring outside a country's borders (41 percent, N=3,020). In contrast, within SWB and FBIS/OSC, stories with a national geographic focus are the most common (39 percent in FBIS/OSC, 48 percent in SWB) with significantly less coverage given to local news and about events outside a country.

Table 5: Differences in Topical Coverage of FBIS and SWB Relative to Constructed-Week Sample

	<i>Dependent variable:</i>				
	Government	Elections	Domestic	Foreign	Misc
	(1)	(2)	(3)	(4)	(5)
FBIS	0.603*** (0.101)	0.383 (0.269)	0.388*** (0.084)	1.201*** (0.086)	-7.115 (5.425)
SWB	1.971*** (0.262)	1.562*** (0.555)	1.196*** (0.257)	2.167*** (0.253)	-8.047 (32.022)
(Constant)	-0.927*** (0.038)	-3.167*** (0.101)	-0.021 (0.029)	-0.797*** (0.036)	-1.884*** (0.056)
Akaike Inf. Crit.	26,317.520	26,317.520	26,317.520	26,317.520	26,317.520

Notes: *p<0.1; **p<0.05; ***p<0.01

The table presents the results of a multinomial logistic regression comparing topical coverage in FBIS and SWB to the constructed week sample with economic stories as the excluded category for the dependent variable.

STORIES THAT MENTION THE US OR UK

Given their clients in the American and British governments, it seems reasonable to expect that SWB and FBIS/OSC might give disproportionate attention to stories that mention the U.S. or U.K. The top row in Figure 3 suggests, in the aggregate, these differences are relatively small. Stories mentioning the U.S. (or figures and entities related to the U.S.) appear in 31 percent (N=2,308) of the content in our constructed-week sample, compared to 35 percent (N=461) of the content in FBIS/OSC and 28 percent (N=62) of the content in SWB. Similarly 23 percent (N=1,705) of the content our constructed-week contained a mention of the U.K. (or U.K. related interests), compared to 19 percent (N=236) of the content in FBIS/OSC and 16 percent (N=35) of the content in SWB. However, as we see from the bottom pair of charts in Figure 3, stories that mention the U.S. with a clear evaluative tone appear much more likely to be included in FBIS/OSC and SWB. Table 6 provides a more formal test of this association. The differences between SWB and FBIS/OSC and the constructed-week sample appear to be driven by these organizations' greater propensity to select content that mentions the U.S. or U.K. with a clear negative evaluative tone. Interestingly, for SWB but not FBIS, stories which contain a positive evaluative tone when mentioning the U.K. are also more likely to be included.

Figure 3: Mentions of US and UK and Story Tone

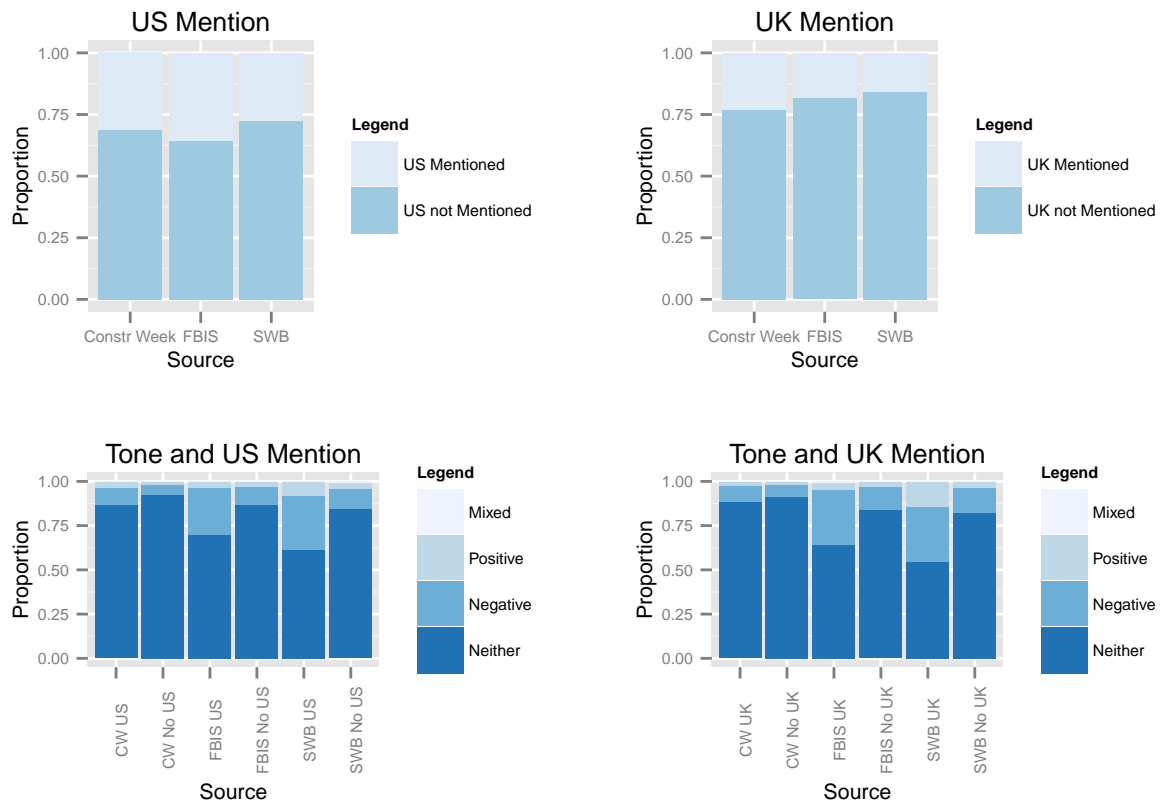


Table 6: Coverage of the U.S. and U.K. in FBIS and SWB by Evaluative Tone

	<i>Dependent variable:</i>	
	U.S. Mentioned	U.K. Mentioned
	(1)	(2)
FBIS	-0.004 (0.064)	-0.549*** (0.078)
SWB	-0.363** (0.183)	-0.987*** (0.239)
Negative Tone	0.441*** (0.078)	0.276*** (0.082)
Positive Tone	0.568*** (0.124)	0.126 (0.136)
FBIS*Neg	0.644*** (0.164)	0.894*** (0.174)
SWB*Neg	0.805** (0.379)	0.849* (0.437)
FBIS*Pos	-0.191 (0.304)	0.114 (0.360)
SWB*Pos	0.259 (0.630)	1.517** (0.647)
Constant	-0.904*** (0.022)	-1.086*** (0.023)
Observations	12,871	12,871
Log Likelihood	-7,817.392	-7,142.693
Akaike Inf. Crit.	15,652.780	14,303.390

Note: *p<0.1; **p<0.05; ***p<0.01
The table presents the results of logistic regressions comparing coverage of the U.S. and U.K. by evaluative tone in FBIS and SWB to the constructed week sample.

ESTIMATING THE DEGREE OF OVERLAP IN SWB AND FBIS/OSC

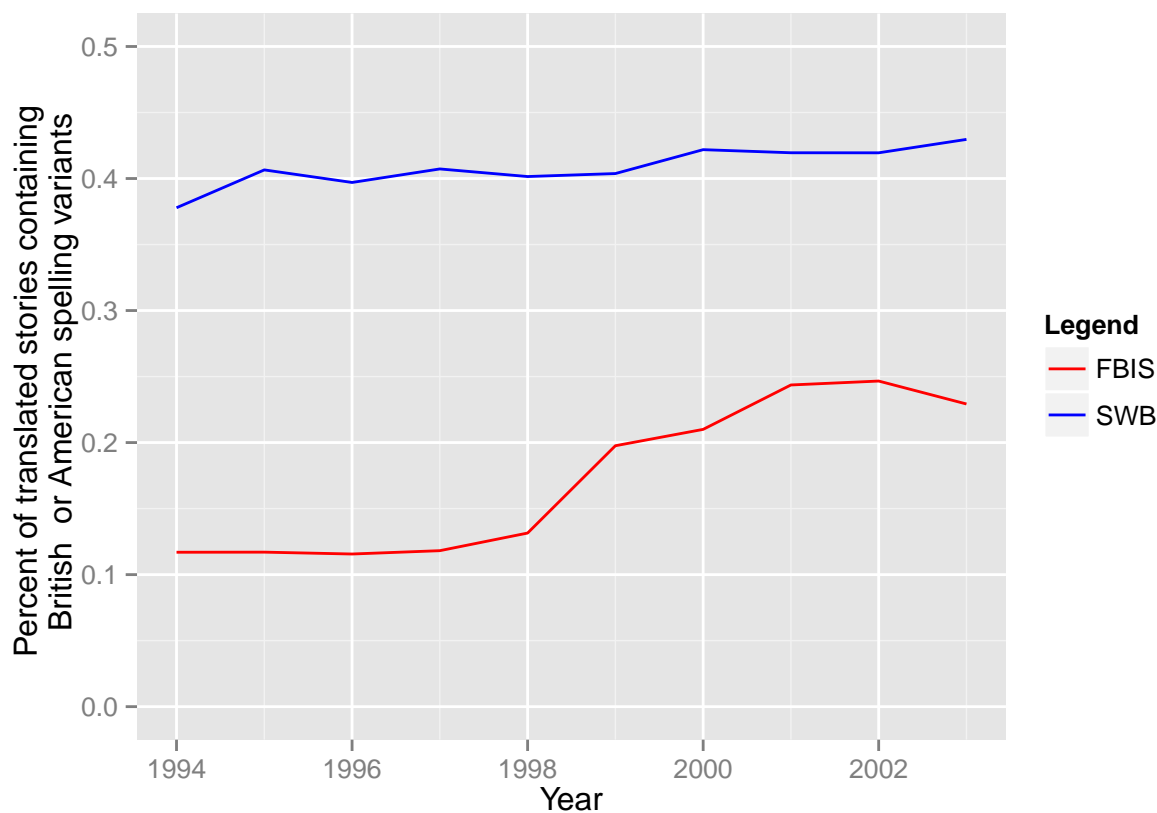
An important question for researchers using these data sets is whether each provides a fully global picture of news coverage, or whether they were intended to be used together to provide that fuller picture. Our interview subjects confirmed that SWB and FBIS/OSC have a rough division of labor in monitoring geography to avoid duplicating one another's efforts. What we don't know is whether data-sharing arrangements serve to "fill the gaps" in each data set.

Understanding the degree of content overlap between these sources can help scholars determine how to use them in their own analysis. If most of the content in SWB is reproduced in FBIS, then maybe there is little need to study both. If instead SWB and FBIS/OSC provide completely different coverage, the two sources may be better used in conjunction as supplements to each other. Comparing the full text of every item in FBIS/OSC to SWB is beyond the scope of this paper. Instead, we provide an initial estimate of the degree of overlap between these two sources by searching translated material for common words that are spelled differently in British and American English. We know that SWB and FBIS/OSC share content, and we assume that when they do, they are unlikely to take the time to change the spelling of labor to labour or colour to color. Using a list of close to 1,800 root words and their derivatives in which American and British spelling differs, we searched the full text of every item in SWB and FBIS/OSC from 1994 to 2003.⁴¹ Since roughly half the coverage in SWB and FBIS/OSC comes from English language sources which may use either British or American spellings, we conducted our analysis on the sample of translated texts only. This narrows our focus to only those news reports that were translated by one organization or the other, rather than including news reports that originally appeared in English with British or Americanized spelling. The results are displayed in Figure 4 with the full counts and proportions provided in the appendix. We find that between 37 to 42 percent of the translated content in SWB contains Americanized spellings, while between 11 and 25 percent of FBIS/OSC translated content contains at least one British spelling. These estimates likely provide an upper bound to the degree of overlap between the two sources, suggesting that both SWB and FBIS/OSC contain relatively large amounts of translated content originating with the other organization. This indicates that

⁴¹The list of "shibboleth" words is available here:<http://www.tysto.com/uk-us-spelling-list.html>

these data sources are intended to be stand-alone resources for understanding global news content, rather than as partial collections that first must be joined together to gain an accurate picture of the whole.⁴²

Figure 4: Estimating the potential overlap in FBIS and SWB using common variants in British and American spelling



Note: The red line shows the proportion of stories in FBIS that contained British spellings. The blue line shows proportion of stories in SWB that contained Americanized spellings

⁴²Given FBIS/OSC's general preference for full-text translations, it also seems possible that what excerpted content that exists in FBIS/OSC may have originated in SWB. Similar searches suggested about 8 to 10 percent of the content in FBIS/OSC from 1994 to 2003 contained the excerpt indicators used by SWB. Since FBIS/OSC is likely to draw on both full text and excerpted content from SWB, these estimates may provide a lower bound for the degree of overlap in FBIS/OSC.

CONCLUSION

The global media content of SWB and FBIS/OSC represent a goldmine of untapped data for scholars. The primary challenge to realizing this potential comes from a lack of understanding about what these archives of open source information actually contain. Through our qualitative interviews and quantitative content analysis, we have provided some initial answers to these questions that can pave the way for use of these datasets in a wide range of research projects. Below, we offer three main conclusions from our analysis. We then address some of the limitations of our study, and conclude by offering five broad recommendations for how scholars can use SWB and FBIS/OSC in their research.

Our findings can be summarized in three key points discussed in further detail below. First, the unique structure and operation of SWB and FBIS/OSC ensures that the content they provide is a non-random, thin slice of global media. Second, this non-random collection produces a sample that is nonetheless fairly representative of global media content. Third, the deviations that do arise are relatively small in size and predictable in nature, making these data suitable for a broad range of inquiries.

All datasets contain potential biases, and SWB and FBIS/OSC are no different. While the unique structure and operation of these organizations ensures that the content they provide is a non-random sample of global news content, our analysis suggests these data are not as unrepresentative of global media as some might suspect. Although statements by the major players of the Cold War are well represented in these archives, SWB and FBIS/OSC are not simply a compilation of Politburo speeches or Central Party bulletins. Instead, these open-source information collection operations are intended to be strategic rather than tactical resources of national intelligence communities ([Leetaru, 2010](#)). The broad interests of the U.S. and British governments lead these organizations to cast a wide net in the monitoring of global media for open source information.

SWB and FBIS/OSC place a premium on obtaining content from a wide range of sources representing diverse views within a country. While designed to reflect the strategic priorities of the U.S. and British governments, the news data collected by these efforts do not appear to over-

represent news items that mention the U.S. or the U.K. However, when SWB and FBIS/OSC include items mentioning their respective countries, they tend to select content with clear evaluative tone. They emphasize front-page news and original content. As a consequence, economic and local news coverage appear somewhat less frequently in these data than in the population of news items from which they were drawn. In contrast, stories with a non-local focus, or that mention the government, elections, other domestic and/or foreign affairs, appear somewhat overrepresented in the data. While this may be a limitation for scholars focused on characterizing the overall media environment in a given country at a given point in time, the types of stories and content likely to be picked up by SWB and FBIS/OSC are the items that political scientists tend to be most interested in.

At least two limitations of our study may qualify these conclusions. First, for practical reasons our quantitative analysis was limited to the post-Cold War period and the data we presented focused only on June 2003. It is possible that some unknown biases may exist in SWB and FBIS/OSC coverage for earlier or later time periods. However the results from our qualitative interviews—some of whom started working for these organizations in the 1960s—suggests continuity rather than radical change was the norm in behavior of SWB and FBIS/OSC over time. Similarly, our analysis was limited to English-language newspapers. Broadcast coverage in SWB and FBI/OCS may be more likely to contain transcripts of speeches and summaries of live events, but again, nothing from our qualitative work suggests the general collection criteria of these organizations vary across sources.

In conclusion, we offer five broad recommendations and best practices to guide the use of SWB and FBIS/OSC in scholarly research. First and foremost, each dataset can be used as a stand-alone window on the world. While SWB and FBIS/OSC capture only portions of global news content, they do so in ways that are largely representative of the underlying population of news content. Second, if scholars wish to combine SWB and FBIS/OSC for analysis, they will need to consider how to avoid duplicate content that would then be included in both. Given the high degree of content-sharing across these organizations (especially for SWB), researchers should be cautious about combining these data sets for analysis without first removing duplicate content that originated in one but was shared with the other or independently selected for translation or inclusion by both. Third, in

terms of the type of analysis these data are best suited for, we believe the emphasis on “newsworthy items” make SWB and FBIS/OSC well-suited for many of the event based analyses associated with the study of conflict in International Relations and social movements within Comparative Politics. Fourth, when using these data for cross-national analysis, scholars should recognize the inherent variability in coverage by countries and time periods. This may require an initial validity test of the data prior to full-fledged analysis, to identify any surges or gaps in coverage. This variability may also require more complicated analysis strategies using models that explicitly account for heteroskedasticity in the data. In general, the more recent the time period, the more extensive and geographically diverse the coverage. Finally, given the relative emphasis on national and international coverage, scholars interested in either country-specific studies, subnational analysis, or topic-specific research questions should proceed with caution. SWB and FBIS/OSC might still be useful resources, but scholars may wish to supplement this data with other sources. In particular, scholars should recognize that SWB and FBIS/OSC do not contain coverage of domestic news sources from their respective countries.⁴³

For years, SWB and FBIS/OSC have provide the U.S. and British governments with a “window on the world.” Yet full potential of these data for scholarly research remains largely untapped. Having established a baseline of what is and is not in these data, we believe it is time scholars see what’s out there.

REFERENCES

Althaus, Scott L and Kalev Leetaru. 2011. “Do “We” Have a Stake in This War? A Worldwide Test of the In-Group Out-Group Hypothesis Using Open-Source Intelligence.” Paper presented at the 2011 Annual Meeting of International Studies Association, Montreal Canada.

Bean, Hamilton. 2011. *No More Secrets: Open Source Information and the Reshaping of US Intelligence: Open Source Information and the Reshaping of US Intelligence*. ABC-CLIO.

⁴³As noted above, SWB does not monitor British or American sources, while FBIS/OSC does contain content from British sources.

- Boydston, Amber E, Shaun Bevan and Herschel F Thomas III. 2014. "Policy Agendas Project."
- De Beer, Arnold and John Merrill. 2004. "Global journalism: Survey of international communication."
- Dittmer, Jason. 2015. "Everyday Diplomacy: UKUSA Intelligence Cooperation and Geopolitical Assemblages." *Annals of the Association of American Geographers* 105(3):604–619.
- Enders, Walter and Todd Sandler. 1993. "The Effectiveness of Antiterrorism Policies: A Vector-Autoregression-Intervention Analysis." *American Political Science Review* 87(04):829–844.
- Frost, Howard E. 1989. "A content analysis of recent soviet party-military relations." *American Journal of Political Science* pp. 91–135.
- Hachten, William A and James F Scotton. 2015. *The World News Prism: Digital, Social and Interactive*. John Wiley & Sons.
- Hayden, Michael V. 2008. "Remarks and Q & A by the Director of the CIA." Open Source Conference, Washington D.C.
- Hester, Joe Bob and Elizabeth Dougall. 2007. "The efficiency of constructed week sampling for content analysis of online news." *Journalism & Mass Communication Quarterly* 84(4):811–824.
- Johnson, Thomas H, Robert O Slater and Pat McGowan. 1984. "Explaining African military coups d'etat, 1960-1982." *The American Political Science Review* pp. 622–640.
- Krippendorff, Klaus. 2012. *Content analysis: An introduction to its methodology*. Sage.
- Leetaru, Kalev. 2008. "An open source study of international media coverage of the WorldCom scandal." *The Journal of International Communication* 14(2):66–86.
- Leetaru, Kalev. 2010. "The scope of FBIS and BBC open source media coverage, 1979–2008." *Studies in Intelligence* 54(1):51–71.
- Lefebvre, Stéphane. 2003. "The difficulties and dilemmas of international intelligence cooperation." *International Journal of Intelligence and Counterintelligence* 16(4):527–542.

- McCrum, Robert. 2011. *Globish: How the English language became the world's language*. Anchor Canada.
- Narduli, Peter and Michael Slana. 2013. "SPEED's Global News Archive: An Overview and Assessment." *Cline Center of Democracy, White Paper. Champaign IL* .
- Nardulli, Peter F, Buddy Peyton and Joseph Bajjalieh. 2015. "Climate Change and Civil Unrest The Impact of Rapid-onset Disasters." *Journal of Conflict Resolution* 59(2):310–335.
- Nardulli, Peter F, Scott L Althaus and Matthew Hayes. 2015. "A Progressive Supervised-learning Approach to Generating Rich Civil Strife Data." *Sociological Methodology* p. 0081175015581378.
- Olcott, Anthony. 2012. *Open source intelligence in a networked world*. Vol. 7 Bloomsbury Publishing.
- Reisner, Ann E. 1992. "The news conference: How daily newspaper editors construct the front page." *Journalism & Mass Communication Quarterly* 69(4):971–986.
- Richelson, Jeffrey T. 2015. *The US Intelligence Community*.
- Ridout, Travis N, Erika Franklin Fowler and Kathleen Searles. 2012. "Exploring the validity of electronic newspaper databases." *International Journal of Social Research Methodology* 15(6):451–466.
- Riffe, Daniel, Charles F Aust and Stephen R Lacy. 1993. "The effectiveness of random, consecutive day and constructed week sampling in newspaper content analysis." *Journalism & Mass Communication Quarterly* 70(1):133–139.
- Roop, Joseph E. 1969. Foreign Broadcast Information Service. History. Part 1: 1941-1947. Technical report DTIC Document.
- Rotheray, Brian. 2007. "A History of BBC Monitoring." *BBC Monitoring, BBC Written Archives Centre* .
- Shank, David. 1991. "Listening to the World for 50 Years." *Studies in Intelligence* 33 (Spring):19:33.

Shelef, Nadav G. 2015. "Unequal Ground: Homelands and Conflict." *International Organization* pp. 1–31.

Warner, Geoffrey. 1989. "The Anglo-American Special Relationship." *Diplomatic History* 13(4):479–499.

APPENDIX

CONTENTS

Codebook	42
Coding Categories for Story Structure	42
Story Location	42
Story Type	43
Story Origin	43
Story Completeness[For FBIS/SWB items only]	44
Evaluative Tone of Story	45
Primary Topical Focus of Story Contents	48
Primary Geographic Focus of Story Contents	62
Story Mentions the United States	64
9. Story Mentions the United Kingdom	64
Reliability Tests	66
Chi-Squared Tests for Content Analysis	66
Descriptive Tables from Content Analysis	70

CODEBOOK

CODING CATEGORIES FOR STORY STRUCTURE

STORY LOCATION

1. Primary front page of the entire paper
2. Secondary front page (of an inside section)

3. Other page

“Secondary front page” should be used when a separate section actually exists (e.g. page B1 for the business section).

STORY TYPE

1. Editorial
2. Op-Ed story inside Opinion Section
3. Labeled “News Analysis” story outside Opinion Section
4. Unlabeled “Straight News” story outside Opinion Section

Editorial stories are unsigned stories, express the opinions of the newspaper’s editorial board, and are placed in the Opinion Section of a newspaper.

Op-ed stories are bylined stories, express the opinion of the bylined author and are placed in the Opinion Section of a newspaper.

Labeled “News Analysis” stories are placed outside the Opinion Section, (e.g. next to “Straight News” articles), express the opinion or commentary of the author as opposed to reported facts and observations, and are often marked with a specific label like “News Analysis,” “Commentary,” “From the Gallery.” Interviews, question and answer columns, and write-in columns are also included in this category.

Unlabeled “Straight News” stories are placed outside the Opinion Section, and report facts and observations. This is the default category unless the story explicitly belongs in one of the other categories

STORY ORIGIN

1. Bylined story written by newspaper staff
2. Unsigned story written by newspaper staff

3. Wire service story
4. Story written by another news outlet
5. Guest author
6. Unable to tell (ambiguous origin)

Stories that combine wire service reports with the work of a newspaper staffer are coded as “Bylined story written by newspaper staff” if bylined or “Unsigned story written by newspaper staff” if unsigned. For example “By Newspaper Staff with wire services” would be coded as “Unsigned story written by newspaper staff”. “By John Smith, Reuters” would be coded as “Wire service.”

Occasionally stories will not have a separate byline but instead end the first sentence with “writes John Smith” or “...(John Smith writes).” Code these as a “Bylined story written by newspaper staff,” unless the article indicates the author is from some other source.

“Wire service stories” includes stories written by wire service agencies and syndication services such as the Associate Press (AP), Reuters, Agence France Press (AFP), Dow Jones News Wire, Xinhua, Canada Press, etc.

“Story written by another news outlet” includes stories that originally appeared in another paper such as The New York Times or the Wall Street Journal.

Guest author is used primarily for opinion pieces, with the column explicitly indicating the author holds a position outside of the newspaper’s staff. If no indication is made, code as “Bylined story written by newspaper staff.”

News briefs that give no author and no indication that they are compiled from wire services are coded as 6 “Unable to tell.”

Photo attributions are not included in the coding of story origin. For example, a news brief listed a photo that has “Reuters” listed under the source is coded under “unable to tell” unless the brief’s text is also attributed to “Reuters.”

STORY COMPLETENESS[FOR FBIS/SWB ITEMS ONLY]

1. Full text of entire story

2. Extract, excerpt or summary of full text story

EVALUATIVE TONE OF STORY

1. At least somewhat positive
2. Not even somewhat positive
3. At least somewhat negative
4. Not even somewhat negative

Evaluative tone is coded only from the story's headline, or if there is no headline (e.g. such as in the case of short news briefs), then tone is coded from the story's first sentence. Subheadings are included in the coding as part of the story's headline. If a story's headline has no verb (e.g. it is just the name of a company like "Microsoft") code the first sentence of the story for evaluative tone.

For FBIS/SWB content do not code the summary title provided by FBIS/SWB. Only code an original title denoted in the text with quotation marks. If no title is present code the first sentence.

Evaluative tone is coded separately for positive and negative tone. Not even somewhat positive or negative are the default categories.

The threshold for distinguishing between "at least somewhat" and "not even somewhat positive" is based primarily on the use of linguistic indicators, verbs and adjectives that convey an added positive or negative meaning beyond the mere occurrence of an event. Content alone, such as death or economic growth, is not enough to determine the evaluative tone of article.

A useful rule of thumb is to ask if a given headline could have been rewritten to only convey the occurrence of an event, and if so, then does the current phrasing of the headline convey a different meaning from that neutral baseline. If the phrasing does not change the interpretation of the event, then the headline fails to cross the threshold into "at least somewhat positive/negative."

Another indicator that evaluative tone may be occurring is when a headline makes relative comparisons. While this is neither necessary nor sufficient to be coded as "at least somewhat positive/negative" it is often an indicator that evaluative tone may occur.

Finally, if a key linguistic indicator can have multiple meanings or interpretations some positive or negative and some not, this does not reach the threshold for coding evaluative tone.

Examples of “at least somewhat positive:”

Below are some examples to help illustrate the threshold to be coded “at least somewhat positive:”

Government/Elites

- At least somewhat positive: “Tony Blair is no John Major. Thank heavens.”
- Not even somewhat positive “Third of Labour MPs support Blair”

The first headline might simply state “Blair is no John Major.” The “Thank heavens” provides a positive linguistic indicator for Blair’s qualifications relative to John Major. Similarly, when coding for negative tone, “Thanks Heavens” would be a negative indicator for Major.

“Third of Labour MPs support Blair” does not provide any clear indication of positive tone. A modifier like “enthusiastically support” or additional information as “...as PM with strong vision for party,” would cross the threshold of “at least somewhat positive.”

Economy

- At least somewhat positive “US consumers brimming with confidence”
- Not even somewhat positive “US confidence in economy rises”

“Brimming” is the key linguistic indicator distinguishing the first headline as “at least somewhat positive.” Additionally, standard economic terms like “bear” and “bull” markets do not, by themselves, indicate evaluative tone.

Campaigns and Elections

- At least somewhat positive “Voters overwhelmingly back Blair as next PM”
- Not even somewhat positive “55% of voters support Blair as next PM”

“Overwhelmingly” is the key linguistic indicator distinguishing these two headlines

Other Domestic Issues

- At least somewhat positive “Region ‘ripe for tourism’”
- Not even somewhat positive “Association launched to boost tourism”

“Ripe” is the key linguistic indicator distinguishing these two headlines. The tone of a quotation within a headline is treated as reflecting the overall tone of the headline.

Other Foreign Issues

- At least somewhat positive “N. Korea: Carter produces a breakthrough in crisis”
- Not even somewhat positive “Carter negotiates deal with N. Korea”

Here it is the combination of “breakthrough” and “crisis” that distinguishes the first headline from the neutral basis of the second headline.

Examples of “at least somewhat negative:”

Government/Elites

- At least somewhat negative “Blair cannot stand comparison with Attlee. And his time is running out.”
- Not even somewhat negative “Government to resist opposition pressure”

The combination of “cannot stand” (suggesting a negative comparison) and time running out (suggesting the comparison is damaging) makes this headline cross the threshold into negative tone.

Nothing in the second headline conveys any negative tone toward either the government or opposition. Without additional context or linguistic indicators, this title would be coded “not even somewhat negative”

Economy

- At least somewhat negative “Stock market plummets after dismal forecast”
- Not even somewhat negative “U.S. growth revised”

“Plummets” and “dismal” are the key linguistic indicators suggesting the first headline should be coded “At least somewhat negative.”

Nothing about “revised” in the second headline is inherently negative and thus should be coded “Not even somewhat negative.” Indeed, based on the headline alone it may have been revised upward or downward.

Campaigns and Elections

- At least somewhat negative “Why should the voters trust Blair when clearly he doesn’t trust them?”
- Not even somewhat negative “Third of Tory voters want Major to go”

The construction of the first headline ascribes a negative characteristic, untrustworthiness, to Blair. The second headline merely reports Major’s standing among Tory voters. If it ascribed motives, or reason for their dissatisfaction, it might cross the threshold.

Other Domestic Issues

- At least somewhat negative “MPs attack ‘inadequate’ pension plan”
- Not even somewhat negative “MPs debate pension plan”

Hear “attack” and “inadequate” are the key linguistic indicators that separate the first headline from the second.

Other Foreign Issues

- At least somewhat negative “Muslims gamble on prolonging conflict”
- Not even somewhat negative “Muslim factions battle”

Here “gamble” (a risky action, suggesting questionable motives) and “prolonging conflict” (a negative outcome) are the key linguistic indicators.

PRIMARY TOPICAL FOCUS OF STORY CONTENTS

1. Government/Elites
2. Economy

3. Elections
4. Other Domestic Issues
5. Other Foreign Issues
6. Miscellaneous

1. GOVERNMENT/ELITES

This section includes everything pertaining to the state and general operation of governments and the specific actions and personalities of political elites. Stories that are about specific policies and not the actors making those specific policies should be coded under the relevant topical area.

Examples:

- Government spending/budget: Government spending in general or comparison across issues
- Taxes: All national, state, and local taxes, raising and cutting taxes.
- Federalism: Relations between levels of government
- Overall state of country: General evaluations of the state of the country, where it is and is going.
- Strategies and evaluation of government officials/appointees/nominees: Elite or public reactions to members of the government without reference to re-election or to nominees for government positions. Includes approval ratings, governing strategies, confirmation hearings. For evaluations of specific policies, code under those policies (e.g. Domestic or Economic)
- Strategies and evaluation of interest groups, non-governmental organizations (NGOs): Evaluation of nongovernmental elites, like lobbyists, interest groups, and NGOs. Code international organizations like the EU and United Nations under foreign affairs.
- Political corruption: General trust in government, political scandals, bribery, ethics violations, etc. Focus on integrity of both elected and un-elected officials.

- Actions and evaluations of political parties, government agencies, bureaus, departments: Includes all actions not covered in a specific policy areas
- Health and welfare of political actors: Deaths, illnesses, speculations about the health of political leaders.
- Other: Includes any story not covered above, such as stories relating to private lives or personal histories of public officials.

Additional Coding rules:

This category includes the actions of government elites in other countries, and also at other levels (e.g. UN, EU, NATO) For example, a report on the activities of Zimbabwe's president, Robert Mugabe, in Kenya's Daily Nation would be coded under "Government/Elites."

In the case of stories about government elites and international organizations, if the primary cause of the story is the specific action of an elite (e.g. Clinton gives speech to UN) code under "Government/Elites." If the story is about the international organization as whole (e.g. NATO intervenes in Bosnia) code under "Other foreign issues." If the story mentions a meeting of foreign ministers, but the focus is really on the relations of the two countries and not the figures themselves, code as "Other foreign issues."

Stories about the general state of a political party (e.g. Labour unites behind Blair) are coded under "Government/Elites." Stories about a specific policy of a party should be coded under either "Other Domestic Issues" (e.g. Republicans propose Social Security reforms) or "Other Foreign Issues" (e.g. Democrats oppose war in Iraq). If the story discusses the fate of party or party's policy with regard to a specific election, code under "Campaigns and Elections."

Stories about taxes and budget in general (e.g. the specific budget for a given year) are coded under "Government/Elites". Stories about funding for a specific program (e.g. Social Security) are coded under "Other Domestic Issues." Stories about the economic impact of the budget (e.g. the effects of the national debt on economic growth) are coded under "Economy." In general, unless the domestic or economic implications of story are discussed in passing (e.g. one sentence or short paragraph) code stories that mention these topics in their respective categories.

2. ECONOMY

This section includes everything on the economy, both domestic and international, from daily market reports to labor strikes to general economic indicators. Any government regulation or actions relating to one of these subcategories is coded under “Economy.”

Examples:

- Un/employment: employment and unemployment statistics; economic and social impact of un/employment; national employment priorities; employment and labor market development
- Inflation/cost of living: inflation control and reduction; anti-inflation programs; economic statistics including inflation; interest rates; and price index; consumer price index; food prices; cost of living; interest rates; reports on previous economic indicators; effects of previous economic indicators
- Banking, finance, and domestic commerce: federal regulation of the economy and banking system; financial institution regulation; regulation of securities and commodities; consumer finance, mortgages, and credit cards; insurance regulation; reform of bankruptcy laws; corporate mergers; antitrust regulation; corporate management issues; small business issues; copyrights and patents; tourism and status of tourism industry; consumer safety and fraud
- Labor issues: worker safety and protection; employment training and workforce development; employee benefits; employee relations and labor unions; strikes; wage issues; child labor
- Trade relations: trade negotiation, disputes, and agreements; imports and exports; regulation of imports/exports; restrictions on imports/exports; foreign acquisition of a country’s company/companies; domestic acquisition of foreign company/companies; foreign investments and policies aimed to increase or decrease them; productivity and competitiveness of domestic industries relative to their foreign counterparts; exchange rates and their domestic impact; international financial policies; currency manipulation and foreign exchange rates; exchange values of domestic currency; international monetary reform

- Government-business relationships: All government regulation with an economic impact; industrial policy; government programs designed to help specific industries; lawsuits between companies and government.
- Agriculture: economic conditions in agriculture; state of agriculture; agricultural of trade; inspection of agricultural imports; import trends and their impact; import or export quotas; government subsidies to agricultural producers; food inspection and safety; agricultural marketing, research, and promotion; animal/crop disease; pest control; agricultural research and development
- Industry specific news: News relating to a specific industry or companies within an industry such as telecom, steel, pharmaceuticals, etc.
- Commodities: News relating to specific commodity prices such as gold or oil.
- Other: Includes any story not covered above.

Additional Coding Rules

General tax policy is coded under “Government/Elites.” Tax policy relating to a specific industry (e.g. a subsidy for manufacturers) is coded under “Economy.”

Stories about international economic organizations such as the Group of 8 (G-8), the International Monetary Fund (IMF) and World Bank, are coded as economic. Stories about the EU with economic implications (e.g. “Leaders meet to coordinate agricultural policy”) are coded under “Economy.” If the story is primarily about the EU as a diplomatic body (e.g. the drafting of the EU constitution) code as “Other foreign issues.”

Stories about the national budget are coded under “Government/Elites.” Stories about the economic impact of issues like the national debt, are coded under “Economy.”

Agricultural stories that make no reference to the economic impact of agriculture, such as the development of a new pesticide by government researchers, should be coded as “Other Domestic Issues” because the primary cause of the story is a scientific/technological development. However,

if the story is primarily about increased production from a new pesticide, then it should be coded under “Economy” because the primary cause of the story is increased agricultural output.

Stories about the stock market or specific stocks are coded as “Economy”

Stories about tourism must refer to the economic impact of the industry. Travel stories such as “What to do on vacation in Mumbai” should be coded under “Miscellaneous.”

In general if a story can be construed to have any broader economic implications (e.g. job creation) outside the company it should be coded as “Economy”

3. CAMPAIGNS AND ELECTIONS

This section contains all articles pertaining to elections, electoral activities, and electioneering, either domestic or foreign.

Examples:

- Campaign activities: elections, votes, campaign speeches, debates, international monitoring of elections, instances of voter fraud, other campaign abuses.
- Candidate character evaluations: articles about a candidate’s character, their ability to lead, broker compromise, trustworthiness, history of corruption, past accomplishments.
- Candidate policy evaluations: where candidates stand on specific issues.
- Legislative evaluations: descriptions of particular legislative outcomes with reference to past or upcoming elections.
- Political parties: state of the party, prospects in coming elections, legislative agendas.
- Political participation of citizens/interest groups/voting blocs: political rallies, support for specific candidate or policies.
- Horse race/Ballot test: how candidates are doing in the polls, votes of no-confidence.
- Political strategy: descriptions of political and electoral strategy, targeting of specific blocks.
- Other: Any story with reference to an election not covered above

Additional Coding Rules: To be coded in “Campaign Elections” the story must reference a specific election. The primary cause of a story should relate to some electoral outcome. If a story discussing health reform makes references to both the details of the proposed policy and its electoral impact, code as “Other Domestic Issue” if the majority of the references are about the details of the policy and “Campaigns and Elections” if the majority of the references are about the electoral impact of the policy. If the article gives equal weight to both topics, code under “Campaigns and Elections.” Rallies in support of a particular candidate or party should be coded under “Campaigns and Elections.” Rallies or demonstrations against the government, or some particular policy should be coded under “Other Domestic Issues,” unless the rally is specifically protesting the results of an election, then it should be coded under “Campaigns and Elections.”

4. OTHER DOMESTIC ISSUES

This section includes articles pertaining to general domestic issues.

Examples:

- Anti-government activities: subversive activities; investigation of the activities of those identified as subversive or of protestors; internal security laws; establishing agencies to educate the public on the tactics of subversives;
- Law and order: violent crimes such as murder or rape, emerging criminal justice issues; administration of criminal justice; revision of the criminal justice system; white collar crime and organized crime; illegal drug production, trafficking, and control; court administration; prisons and prison programs; national correction standards; penal reform; juvenile crime; juvenile justice system; violence and exploitation of children; police; weapons control; criminal and civil code; crime statistics; riots and crime prevention; domestic terrorism
- Health care: health and wellness, including physical and mental health; proposals to reform broader health system; health insurance and health insurance reform, insurance availability and cost; regulation of drug industry, medical devices, and clinical labs; construction of health care facilities (hospitals, laboratories, health centers and nursing homes); regulation of and

payments to healthcare facilities; medical liability, fraud and abuse; manpower and training in the healthcare field; disease and communicable disease prevention; health promotion; long-term care; home health; care for terminally ill, and rehabilitation services; prescription drugs, coverage and costs; disease treatment; health communication; controlled and illegal drug abuse, treatment, and education; research and development related to health and healthcare

- Social welfare: poverty; homelessness; effectiveness of public welfare programs; social services proposals; public assistance programs, including financial and housing assistance for low-income families, the elderly, and the disabled; effects of economic and social deprivation; social service and volunteer associations
- Energy: energy supply and conservation; current and future energy needs; regulation of energy sources like natural gas and electricity; energy development projects; nuclear energy; electricity and hydroelectricity; natural gas and oil; coal; alternative and renewable energy, including solar, geothermal power, ethanol, biomass fuel, and wind energy; development of alternative and renewable energy technologies; construction of plants for energy providers; offshore gas and oil leasing; oil prices and demand, gasoline prices, oil shortages; long-term outlook of world oil supply; regulation, licensing, and safety for various sources of energy; energy conservation; programs designed to promote energy conservation; research and development centered on energy
- Environment: environmental issues; environmental protection; implementing environmental protection legislation; land and water conservation; water pollution; drinking water safety; drinking water availability; waste disposal; waste management; hazardous waste and toxic chemical regulation, treatment and disposal; air pollution, and noise pollution; air quality criteria; plans for reducing greenhouse emissions; ozone layer depletion; acid rain; regulation of automobile emissions; recycling; indoor environmental hazards, including lead, asbestos, and radiation; research and development centered on the environment; species and forest protection, except illegal trade in plants and animals
- Education: state of education in a country; education programs development; education

quality; national education methods; elementary, secondary education, and higher education; educational statistics like dropout rates; teaching disadvantaged students; adult literacy programs; combating illiteracy; vocational education; special education programs for disabled; promotion of excellence in education

- Immigration: immigration of refugees; refugee resettlement appropriations; legalization procedures for immigrants; visa allocations; denial of visas to political refugees; citizenship issues; migrant workers; economic and social problems related to migration; effects of immigration and migration on domestic labor conditions
- Civil rights and moral issues: civil rights violations like those concerning free speech and religion or voting rights; discrimination against women, minorities, the handicapped or persons with health conditions; civil rights enforcement; employment discrimination involving several communities (age, gender, race, sexual orientation etc. in combination); taking private property; impact on private property rights; employment discrimination due to race, ethnicity, color, and religion; policies aimed at protecting civil rights and liberties; affirmative action policies like quotas or preferential admissions to jobs or schools aimed at repressed groups or minorities
- State, provincial and local politics: local maintenance of buildings, museums, parks, landmarks, historical locations; state and local procurement and contracts; urban planning between cities; state and local surveys (water supply, street cleaning); state and local statutes and ordinances; legislative action
- Transportation: budget requests and appropriations related to transportation programs; national transportation policy, national transportation needs; adequacy of transportation system; highway and mass transit programs, high-speed ground transportation systems; mass transportation and safety; airports, airlines, air traffic control and safety; railroad transportation and safety; truck and automobile transportation and safety; maritime issues; infrastructure development; transportation related research and development

- Science and technology: budgets and appropriations concerning technology; national engineering and science policy, including infrastructure; technology transfer barriers and limitations; cooperation between national entities for technological advancements; weather forecasting and warning technologies; research and development on technologies unrelated to those captured in other categories
- Disasters: natural disasters such as floods, earthquakes, droughts, etc.
- Internal conflict, protests, and terrorism: civil war; ethnic conflict; acts by domestic terrorists organizations; rallies and demonstrations against the government
- Other: Any story referencing a domestic policy issue not covered above.

Additional Coding Rules:

Articles discussing the economic impact of domestic policies and government regulation are coded under “Economy.”

If a story discusses domestic issues that are also transnational, such as the trafficking of drugs or children, determine whether primary cause of the story is about the domestic impact or its linkage to international causes. For example a story about rising drug trafficking in Pakistan would be coded as “Other Domestic Issues” but a story that focused on the influx of drugs into Pakistan from Afghanistan would be coded as “Other Foreign Issues.”

For stories that discuss domestic issues by making international comparisons, for example a piece about the acquired immune deficiency syndrome (AIDS) rates in South Africa relative to other African nations, determine whether the primary cause of the story is AIDS in South Africa, or the broader issue of AIDS in Africa. For example, the story may focus on the actions of the South African government and use the international comparison to give baseline to assess the government, making it a domestic story. Alternatively, the article may use low or high AIDS rates to suggest that South Africa is leading or lagging behind its peers, in which case the story is more about the regional issue of AIDS and South Africa’s standing in the region. If the story gives equal weight to both aspects, code as “Other Domestic Issue.”

Environmental issues refer only to those within a country, for example water quality in Beijing. Stories about the broader environmental issues, such as global warming should be coded under “Other Foreign Issues.” However, a story about China’s specific efforts to reduce greenhouse gases would be coded under “Other Domestic Issues,” because it is the actions of one country and concerns behaviors only within that country.

Stories about science must have some broader public or policy implications (e.g. the discovery of a new drug). Science stories that are primarily human-interest stories (e.g. “Dinosaur bones discovered in back yard”) should be coded as “Miscellaneous”

Stories about immigration policy should be coded as “Other Domestic Issues,” unless the focus is on immigration as an issue that affects the relations between two governments. For example, a story about Turkish immigrants in Germany would be coded in “Other Domestic Issues” but a story about negotiations between Germany and Turkey over a new guest-worker program would be coded under “Other Foreign Issues.”

Stories about local and state taxes and budgets follow the same coding rules as outlined for “Government and Elites.” If it is a general story, coded under “Government and Elites.” If it is a policy specific story (e.g. local school taxes) code under “Other Domestic Issues.” If it is about the economic impact of this budget (e.g. the effect of local school taxes on small businesses and job creation) code under “Economy.”

For stories about religion to be coded under “Other domestic issues” they must have some broader societal or political impact.

Stories about the natural disasters are generally coded as “Other Domestic Issues,” unless the focus is on how a disaster has affected multiple countries. For example, a story about SARS in China is coded as “Other Domestic Issues,” but a story about how SARS in China has led to travel bans from other nations would be coded as “Other Foreign Issues.”

5. OTHER FOREIGN ISSUES

This section includes all stories pertaining to foreign affairs that do not have an economic focus.

Examples:

- Foreign policy: Agreements about international resource use; international law; agreements made within international organizations; diplomats, embassies, a citizens abroad, and passports; foreign visits; military aid and weapons sales to other countries; cooperation in areas like science or healthcare
- Conflict: Issues related to outbreaks of conflict between nations, with or without a declaration of war; violations or claims of violations of Geneva Conventions during conflicts, not related to civilian human rights abuses
- Defense: Defense requests and appropriations; defense alliances; national stockpiles of strategic materials; civil defense and security; civilian employment by defense industry, defense contracts and contractors; defense related research and development
- Foreign aid: Foreign aid requests and appropriations; international disaster relief; developing countries, economic development, including institutions such as the International Monetary Fund (IMF) and Agency for International Development (AID), international organizations such as the United Nations (UN), UNESCO, International Red Cross
- Human rights: Human rights abuses war crimes tribunal hearings; international human rights policy; Helsinki Accords human rights agreements; UN reports on human rights; government use of torture; human rights violations; worldwide religious persecution; crimes associated with genocide and crimes against humanity
- International terrorism: terrorist acts, including hijackings and kidnappings; security of domestic targets from terrorist attacks; impact of international terrorism; combating terrorism; political killings in foreign countries and international responses
- Military: military operations and maintenance; status of the national military establishment; military readiness; coordination of different branches of armed services; manpower, current and former military personnel; military procurement and weapons system acquisitions; military installations, construction, and land transfers; military nuclear and hazardous waste disposal, military environmental compliance; claims or settlement of claims against military

- Peacekeeping: need for peacekeeping missions; operation and maintenance of peacekeeping missions; coordination among peacekeepers; actions of peacekeepers; results of peacekeeping missions; attacks on peacekeepers
- Proliferation/non-proliferation: nuclear proliferation, including UN reports on proliferation; arms trade; nuclear exports; arms control and nuclear nonproliferation of weapons of mass destruction; destruction of nuclear weapons; nonproliferation of chemical weapons; nuclear testing moratoriums; export controls of nuclear production materials; arms export controls; arms reduction agreements; bans on chemical or biological weapons; prevention of weapons sales to foreign nations; arms control treaties and violations of these treaties; conventional forces reduction; atomic weapons research and development
- Other: Any story about foreign affairs which does not fit into the above sub-categories or has to do with a tradeoff of two or more of the above sub-categories

Additional Coding Rules:

Domestic conflicts (civil wars; ethnic violence) are coded under “Other Domestic Issues.”

Human rights abuses refer to stories about abuses in other countries. Human rights abuses within the newspaper’s country of origin should be coded as “Other Domestic Issues.”

Stories about the military’s actions within a newspaper’s country of origin, such as the use of the army as a police force, are coded as “Other Domestic Issues.” Similarly stories about peacekeeping within a newspaper’s country of origin are also coded as “Other Domestic Issues,” unless the primary cause of the story is external to the story, for example another country deciding to pull out its peacekeepers.

International terrorism refers primarily to the activities of states or terrorist organizations based outside the newspaper’s country of origin. For example, a bombing in Jerusalem by Hamas would be coded as “Other Domestic Issues” while a bombing by Hezbollah, which is based in Lebanon, would be coded as “Other Foreign Issues.”

Stories about international trade and the economies of other nations are coded under “Economy.”

In general stories about foreign aid between countries are coded under “Other foreign issues.”

Stories about development assistance from international economic organizations, like the World Bank or IMF, are coded under the “Economy.”

Stories involving economic International Economic organizations without a specific discussing a specific economic event or impact are coded under “Other Foreign Affairs.” For example, a World Trade Organization (WTO) meeting would be coded a under “Other Foreign Affairs.” A ruling by the WTO on a specific trade issue would be coded under “Economy.”

6. MISCELLANEOUS

A catch-all category for articles which do not fit into the above topics. To be used sparingly. Includes human-interest stories and pieces about celebrities or sports.

Examples:

- Human interest stories
- Sports
- Profiles of non-government actors (e.g. “Apple’s Founder Steve Jobs Dies”)

Additional Coding Rules:

In general, this should be a category of last resort, used when a story has no clear categorization or is completely devoid of hard news content. If a plausible case can be made that a story has economic implications or domestic policy implications code under those categories.

Announcements of events that have no broader policy implication (e.g. New Book Club meets) are coded as “Miscellaneous,” while events like a public health clinic would be coded “Other domestic issues.”

Stories about religion that focus primarily on individual spirituality or scriptural interpretation are coded as “Miscellaneous.” For stories about religion to be coded under “Other domestic issues” they must have some broader societal or political impact.

Brief corrections to previous stories (e.g. correcting the previous misspelling of a name) are coded as “Miscellaneous.”

HIERARCHY OF TOPICS WHEN STORIES ARE EVENLY SPLIT

If a story could be categorized under multiple categories, code under the category that most frequently applies. If a story is evenly split between two topics, use the following hierarchy to decide its categorization:

- Campaigns and Elections
- Government/Elites
- Economy
- Other Domestic Issues
- Other Foreign Issues

For example, a story split between the economy and other domestic issues should be coded under “Economy.” A story split between the economy and campaigns and elections should be coded under “Campaigns and Elections.” A story split between domestic and foreign issues should be coded under “Other Domestic Issues.”

PRIMARY GEOGRAPHIC FOCUS OF STORY CONTENTS

Primary geographic focus is determined primarily by where the story takes place. Any story that takes place outside of the newspaper’s country of origin is coded as “International.” Any story concerning International Organizations, such as the United Nations or the World Trade Organization is also coded as International. However, an the meeting of an international organization that takes place within a newspaper’s country of origin is coded as “National Domestic”

1. International
2. National domestic (pertaining to the entire country)
3. Sub-national domestic (e.g., state, local, pertaining to a particular city)

Additional Coding Rules

Any story that takes place inside a newspaper's country of origin is coded as either "National Domestic" or "Sub-national domestic." The default for these stories is "sub-national domestic." To be coded as "National domestic" a location of the story must be viewed as relevant to the country as a whole.

Stories about domestic stock and bond markets, an industry as a whole (e.g. tourism or logging), or a specific state owned industry are coded as "National domestic." Stories about specific stocks or companies or a an industry in a specific sub region (e.g. Tourism in London or logging in Western Canada) are coded as "Sub-national domestic." If a company is a multinational corporation with headquarters outside the country, code "International," unless the story is primarily about the domestic branch of that multinational corporation in the newspaper's country of origin. When a company has joint state-private ownership code as "Sub-national domestic." If the origin of a company is unclear the default is "Sub-national domestic".

To be coded as sub-national domestic, the primary focus of a story must be on a specific city or region within a country. Stories about the national government's efforts to reduce SARS in Beijing would be coded as Sub-national Domestic, whereas stories about the total incidence of SARS in China would be coded as "National Domestic." Stories about the international reaction of other countries like Japan, or international bodies, like the World Health Organization, would be coded as "International."

Stories about national leaders, such as the president or prime minister of a country are coded as "National domestic." For example, a story about President Clinton vacationing in Martha's Vineyard is coded as national domestic, because the story is about a country's president taking a vacation and not specifically about Martha's Vineyard. However, a story about President Clinton's visit to China would be coded under "International"

For the South China Morning Post in 1994, all mentions of China are coded as "International," because Hong Kong was still a British Dependent Territory. However, residedents of Hong Kong were not full British citizens so mentions of the United Kingdom (excluding its overseas colonies) are also treated as "International." Similarly dependent territories mentioned in British newspapers

(The Times of London, The Independent, The Daily Telegraph) are coded as “Intenational.”

STORY MENTIONS THE UNITED STATES

Does the story mention the U.S., American citizens, American leaders, or American-based companies?

1. Yes
2. No

9. STORY MENTIONS THE UNITED KINGDOM

Does the story mention the United Kingdom, British citizens, British leaders, or British-based companies?

1. Yes
2. No

Additional Coding Rules:

“British” excludes references to Ireland, Irish citizens, or Irish leaders

If it is ambiguous whether a specific company is based in the U.S. or U.K. the default code is “No.”

Culturally relevant figures, such as Oprah, an American talk show host, or David Beckham, a famous British soccer player, would also be coded as a U.S. or U.K. mention respectively because they speak to the cultural relevance of these countries abroad.

Mentions of the Federal Reserve and Bank of England are also included as “Yes’s” under this category.

Mentions of the Anglican Church are coded as a mention of the U.K.

Stories that simply list a countries currency and give the U.S. or British equivalent in dollars (e.g. The price of milk rises to 1 rupee (\$1.20 USD) are not coded as mentioning the U.S. or Britain, but stories about the U.S. dollar as a global currency would be coded as mentioning the U.S.

Since Hong Kong was still a British Dependent Territory in 1994, The South China Morning Post is coded as if it were a British Paper, like The Times of London, in terms of UK mentions.

For British papers, and Hong Kong's South China Morning Post in 1994, that talk about domestic issues, this is treated as an implicit reference and is coded as a "Yes" for "Story mentions the UK."

Below are lists of the potentially relevant leaders for the United States and United Kingdom for the years of 1994 and 2003:

Table 7: Major Figures in the U.S.

United States	1994	2003
Major Leaders		
President	Bill Clinton	George W. Bush
Secretary of State	Warren Christopher	Colin Powell
Secretary of Defense	Leslie Aspin Jr.	Donald Rumsfeld
Speaker of the House	Tom Foley	Dennis Hastert
Other figures	OJ Simpson Jesse Jackson	Jayson Blair

Table 8: Major Figures in the U.K.

United Kingdom	1994	2003
Major Leaders		
Prime Minister	John Major	Tony Blair
Secretary of State for Foreign Affairs	Douglas Hurd	Jack Straw
Secretary of State for Defence	Malcolm Rifkind	Geoff Hoon
Other Figures	Members of the Royal Family	Members of the Royal Family

RELIABILITY TESTS

Table 9: Reliability Coding Results

	Percent Agreement	Coefficient S	Scott's Pi	Cohen's Kappa	Krippendorff's Alpha
Story Location	99%	0.99	0.91	0.91	0.91
Story type Type	99%	0.99	0.86	0.86	0.86
Story Origin	94%	0.93	0.92	0.92	0.92
Positive Evaluative Tone	97.50%	0.95	0.69	0.69	0.69
Negative Evaluative Tone	97%	0.94	0.69	0.69	0.69
Primary Topical Focus	88%	0.86	0.84	0.84	0.84
Primary Geographic Focus	88%	0.82	0.82	0.82	0.82
Mentions U.S.	98%	0.96	0.95	0.95	0.95
Mentions U.K.	98.50%	0.97	0.95	0.95	0.95

Note:Results from coding a sample of 200 articles from 10 papers in 1994 and 2003

CHI-SQUARED TESTS FOR CONTENT ANALYSIS

Table 10: Chi-Squared and Fisher-Exact Tests for 1994 vs 2003 Constructed Week Sample

	χ^2	p-value	Fisher Test p-value
Story Location	31.96	0.00	0.00
Story Type	19.34	0.00	0.00
Story Origin	255.47	0.00	0.00
Story Completeness	0.65	0.42	0.31
Positive Tone	5.82	0.02	0.01
Negative Tone	0.26	0.61	0.60
Evaluative Tone	7.87	0.05	0.03
Primary Topic	207.30	0.00	0.00
Primary Geographic Focus	120.61	0.00	0.00
U.S. Mentioned	6.42	0.01	0.01
U.K. Mentioned	57.96	0.00	0.00

Table 11: Chi-Squared and Fisher-Exact Tests for 1994 vs 2003 Constructed Week Sample

	χ^2	p-value	Fisher Test p-value
Story Location			0.00
Story Type	19.34	0.00	0.00
Story Origin	255.47	0.00	0.00
Story Completeness	0.65	0.42	0.31
Positive Tone	5.82	0.02	0.01
Negative Tone	0.26	0.61	0.60
Evaluative Tone	7.87	0.05	0.03
Primary Topic	207.30	0.00	0.00
Primary Geographic Focus	120.61	0.00	0.00
U.S. Mentioned	6.42	0.01	0.01
U.K. Mentioned	57.96	0.00	0.00

Table 12: Chi-Squared and Fisher-Exact Tests for FBIS vs. SWB

	χ^2	p-value	Fisher Test p-value
Story Location	61.95	0.00	0.00
Story Type	10.38	0.02	0.00
Story Origin	291.25	0.00	0.00
Story Completeness	228.26	0.00	0.00
Positive Tone	1.21	0.27	0.19
Negative Tone	0.79	0.37	0.32
Evaluative Tone	3.38	0.34	0.24
Primary Topic	40.45	0.00	0.00
Primary Geographic Focus	6.68	0.04	0.03
U.S. Mentioned	2.20	0.14	0.13
U.K. Mentioned	1.88	0.17	0.15

Table 13: Chi-Squared and Fisher-Exact Tests for FBIS-94 vs. FBIS-03

	χ^2	p-value	Fisher Test p-value
Story Location	1402.40	0.00	0.00
Story Type	31.23	0.00	0.00
Story Origin	25.88	0.00	0.00
Story Completeness	6.77	0.01	0.01
Positive Tone	1.07	0.30	0.24
Negative Tone	18.88	0.00	0.00
Evaluative Tone	20.80	0.00	0.00
Primary Topic	77.83	0.00	0.00
Primary Geographic Focus	39.14	0.00	0.00
U.S. Mentioned	21.63	0.00	0.00
U.K. Mentioned	8.78	0.00	0.00

Table 14: Chi-Squared and Fisher-Exact Tests for FBIS-03 vs. SWB

	χ^2	p-value	Fisher Test p-value
Story Location	7.11	0.03	0.01
Story Type	14.05	0.00	0.00
Story Origin	277.17	0.00	0.00
Story Completeness	187.34	0.00	0.00
Positive Tone	1.72	0.19	0.17
Negative Tone	0.01	0.91	0.84
Evaluative Tone	2.65	0.45	0.29
Primary Topic	88.90	0.00	0.00
Primary Geographic Focus	10.46	0.01	0.01
U.S. Mentioned	4.97	0.03	0.02
U.K. Mentioned	0.72	0.40	0.40

Table 15: Chi-Squared and Fisher-Exact Tests for FBIS vs. Population

	χ^2	p-value	Fisher Test p-value
Story Location	9136.40	0.00	0.00
Story Type	305.66	0.00	0.00
Story Origin	600.86	0.00	0.00
Story Completeness	598.27	0.00	0.00
Positive Tone	6.24	0.01	0.01
Negative Tone	109.59	0.00	0.00
Evaluative Tone	123.22	0.00	0.00
Primary Topic	371.50	0.00	0.00
Primary Geographic Focus	272.79	0.00	0.00
U.S. Mentioned	11.82	0.00	0.00
U.K. Mentioned	33.93	0.00	0.00

Table 16: Chi-Squared and Fisher-Exact Tests for FBIS-94 vs. Population

	χ^2	p-value	Fisher Test p-value
Story Location	184.94	0.00	0.00
Story Type	45.85	0.00	0.00
Story Origin	187.15	0.00	0.00
Story Completeness	57.41	0.00	0.00
Positive Tone	2.28	0.13	0.10
Negative Tone	0.00	1.00	1.00
Evaluative Tone	3.02	0.39	0.28
Primary Topic	51.88	0.00	0.00
Primary Geographic Focus	97.76	0.00	0.00
U.S. Mentioned	4.04	0.04	0.04
U.K. Mentioned	3.59	0.06	0.06

Table 17: Chi-Squared and Fisher-Exact Tests for FBIS-03 vs. Population

	χ^2	p-value	Fisher Test p-value
Story Location	7810.73	0.00	0.00
Story Type	243.63	0.00	0.00
Story Origin	404.98	0.00	0.00
Story Completeness	469.89	0.00	0.00
Positive Tone	5.26	0.02	0.02
Negative Tone	122.89	0.00	0.00
Evaluative Tone	137.94	0.00	0.00
Primary Topic	273.38	0.00	0.00
Primary Geographic Focus	178.58	0.00	0.00
U.S. Mentioned	18.96	0.00	0.00
U.K. Mentioned	21.04	0.00	0.00

Table 18: Chi-Squared and Fisher-Exact Tests for SWB vs. Population

	χ^2	p-value	Fisher Test p-value
Story Location	7031.00	0.00	0.00
Story Type	52.00	0.00	0.00
Story Origin	488.10	0.00	0.00
Story Completeness	2526.55	0.00	0.00
Positive Tone	8.14	0.00	0.01
Negative Tone	30.61	0.00	0.00
Evaluative Tone	68.16	0.00	0.00
Primary Topic	126.29	0.00	0.00
Primary Geographic Focus	88.05	0.00	0.00
U.S. Mentioned	0.26	0.61	0.60
U.K. Mentioned	8.21	0.00	0.00

DESCRIPTIVE TABLES FROM CONTENT ANALYSIS

Table 19: Counts and Proportions for Story Location

	Population		FBIS		FBIS-94		FBIS-03		SWB	
Front Page	466	(0.04)	76	(0.05)	75	(0.24)	1	(0)	0	(0)
Section Front	244	(0.02)	17	(0.01)	17	(0.05)	0	(0)	0	(0)
Other	9735	(0.93)	266	(0.16)	227	(0.71)	39	(0.03)	0	(0)
Online	0	(0)	1260	(0.78)	0	(0)	1260	(0.97)	225	(1)

Table 20: Counts and Proportions for Story Type

	Population		FBIS		FBIS-94		FBIS-03		SWB	
Editorial	161	(0.02)	142	(0.09)	17	(0.05)	125	(0.1)	18	(0.08)
Op-Ed	356	(0.03)	68	(0.04)	2	(0.01)	66	(0.05)	0	(0)
Analysis	806	(0.08)	130	(0.08)	13	(0.04)	117	(0.09)	17	(0.08)
News	9122	(0.87)	1279	(0.79)	287	(0.9)	992	(0.76)	190	(0.84)

Table 21: Counts and Proportions for Story Origin

	Population		FBIS		FBIS-94		FBIS-03		SWB	
Byline	4237	(0.41)	1008	(0.62)	225	(0.71)	783	(0.6)	35	(0.16)
Staff	1105	(0.11)	211	(0.13)	20	(0.06)	191	(0.15)	23	(0.1)
Wire	2657	(0.25)	21	(0.01)	0	(0)	21	(0.02)	0	(0)
Other Paper	299	(0.03)	0	(0)	0	(0)	0	(0)	0	(0)
Guest	226	(0.02)	70	(0.04)	9	(0.03)	61	(0.05)	7	(0.03)
Ambiguous	1921	(0.18)	309	(0.19)	65	(0.2)	244	(0.19)	160	(0.71)

Table 22: Counts and Proportions for Story Completeness

	Population		FBIS		FBIS-94		FBIS-03		SWB	
Complete	10428	(1)	1507	(0.93)	308	(0.97)	1199	(0.92)	133	(0.59)
Excerpt	17	(0)	112	(0.07)	11	(0.03)	101	(0.08)	92	(0.41)

Table 23: Counts and Proportions for Positive Tone

	Population		FBIS		FBIS-94		FBIS-03		SWB	
Positive	259	(0.02)	58	(0.04)	15	(0.05)	43	(0.03)	12	(0.05)
Neutral	10186	(0.98)	1561	(0.96)	304	(0.95)	1257	(0.97)	213	(0.95)

Table 24: Counts and Proportions for Negative Tone

	Population		FBIS		FBIS-94		FBIS-03		SWB	
Negative	714	(0.07)	233	(0.14)	21	(0.07)	212	(0.16)	38	(0.17)
Neutral	9731	(0.93)	1386	(0.86)	298	(0.93)	1088	(0.84)	187	(0.83)

Table 25: Counts and Proportions for Evaluative Tone

	Population		FBIS		FBIS-94		FBIS-03		SWB	
Positive	258	(0.02)	56	(0.03)	15	(0.05)	41	(0.03)	11	(0.05)
Negative	713	(0.07)	231	(0.14)	21	(0.07)	210	(0.16)	37	(0.16)
Mixed	1	(0)	2	(0)	0	(0)	2	(0)	1	(0)
Neither	9473	(0.91)	1330	(0.82)	283	(0.89)	1047	(0.81)	176	(0.78)

Table 26: Counts and Proportions for Primary Topic

	Population		FBIS		FBIS-94		FBIS-03		SWB	
Government	1242	(0.12)	236	(0.15)	37	(0.12)	199	(0.15)	57	(0.25)
Economic	3746	(0.36)	416	(0.26)	141	(0.44)	275	(0.21)	20	(0.09)
Elections	184	(0.02)	25	(0.02)	8	(0.03)	17	(0.01)	4	(0.02)
Domestic	3332	(0.32)	471	(0.29)	74	(0.23)	397	(0.31)	65	(0.29)
Foreign	1380	(0.13)	471	(0.29)	59	(0.18)	412	(0.32)	79	(0.35)
Misc	561	(0.05)	0	(0)	0	(0)	0	(0)	0	(0)

Table 27: Counts and Proportions for Primary Geographic Focus

	Population		FBIS		FBIS-94		FBIS-03		SWB	
International	4671	(0.45)	522	(0.32)	134	(0.42)	388	(0.3)	69	(0.31)
National	2242	(0.21)	651	(0.4)	140	(0.44)	511	(0.39)	109	(0.48)
Sub-National	3532	(0.34)	446	(0.28)	45	(0.14)	401	(0.31)	47	(0.21)

Table 28: Counts and Proportions for U.S. Mentioned

	Population		FBIS		FBIS-94		FBIS-03		SWB	
US Mentioned	2980	(0.29)	530	(0.33)	69	(0.22)	461	(0.35)	62	(0.28)
US not Mentioned	7465	(0.71)	1089	(0.67)	250	(0.78)	839	(0.65)	163	(0.72)

Table 29: Counts and Proportions for U.K. Mentioned

	Population		FBIS		FBIS-94		FBIS-03		SWB	
UK Mentioned	2764	(0.26)	318	(0.2)	82	(0.26)	236	(0.18)	35	(0.16)
UK not Mentioned	7681	(0.74)	1301	(0.8)	237	(0.74)	1064	(0.82)	190	(0.84)