4-18-2011

Bias in Hard News Articles from Fox News and MSNBC: An Empirical Assessment Using the Gramulator

Brock Ashton Terwilleger

To the University Council:

The Thesis Committee for Brock Terwilleger certifies that this is the final approved version of the following thesis: "Bias in Hard News Articles from Fox News and MSNBC: An Empirical Assessment Using the Gramulator."

_____

Philip M. McCarthy, Ph.D. Major Professor

We have read this thesis and recommend its acceptance:

_____

Charles E. Hall, Ph.D.

_____

Reginald Martin, Ph.D.

Accepted for the Graduate Council:

_____

Karen D. Weddle-West, Ph.D.
Vice Provost for Graduate Program

BIAS IN HARD NEWS ARTICLES FROM FOX NEWS AND MSNBC:
AN EMPIRICAL ASSESSMENT USING THE GRAMULATOR

By

Brock Terwilleger

A Thesis

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Arts

Major: English

The University of Memphis

May 2011

Abstract

Terwilleger, Brock Ashton. MA. The University of Memphis. May 2011. "Bias in Hard News Articles from Fox News and MSNBC: An Empirical Assessment Using the Gramulator." Major Professor: Philip M. McCarthy.

Hard news articles can reflect a media organization's bias. This study assesses bias in the hard news articles published by Fox News and MSNBC. More specifically, the study examines bias in the two networks' coverage of elections and immigration. Indicative linguistic features identified by the Gramulator reveal biases in corpora of texts pertaining to these two news topics. The results suggest that the networks exercise selection bias. Furthermore, the findings show that the networks use the same collocations differently and also use different collocations to refer to the same referents.

Table of Contents

List of Tables

Introduction

My study examines bias in articles found on the websites of cable news networks. More specifically, this study focuses on the linguistic features that distinguish the coverage of news topics on Fox News and MSNBC. My research question is, "Do the different linguistic features in online hard news articles reflect common views of Fox and MSNBC's biases?" If so, where and how do the corpora refer to topics associated with the networks' ideologies? My primary hypothesis is that the corpora will reflect conservative and liberal biases respectively, because Fox News and MSNBC take these positions in their televised content.

Overall, the goal of this study is to identify differences in lexical choices and story selection that manifest a network's bias in its hard news online. Finding bias in electronic hard news will provide another angle from which to assess the biases of these networks, strengthening the case for institutional bias at Fox News and MSNBC.

Literature Review

In this study I combine two linguistic approaches to analyzing biased language in news articles: *computational corpus linguistics* and *discourse analysis*. The study uses computational tools to identify key linguistic features within contrastive corpora. Using discourse analysis, I interpret how these features reflect news discourse in general and how they amount to bias within the specific discourses of Fox News and MSNBC.

Computational Corpus Linguistics

In computational corpus linguistics, researchers use software to study language throughout groups of texts (McCarthy & Boonthum, in press). *Corpus* means *body* in Latin, and in corpus linguistics texts are grouped into a body of texts. Researchers are

concerned with language use in the body of texts as a whole, not with the language of an individual text (Gries, 2009). Software can analyze the language in a group of texts and identify key linguistic features and patterns in the corpus overall. Gries explains that such software tools describe linguistic patterns in a way that is quantifiable and highly generalizable to the corpus. Given the goals of the study, researchers then interpret how these descriptions apply to their hypotheses and theoretical frameworks.

Discourse Analysis

Scholars have differing definitions of discourse, and therefore differing approaches to discourse analysis. In this study, I draw on the scholarship of Johnstone (2008), Kent and McCarthy (in press), and van Dijk (1988). Johnstone explains the general approach of discourse analysis. van Dijk gives an overview of news discourse specifically, including methods for analysis. And Kent and McCarthy discuss the role of computational language assessment tools in discourse analysis.

Johnstone (2008) explains the origins and persistence of discourse. To her, discourse consists of the conventions of communicating in a certain situation. And those conventions are based on past experiences of communicating in similar situations. These conventions are reproduced because they provide people the background knowledge necessary to interpret and engage in a communicative situation. Mutual expectations facilitate the exchange of information, because the participants waste no energy negotiating the form of their exchange. Instead, the form is predetermined, allowing the participants to focus on the new information being exchanged.

According to van Dijk (1988), discourse is contradictory; it is both systematic and irregular. In other words, discourse follows patterns, but there is an ample amount of space for variation within those patterns. For example, news discourse is characterized by certain patterns of language use that differentiate it from scientific discourse. Yet, individual pieces of discourse are always embedded in a particular context that make them unique. For example, a news article about a candidate for the city council of Memphis, TN in 2011 would be different from an article about a candidate for the state Senate of Oregon in 1910. The details of the two articles would differ in terms of subject, place, and time. However, the two would exhibit many of the same characteristics of discourse in American print news, such as organization and style. Therefore, news discourse follows universal patterns and varies according to space and time.

For Johnstone (2008), the objective of discourse analysis is to answer a few global questions. Those questions are *What is in this text and why? What is not in this text and why not? Why is this information presented in this way and not another?* To answer these questions, analysts must consider the author, purpose, audience, medium, genre, and context of a text.

Discourse analysis requires researchers to isolate pieces of discourse for examination. Isolating examples means they are removed from their real world context. Johnstone (2008) acknowledges that extracting examples of discourse from their context compromises their integrity. Nevertheless, she concedes that this step is a necessary part of discourse analysis.

Description is the foundation of the discourse analysis process. Johnstone (2008) notes that this approach assumes that descriptions of texts can be valid. And van Dijk

(1988) emphasizes that the challenge is to provide descriptions that are *explicit* and *systematic*. There are two types of descriptions: textual and contextual. Textual descriptions pertain to linguistic structures within the text, whereas contextual descriptions link structures within the text to factors outside the text.

In my study, I address textual and contextual factors. I use computational tools to provide textual descriptions, and I group texts into corpora according to their context. I use the textual analysis tool the Gramulator, which Kent and McCarthy (in press) argue describes discourse by segmenting texts into collocations. This method adheres to van Dijk's prescription that textual descriptions be systematic and explicit. That is, the Gramualtor *systematically* extracts the distinctive collocations from a corpus, thereby making underlying patterns in that corpus *explicit*. In addition, the texts are grouped into corpora according to contextual factors, i.e., the current events that the texts address, such as elections and immigration policy. Furthermore, I group texts by network, and the networks are the major contextual factors that I hypothesize will influence bias. In this way, my study addresses the two levels of discourse, textual and contextual, by making the linguistic phenomena of bias explicit through systematic analysis.

When analyzing textual elements, it is important to consider semantics. More specifically, van Dijk (1988) holds that a complete discourse analysis addresses both the meaning of language and the reality that language represents. Language cannot accurately describe reality in its entirety. Therefore, language requires interpretation. Discourse analysis is concerned with the way language leads to incomplete or inaccurate interpretations of reality. My study focuses on this disconnect between reality and its textual representation as an opportunity for bias. In other words, this study will assess

what textual and contextual factors in the corpora encourage readers to interpret texts according to Fox News and MSNBC's biases.

Discourse analysis deals with semantics and syntax, but van Dijk (1988) calls for analysts to consider two concepts he terms *macrosemantics* and *macrosyntax*. Macrosemantics goes beyond the microsemantics of a word or phrase in a sentence. Instead, macrosyntax concerns the overall meaning conveyed over sections of a text. Likewise, macrosyntax reaches beyond the microsyntax of the arrangement of words in a sentence. Instead, macrostynax pertains to the patterns typical of a discourse.

In my study, bias falls in the domain of macrosemantics, and the language of bias falls in the domain of macrosyntax. Bias can be one of several macrosemantic messages in a text. For example, an article about health care legislation may have the overt macrosemantic message that the bill is moving closer to a vote. At the same time, the article may have the underlying macrosemantic message that the bill is good or bad. Similarly, bias may be expressed systematically throughout a text in similar ways, amounting to a form of macrosyntax. For example, the repeated use of the phrase *health care reform* may be a macrosyntactic pattern indicating the macrosemantic message that health care legislation is positive. In contrast, the repeated use of the phrase *health care overhaul* may be a macrosyntactic pattern indicating the macrosemantic message that health care legislation is too dramatic. Bias lies in the different macrosemantic messages created by the differing uses of *health care reform* and *health care overhaul*.

At the sentence level, words and word order, i.e. semantics and syntax, combine to create a *proposition*. By proposition, van Dijk means a statement or an idea. Multiple propositions combine to form a *macroproposition*, or the big idea of the text. In the

hypothetical example above, the macroproposition of the article is that the piece of health care legislation will soon come to a vote. And in my study, the bias of the article, i.e., whether the legislation is good or bad, will be also considered a macroproposition.

As noted by van Dijk (1988). macropropositions can be subjective, because they must be interpreted by the reader. The author may emphasize certain parts of a text, but ultimately the macroproposition depends on the reader. This equation leads to a different result for each reader, because readers focus on different parts of the text. This study focuses on the author's role in forming macropropositions as an opportunity for the manifestation of bias.

Overall, my study builds on Johnstone's (2008) definition of discourse as a reproduced pattern of communication expected by author and audience. This study also builds on van Dijk's observation that discourse is both consistent and irregular. Both researchers agree that the objective of discourse analysis is description. This study uses computational tools to provide descriptions of discourse, as explained by Kent and McCarthy (in press). With these descriptions, this study aims to identify patterns in Fox and MSNBC's depictions of news events. Descriptions make explicit the linguistic phenomena aimed at elliciting macropropositions in the reader that coincide with the networks' ideological stances, revealing bias.

News Language

News media covers multiple types of discourses. For example, televised evening newscasts, daytime political talk shows, NPR stories, and op-ed articles are all variations of news discourse. This study is concerned specifically with the discourse of hard news articles found on the Internet. More specifically, this study assesses the presence of bias

in news discourse by comparing the linguistic features that differentiate corpora. Therefore, I will focus less on the structure of news articles and more on their language. Bell's *The Language of News Media* (1991) and van Dijk's *News as Discourse* (1988) both provide insights into the language of written news discourse.

      Quotes. Quotes are a prevalent feature of news discourse. As Bell (1991) notes, quotes are often the news item itself. That is, a comment made by a politician, for example, may amount to a news event and be the subject of an article. Used in other ways, quotes illustrate or support the points of articles. According to Bell (1991), direct quotes have an air of authority and objectivity. They seem objective because they are pure data. Therefore they cannot be challenged, and they strengthen the journalist's message with real world support.

      Quotes introduce uncharacteristic elements into news discourse. For example, hard news articles ideally attempt to be objective. However, both researchers highlight that quotes provide journalists opportunities to present opinions and different perspectives into news articles. And as van Dijk (1988) notes, the journalists may choose quotes to fulfill their own agendas. In addition to adding opinion, direct quotes can also incorporate linguistic features that are uncharacteristic of news discourse, such as first person and vernacular speech, into news articles.

      Bell (1991) further examines the different types of quotes and their usage. There are three types: direct quotes, indirect quotes, and semi-direct quotes. Direct quotes are given verbatim in quotation marks; indirect quotes are paraphrased and attributed to the speaker; and semi-direct quotes consist of parts of an utterance quoted verbatim and embedded in a sentence. Bell explains that the credibility and status of a speaker dictates

11

how they are quoted. The higher a speaker's status, the higher the chances that the utterance will be directly quoted in its entirety. For example, the whole transcript of a presidential speech may be published as an article online. However, a journalist may directly quote only one sentence from a senator's response to the speech, and may only paraphrase the comments made by a scholar of presidential speeches.

The speech verb that contextualizes a quote also influences how readers interpret it. Bell (1991) explains that the most common speech verbs such as *say* and *tell* are more neutral, thereby giving the statements quoted more legitimacy. On the other hand, more connotative words like *claim* call into question the credibility of the statement and its source, according to Bell.

Sources. The quality of a journalist's sources influences the credibility of the article. Therefore, it is in the best interest of the journalist to have authoritative sources. Bell (1991) notes that the status of a source affects the status of the news story. For example, even mundane cables released by Wikileaks became newsworthy because their sources were high ranking officials in the State Department. Furthermore, politicians and officials are generally disproportionately quoted, even when they are not experts on an issue. Another often-quoted source is professors, who are generally invoked to clarify situations or validate facts.

No matter the status of the source, citing that source in the respected genre of a news article gives credibility to that source. And given the hierarchy of quotes explained above, a direct quote can boost the credibility of a source. Directly quoting a questionable or low-status source may give that person's opinion more credibility in the eyes of the reader.

Figures. In news discourse, numbers are used to support claims. Bell (1991) suggests that figures also increase the seeming objectivity of a piece. He argues that they are sometimes unnecessary. In addition, readers are rarely informed enough to evaluate the numbers. Without sufficient background knowledge, readers cannot tell if numbers are relatively big or small, if they are significant or not. Therefore, Bell concludes that the purpose of using numbers is not to support claims but to enhance the newsworthiness of an event, exaggerate an event's consequences, or vividly illustrate a situation. Furthermore, the way numbers are presented can also have a rhetorical effect. For example, *1,000,000,000* is more dramatic than *1 billion*.

Time and Place. Setting is a basic element of news articles. However, setting is not the focus of the article, therefore time and place are only explained briefly (Bell 1991). Time is important in establishing the relationship between news events. In articles, journalists must often relate events to each other in a small amount of space using words such as *after* and *meanwhile*. Location is important in establishing the relationship between the event and readers. More importantly for this study, mentioning a place can elicit associations between a place, its people, and its history. Bell refers to the suggestiveness of a place as its "consonance with stereotypes" (p. 198).

In this study, the goal is to describe and analyze the language of news media in order to assess bias. Therefore, the characteristic features of news language are central to this study. It is important to be aware of these key features, because many are present in the results generated by the computational linguistics tools I use to process my corpora.

Bias in News Media

In this project, the goal of analyzing news discourse is to find linguistic features in the corpora that manifest Fox and MSNBC's biases. Therefore, this study builds on a tradition of assessing news outlets' biases through corpus-based approaches. Several studies have examined the biases of Fox and MSNBC using corpora of transcripts from their televised content. However, few appear to have studied their hard news print articles. Addressing this deficit in the literature is this study's contribution to the field of bias assessment.

Corpus approaches to assessing news bias. Examining bias in hard news articles with corpora is an established approach. Corpus-based approaches are appropriate for studying bias, because they allow researchers to make generalizable assessments of a news outlet's bias. Of course, the ways researchers identify bias within corpora differs greatly.

Some researchers analyze the quantifiable information given in texts. For example, Ansolabehere, Snowberg, and Snyder (2005) analyzed the dollar amounts of campaign contributions in articles to assess which candidates the newspapers sympathized with. Analyzing quantifiable data allows researchers to assess bias without making subjective judgments.

Other researchers choose to code the content of articles in a corpus. For example, Bruno (2009) studied articles about labor relations in the *Chicago Tribune* by breaking them into six variables, such as the lead and the content frame, and identified the variables as positively, negatively, or neutrally slanted. Hayes (2008) examined the portrayal of presidential candidates in terms of issue-ownership by coding articles

14

according to tone, subject of the article, and mentions of issues associated with the candidate's party. Focusing on paragraphs instead of whole articles, Niven (2003) coded paragraphs as positive or negative in articles about party-switching Congressmen. Similarly, Schiffer (2006) analyzes the tone of articles at the paragraph level in his study of Senate election coverage. Focusing on specific words instead of paragraphs, Chen (2005) codes the transitive verbs in articles as depicting something negatively, positively, or neutrally.

Still other researchers use computational content-analysis tools to assess bias in news articles. For example, Clark (2007) uses Wordsmith Tools to analyze the bias of BBC articles on the Iraq war written by reporters embedded with troops on the ground. Similarly, Baker (2010) uses Wordsmith 4 to analyze the vocabulary of articles in British tabloids to examine bias in the depictions of Muslims. By using computation content-analysis tools, researchers were able to identify patterns in word usage across large numbers of texts and make generalizable claims about the biases of certain publications.

Similar methods have been used to examine bias in the televised content from Fox News and MSNBC. Some researchers have mined televised broadcasts for quantifiable data such as statistics. Others have coded segments from shows. And others have used computational content-analysis tools to analyze transcripts of televised programs.

Like Ansolabehere et al. (2005), Groeling (2008) also found a way to quantifiably identify bias by comparing statistics reported in stories. But in Groeling's study, bias was measured by the approval ratings attributed to a presidential candidate on ABC, CBS, NBC, and Fox News. He finds that all the networks reported poll number that favored Clinton, except for Fox, which tended to report unfavorable poll results. This method

makes no claims as to the bias of the network's content, but instead claims that the network's bias determines which statistics are presented as facts.

Coding the transcripts of televised content is an equally as common way of qualitatively identifying bias on news networks. For example, Aday (2010) assessed bias in reporting of the Iraq war on NBC and Fox by coding the quotes in news stories. The quotes in stories were coded according to four categories: the event of a quote, the type of source, the tone of the quote, and the description of enemies. They found that both networks deemphasized the negative aspects of war, but they found that Fox was more sympathetic with the Bush administration than NBC. In another study of Iraq war coverage, Aday, Livingston, and Hebert (2005) coded the tone of articles from ABC, CBS, NBC, CNN, Fox News, and Al Jazeera according to level of editorial detachment and use of connotative language. They concluded that Fox News was the only network that did not report the war in a balanced way, portraying American involvement positively. Fico, Zeldes, Carpenter, and Diddi (2008) analyzed stories covering the 2004 presidential campaign between Bush and Kerry by measuring whose supporters spoke first, who was pictured more in visuals, who was favored more in quotes from partisans, and who had more time devoted to him. Their findings indicate that Fox gave preference to Bush and MSNBC gave preference to Kerry. Conway, Grabe, and Grieves (2007) analyzed the transcripts of Bill O'Reilly's "Talking Point Memos" on Fox using the Institute of Propaganda Analysis's seven propaganda devices and found that O'Reilly portrays the political left negatively and the Bush administration positively. To study the bias in coverage of justice issues, Britto and Dabney (2010) coded the content of segments and the characteristics of guests on political talk shows on CNN, Fox News,

and MSNBC. They found that conservative guests were asked more questions concerning justice issues and given more time to talk than liberal guests, suggesting that both networks lean right in this area.

Using a computational content-analysis tool, Holtzman et al. (in press) analyzed transcripts from CNN, Fox News, and MSNBC with the Contrast Analysis of Semantic Similarity tool (CASS). This tool leads to bias assessment by finding patterns of the associations made between certain words and concepts in a corpus. Hence, the tool can find if a subject is essentially treated positively or negatively. Holtzman et al. found that Republicans are associated with positive words on Fox and Democrats are associated with positive words on MSNBC.

These previous studies of televised content have identified bias at Fox News and MSNBC. For example, Fico et al. (2008) found that MSNBC and Fox covered the 2004 presidential election in an unbalanced way. In addition, Holtzman, Schott, Jones, Balota, and Yarkoni (in press) identified positive associations with Republicans and Democrats on Fox and MSNBC respectively. Many of the studies identified biased reporting from Fox News but did not study MSNBC. For instance, Groeling (2008) shows that Fox tended to report unfavorable approval ratings for Clinton, a Democratic president. Aday (2010) found that Fox's coverage of the Iraq War portrayed the Bush administration more favorably than similar coverage on MSNBC's parent network, NBC. Similarly, Aday et al. (2005) found that Fox was sympathetic with American involvement in Iraq. Conway, et al. (2007) concluded that Fox's Bill O'Reilly depicts the political left and illegal immigrants negatively but depicts the Bush administration, the criminal justice system, and "ordinary" US citizens positively. Overall, these studies conclude that the televised

content covering health care and immigration from Fox has a conservative bias and MSNBC's televised content of these issues has a liberal bias.

Gaps in the Research

Most of the previous research on bias in the reporting from Fox News and MSNBC has focused on televised content (Aday, 2010; Aday, et al. 2005; Conway et al. 2007; Fico et al. 2008; Groeling, 2008). Research on the electronic news articles from the networks is minimal and superficial. For example, Weatherly, Petros, Christopherson, and Haugen (2007) recognized the value of analyzing the network's online content. However, their study is limited to only the headlines of articles. In contrast, my study is a content analysis of articles. Therefore, my study contributes to the field of assessing bias from Fox News and MSNBC by examining the little-researched medium of online hard news articles. By contributing another perspective on these networks, my study helps build a stronger case for institutional bias.

First, my study contributes to the field by examining reporting from MSNBC, which is relatively understudied compared to Fox News. For example, in my review of the literature, I found only three articles investigating content from MSNBC (Britto & Dabney, 2010; Fico et al. 2008; Holtzman et al., in press). In contrast, seven studies examine bias on Fox News (Aday, 2010; Aday et al. 2005; Britto & Dabney, 2010; Conway et al. 2007; Fico et al. 2008; Groeling, 2008; Weatherly et al. 2007). This imbalance indicates a bias in the literature toward assuming Fox is more biased then MSNBC, which is an assumption that this study seeks to avoid.

Second, my study builds on previous research on election coverage and also contributes a study on the little-researched issue of immigration policy. Fico et al. (2008) and Groeling (2008) examined election coverage for bias, but they only studied presidential elections. Because of the time frame of my study, I analyze the coverage of mid-terms elections, which have received little attention in the literature.

Unlike the majority of bias research, my study investigates reporting on immigration policy. Some studies examined issues related to immigration debates, such as nationalism and representations of race (Britto & Dabney, 2010, Conway et al. 2007). But there are few studies of immigration policy. Therefore, this study contributes to the field of bias assessment by focusing on an alternative form of election and opening the domain of immigration policy to research.

Lastly, my study uses a new approach to content analysis. Few studies have used computational tools to assess bias on these networks. Holtzman et al. (in press) used the tool CASS, which can only indicate the words associated with other words in a corpus. However, I use the tool the Gramulator (McCarthy, Watanabe, & Lamkin, in press), which allows researchers to see what collocations are distinctive of one corpus, as compared to another. Furthermore, it allows researchers to see those collocations in context, allowing for deeper understandings of how collocations are used in context. Seeing distinctive collocations identified by a computational tool in context allows my study to quantitatively identify the most revealing words in a corpus and then analyze them qualitatively with a high degree of sensitivity to context. Few studies have been able to combine quantitative and qualitative approaches in this way.

Overall, my study contributes to previous linguistic research on the biases of Fox and MSNBC by exploring three under-researched areas. First, it examines electronic print content, where as a majority of bias studies of Fox and MSNBC focus on televised content. Second, most studies pertain to Fox but not MSNBC. And last, my study processes corpora from these networks using a new tool, the Gramulator. Therefore, this study opens content analysis to a new domain, makes the literature more balanced, and introduces a new tool to the field.

Methods

The Tool

To analyze my corpora, I use the Applied Natural Language Processing (ANLP) tool the Gramulator (McCarthy et al. in press). This tool identifies the indicative linguistic features that differentiate corpora. Using a computational linguistic tool makes my descriptions of discourse *systematic*. The Gramulator also highlights collocations that may contain key linguistic features of a text, thereby making the descriptions *explicit*. These two functions follow the prescription made by van Dijk (1988) that discourse analysis be systematic and explicit.

The Gramulator is appropriate for contrastive corpus analysis. Other ANLP tools, such as Linguistic Inquiry and Word Count (LIWC) (Pennebaker, Booth, & Francis, 2007) and Contrast Analysis of Semantic Similarity (CASS) (Holtzman et al. in press) only draw semantic connections among words within texts in a single corpus. But the Gramulator compares two corpora to each other. Furthermore, the Gramulator compares more than single words; it compares texts in terms of collocations.

In this study, collocations are defined as any words that co-occur one after the other. Traditionally in linguistics, there is a more narrow definition of collocation. This definition considers collocations words that appear together more often than by accident.

The main unit of analysis in this study is the *bi-gram*, or two-word collocation. In some cases, these bi-grams form parts of larger collocations, such as *tri-grams* or *quad-grams*, or three- and four-word collocations. Sometimes, a bi-gram repeatedly co-occurs in a sentence along with a particular word, lexeme, or syntactic structure. These bi-grams are not part of larger collocations, but often appear in similar contexts. This type of pattern is called a *flexi-gram*. For example, *Obama supports deploying more National Guard troops to the border* and *Obama approved legislation to deploy reinforcements for the National Guard* are examples of flexigrams containing the bi-gram *National Guard* and a lexeme of *to deploy*.

Architecture. The Gramulator consists of eight modules. The main module segments texts into collocations and identifies the collocations that are distinctive of each corpus. The remaining modules serve to view differentials or manipulate the corpora.

Main module. The Gramulator identifies *differentials*, or collocations that appear significantly more frequently in one corpus than another. First, the tool finds the collocations of a corpus that appear with above average frequentcy, or *typicals*. Then, it eliminates collocations that are common between the two corpora. The remaining collocations, the differentials, are indicative of that corpus relative to the corpus it is contrasted with, or its *sister corpus*. Thus, the Gramulator highlights indicative features of contrastive corpora.

The Concordancer. The Concordancer allows researchers to see differentials in context. The module has a search function that finds all instances of a collocation in a corpus. These instances are displayed within the context that they appear. Seeing differentials in context allows researchers more nuanced understandings of how the differentials are used in the corpus. Therefore, the Concordancer is key in the qualitative element of this study.

The Sorter. The Sorter randomly distributes texts from a corpus into smaller groups. This function is necessary when performing statistical analysis (such as t-tests) to validate corpora.

The Evaluator. The Evaluator allows researchers to measure the presence of an index in a corpus. An index is a list of items, be they words or collocations. Therefore, the module evaluates how often a group of words or collocations appears in a corpus.

The Corpora

The corpus includes 858 texts collected from Fox News and MSNBC's websites. The Fox News corpus is composed of 696 texts, and the MSNBC corpus is composed of 162 texts (see Table 1). The two main corpora are divided into two sub-corpora.

Table 1

Corpus Composition

Sub-corpus	Fox	MSNBC
Elections	261	65
Immigration	435	97
Total	696	162

I collected texts on the subjects of *elections* and *immigration* policy for two reasons. First, I chose a subject that has been extensively researched, elections, and a subject that is relatively understudied in this field, immigration policy. Second, I chose topics that were likely to include many opportunities for bias: Elections are essentially struggles between ideologies, and the immigration debate of 2010 was a high-profile contentious issue that divided the country along party lines.

Collection of corpora. All texts were collected from the websites of Fox News and MSNBC. The texts were found through each site's search function. The search terms used coincide with the label of each sub-corpus category. However, the terms for the elections sub-corpora are slightly different. Instead of searching for *elections*, *election* with no *s* was used in order to maximize the search results, because *elections* restricted results to articles that mentioned multiple races. On the MSNBC website, a second search specifying *election 2010* was necessary because of the various years provided in the search results.

Selecting texts. From the search results, I chose articles according to their relevance to the search terms listed above. Broad parameters were used when deciding what was relevant. The sub-corpora contain articles concerning the following issues. The election corpora include topics such as the mid-term elections, special elections, primaries, as well as scandals and deaths that left elected offices empty. The immigration corpora include articles concerning immigration legislation, enforcement of immigration laws, border security, and human interest stories. However, these are not exhaustive lists of the topics included. In general, I included all texts that seemed relevant. Texts that were not deemed relevant generally covered international politics, such as foreign

elections and immigration policies. Others include highly localized elections such as municipal and school board elections.

   Cleansing texts. The texts analyzed contain only the body of the articles. All extraneous text was removed. Examples of extraneous text include the title, date, author, city, section, sub-section, captions to pictures, and links embedded in the text.

   Justification of corpora. The corpora are not equal in size because the networks published different numbers of texts during the set timeframes. This difference reflects disparities in coverage influenced by the networks' ideologies. Thus, the corpora will show what each network deemed newsworthy during a given news cycle, revealing selection bias.

   The timeframes that the corpora cover vary because the Fox News archives limit access sooner than the MSNBC website. MSNBC does not allow users to sort search results. Instead, results from various years are presented out of order. Many of the same articles are repeated throughout the search results. And users cannot access beyond the 100th page of results. On the other hand, Fox News allows users to sort search results by date. However, users cannot access beyond the 51st page of results. Therefore, the last date accessible on the Fox News website determined the beginning date of the timeframes. The end date for all timeframes is June 30, 2010.

Table 2

Corpora Timeframes

Sub-corpus	Fox News	MSNBC	Begin Date	End Date
Elections	261	65	2/1/10	6/30/10
Immigration	435	97	8/1/08	6/30/10
Total	696	162		

Even though the corpora contain unequal numbers of texts, they are representative of the target population. The target population was all texts published within the timeframes. It was impossible to collect all the articles published during the timeframes, but it was possible to collect all articles available through the websites' search function. The Fox News results were ordered chronologically, therefore I was sure to have all the available articles before the cut-off date. The MSNBC results were not ordered chronologically, but articles did repeat among the search results before the last accessible page, as if there were no other relevant articles. Therefore, I collected all of the articles available from the timeframes that were available.

I argue that these parameters for the corpora are justified, because they include the largest number of texts possible within a given timeframe. The various cut-off dates for these time frames were not chosen to encompass any particular stories. Instead, they were dictated by the restraints of the network's websites. It is important that contrasting corpora be from the same timeframe in order to assess Fox News and MSNBC's selection bias.

Validation of Corpora

It is important to verify linguistic features are consistent across a corpus. If not, a relatively small number of texts, dissimilar from the rest of the corpus, could substantially influence the results. In the case of the Gramulator, if a corpus is not consistent, then differentials may over represent certain linguistic features of the corpus. Therefore, validating a corpus lends confidence that the identified differentials are indicative of their respective corpus.

Corpus validation can be accomplished using a *t-test*. A t-test can be used to assess the consistency of a corpus by comparing its random parts. To conduct this validation procedure, a corpus is divided into two parts, a *test set* and a *training set*. A training set consists of two-thirds of a corpus, and a test set consists of one third of the corpus. Texts are randomly assigned to each set using the Sorter module of Gramulator. Random selection provides confidence that potential outlying texts are not disproportionally represented in either set. The goal of the validation procedure is to provide evidence that both sub-corpora equally represent the whole corpus.

In Gramulator studies, the main module is used to contrast the two training sets, and from these training sets to provide training differentials. The Evaluator module is then used show the presence of differentials from contrasted training sets in the test set. A t-test assesses in which test set the differentials appear with greater consistency, or whether there is no significant difference between them. Thus, a t-tests can provide evidence for the consistency of a corpus if the differentials from its training set are more predictive of its corresponding test set than in the test set of the sister corpus[1].

The Gramulator, like all textual analysis tools and statistical procedures, assumes that all things are equal (the so called null hypothesis). Therefore, any differences have the potential to affect the analysis. In this study, the difference in size of the Fox and MSNBC corpora is considerable (696 to 162). Therefore, we have to accommodate the disparities.

---

[1] Strictly speaking, in statistical terms, the test set is "explaining" the training set rather than the training set "predicting" the test set; however, in a validation procedure, the distinction in terminology is not important and the term *predictive* better describes to goal of the procedure.

I follow the standard method of validating corpora, as described by McCarthy, Watanabe, and Lamkin (in press); however, my study requires an extra step in order to account for the differences in corpora size. In the standard method, the corpora are equal in size and text length, therefore the unequal number of differentials produced can be explained by the composition of the respective corpora. However, in my study, the difference in the size and text length of the respective corpora are likely to result in corresponding differences in the number of differentials. These differentials are likely to reflect the differences in corpora size and text length, rather than the composition of the corpora. Therefore, only the number of differentials that made a unique contribution to the identification of the corpus were included in the index. For example, XX appears in 20 texts, YY appears in the same 20 texts, and ZZ appears in 10 texts. YY does not contribute to the predictability of an index, because XX predicts the same 20 texts as YY. However, ZZ predicts different texts. Therefore, ZZ contributes to the predictability of the index, whereas YY is redundant. However, the difference in number of texts is still important, because the larger index ($Fox_{training}$) will have more opportunity to make a unique contribution. Thus, in this analysis, we tagged the Fox index to cover the same degree of frequency as the MSNBC corpus.

In the election corpus, the resulting Fox index included 6 Fox differentials and the resulting MSNBC index included 17 MSNBC differentials. Because differentials contribute different values, however, the final analysis was 67% for MSNBC and 69% for Fox. In other words, the MSNBC training differentials occurred in 67% of the training set, whereas the Fox training set differentials occurred in 69% of the Fox training set.

To validate the Fox election corpus, I conducted a within texts t-test to assess the effect of the $Fox_{training}$ differentials and the $MSNBC_{training}$ differentials on Fox_{test}. The result was in the predicted direction and reached a level of significance: $t\,(1,174) = 7.121$, $p < .001$, $d = 0.538$. The effect size of 0.538 can be described as medium. The result suggests that $Fox_{training}$ is more present than $MSNBC_{training}$ in Fox_{test}. The result also validates the Fox index.

To validate the MSNBC election corpus, I conducted a within texts t-test to assess the effect of the $MSNBC_{training}$ and $Fox_{training}$ on the corpus of $MSNBC_{test}$. The result was in the predicted direction and reached a level of significance in a 1-tail test: $t\,(1,20) = 1.861$, $p = 0.078$, $d = 0.406$. The effect size of 0.406 can be described as medium. The result provides some evidence that $MSNBC_{training}$ may be more present than $Fox_{training}$ in $MSNBC_{test}$; and therefore provides some evidence towards validating the MSNBC index.

Results

The Gramulator's Main Module contrasted the election and immigration corpora from Fox and MSNBC to produce differentials, or two word collocations indicative of the corpora as compared with their sister corpora. Concordensing these differentials revealed how the differentials are used in context. Patterns among differentials and patterns among their uses in context emerged in each corpus. These patterns suggest that the two networks focused on different aspects of the immigration debate and mid-term elections in 2010.

Elections

Given the time span covered by the Fox and MSNBC election sub-corpora, February 1st to June 30th, the majority of differentials refer to the 2010 midterm elections and the party primaries that determined the candidates for those elections. Overall, the differentials from the Fox sub-corpus deal with the relationship between political figures, legislation, and public opinion. On the other hand, the differentials from the MSNBC sub-corpus are characteristic of election discourse regarding election strategies and party politics.

Fox. Overall, the differentials from the Fox election corpus link voters' opinions of legislation, politicians' ownership of legislation, and the resulting election battles. For example, Fox frequently mentions *President Obama*, emphasizing his association with contentious legislation, whereas MSNBC mentions the president only twice in its entire corpus on elections. Furthermore, instances of *the bill* and *the law* tend to refer to voters' receptions of legislation in the Fox corpus, unlike instances in the MSNBC corpus. These perceptions are supported by polls in the Fox corpus, as noted by four differentials, in contrast to the MSNBC corpus, which has no differentials relating to polls. And throughout the corpus, Democrats are often grouped into a bloc, *the Democrats*, suggesting little attention to differences within the party. However, MSNBC devoted more attention to *the Republicans* as a bloc than Fox devoted to *the Democrats*.

Table 3

Fox Election Corpus Differentials

Differentials	Fox texts	MSNBC texts	Significance
Fox News	84	3	$p < .001$
told Fox	44	0	$p < .001$
President Obama	78	2	$p < .001$
Harry Reid	36	3	$p = .052$
Senate majority	36	4	$p = .137$
say they	28	3	$p = .160$
Joe Sestak	32	5	$p = .385$
FoxNews. com	26	0	$p = .004$
the American	30	5	$p = .503$
Blanche Lincoln	34	5	$p = .290$
margin of	29	3	$p = .161$
the former	31	4	$p = .262$
the bill	24	4	$p = .621$
lot of	29	5	$p = .503$
Leader Harry	32	3	$p = .114$
President Obama's	31	1	$p = .009$
Sarah Palin	35	7	$p = .682$
the President's	27	2	$p = .086$

Table 3

Fox Election Corpus Differentials (continued)

Differentials	Fox texts	MSNBC texts	Significance
Rep. Joe	30	4	$p = .261$
the way	30	5	$p = .503$
there is	27	4	$p = .356$
Democrats are	26	4	$p = .473$
Sen. Blanche	30	4	$p = .261$
the law	21	3	$p = .435$
former Rep.	30	3	$p = .112$
have the	26	5	$p = .813$
told Foxnews	25	0	$p = .007$
it will	27	4	$p = .356$
the latest	27	4	$p = .356$
in Arkansas	23	3	$p = .318$
not be	26	3	$p = .227$

Content-revealing differentials. Differentials in this sub-section deal with the interplay between political figures, the legislation they support, and public opinion. Overall, they seem to be part of an election narrative explaining what political developments have influenced voter sentiment regarding candidates and issues, which will impact election results.

Politicians. The most frequent bi-grams among the top differentials relate to political figures, the majority of whom are Democratic lawmakers. For example, President Obama figures prominently in the Fox corpus, appearing in two differentials, whereas MSNBC mentions him only twice.

In the Fox corpus, the highest-ranking differential in the Politicians category is also the highest-ranking politician in the corpus, *President Obama* ($p < .001$). As president, *President Obama* figures high among the differentials, because of his central role in the political process. Another differential, *President Obama's* highlights his ownership of certain legislative initiatives, especially legislation regarding health care and economic stimulus. Take the following example, *But Inglis defends his conservatism, pointing to his opposition to President Obama's stimulus package, auto-industry bailout and health care overhaul* (Lemmatized 104 Fox 194).

In contrast, President Obama is mentioned in only two of the 65 texts in the MSNBC corpus. *President Obama* appears in one text and *President Obama's* in the other. Together, the two texts comprise 3.1% of the MSNBC corpus, opposed to 29.9% of the Fox corpus. Furthermore, the instance of President Obama is introduced indirectly into the text through a quote, "I never thought I'd run for office again , but with the direction President Obama is taking the country, (wife) Marsha and I decided we had to stand up," former Sen Dan coats…told supporters in a campaign commercial (Lemmatized 27 MSN 34). In addition, the object *President Obama's* is modifying is vague, *President Obama's first-term agenda*, compared to the instances in the Fox corpus.

Several differentials mark Harry Reid as a major figure in Fox's election coverage. *Majority leader*, *leader Harry*, and *Harry Reid* combine to form the quad-gram *majority leader Harry Reid*. Harry Reid figures among the top differentials, because as majority leader he is a symbol of the status quo, e.g., *"The problem is Harry Reid"* (Lemmatized 159 Fox 245). Notably, instances of *Harry Reid*, independent of the quad-gram *majority leader Harry Reid*, often appear in quotations, as in *"I say Harry Reid you have failed America....I say Harry Reid, you're fired"* (Lemmatized 15 Fox 112). As in the quote above, most of the instances of Harry Reid dealt with his re-election campaign. In 2010, Reid was the incumbent in a high-profile senate race, due to his position, for example, *Angle in Nevada is outpolling a famous and powerful Democrat, Senate majority leader Harry Reid* (Lemmatized 147 Fox 234).

In contrast, Harry Reid is mentioned in only 4.6% of the MSNBC corpus, opposed to 13.8% in the Fox corpus. Also unlike the Fox corpus, instances of Harry Reid never appear in quotes.

Two differentials, *Rep. Joe* and *Joe Sestak*, indicate that the Obama administration's job offer figured prominently in Fox's election coverage. During the 2010 Pennsylvania primaries, the Obama White House offered Joe Sestak a position in the executive, which some commentators considered a move to influence the Democratic primary (New York Times 2010). Some critics and lawmakers questioned the legality of this move, as shown in *A GOP lawmaker says that the White House committed a "crime" if it offered Democratic Rep. Joe Sestak a federal job in exchange for dropping his primary challenge to Sen. Arlen Specter* (Lemmatized 244 Fox 84). That lawmaker was

Rep. Darrell Issa, R-Calif., who led the charge in Congress to find out more about the

Pennsylvania race (Lemmatized 105 Fox 195).

In the MSNBC corpus, Joe Sestak is mentioned in 7.7% of texts, opposed to

12.3% on Fox. One instance deals with the job offer from the White House. However, the

others deal with his campaign.

Coverage of the Arkansas democratic primary race portrayed *Sen. Blanche*

Lincoln as an embattled fiscal conservative on Fox and as a natural victim of anti-

incumbent sentiment on MSNBC.

The differentials *Sen. Blanche*, *Blanche Lincoln*, and *in Arkansas* reveal frequent

coverage of the 2010 Democratic primary in Arkansas. The Arkansas primary was a

high-profile campaign because of the ideological tensions it displayed within the

Democratic Party. As one article explains, Sen. Blanche Lincoln was the fiscal

conservative in the race, *the "blue dog" Democrat is competing in a heated runoff*

against Arkansas Lt. Gov. Bill Halter, who has the backing of labor unions and liberal

groups (Lemmatized 121 Fox 209).

In contrast, MSNBC covered the race less. Mentions of Blanche Lincoln appeared

in 7.7% of MSNBC texts, opposed to 13% of Fox texts. MSNBC tends to downplay the

race as a struggle between ideologies in the Democratic party. Instead, MSNBC mentions

Sen. Lincoln along with other Democrats who were the targets of anti-incumbent

sentiment, as one article states, *It forced incumbent Sen. Blanche Lincoln of Arkansas*

into a primary runoff and turned out 14-term Democratic Rep. Allan Mollohan of West

Virginia (Lemmatized 49 MSN 54).

Sarah Palin receives similar treatment on Fox and MSNBC. The differential *Sarah Palin* generally accompanies details regarding the political figure's support for candidates. Notably, Sarah Palin is most often mentioned alongside other political figures, as in *The candidate has big-name support from GOP luminaries like former Massachusetts Gov. Mitt Romney and former Vice Presidential nominee Sarah Palin* (Lemmatized 177 Fox 262). Overall, references to Sarah Palin emphasize the influence of her support on election results, as in *But a Tea Party surge and Sarah Palin's endorsement propelled her to an easy first-place finish* (Lemmatized 151 Fox 238).

On MSNBC, Sarah Palin received an almost equal amount of attention as on Fox, appearing in 10.8% of texts compared to Fox's 13.4%. She was mentioned in many of the same ways as on Fox, often in a flexigram including the word *backing*, as in *Mark Sanford has the backing of Sanford's ex-wife as well as support from conservative stars Sarah Palin and Mitt Romney* (Lemmatized 50 MSN 55).

The two differentials *the former* and *former Rep.* refer most often to the candidates J.D. Hayworth, Pat Toomey, and Tom Campbell. Hayworth is a Democrat, And Toomey and Campbell are Republicans challenging high profile Democrats. J.D. Hayworth challenged John McCain in the Republican Arizona primary, as shown in the example *In Arizona, Sen. John McCain is in a tough primary fight against former Rep. J.D. Hayworth, a conservative talk-radio host* (Lemmatized 40 Fox 135). Pat Toomey was the Republican candidate running against Joe Sestak for the Pennsylvania Senate seat, as noted in *It would make sense for the White House to lend support to Sestak in his run against his Republican opponent, former Rep. Pat Toomey, as part of its efforts to fend off a GOP attempt to regain control of the Senate in November* (Lemmatized 68 Fox

160). And Tom Campbell was a contender in the California Republican primary, as explained in *Carly Fiorina…has claimed a significant lead over former Rep. Tom Campbell in the Republican primary to challenge California Sen. Barbara Boxer in November* (Lemmatized 119 Fox 207). MSNBC occasionally qualifies these candidates as *former reps*., but it is not as frequent as on Fox.

    Legislation. On Fox, discussions of *the bill* and *the law* tend to refer to the public's perceptions of legislation, whereas MSNBC focuses on bills' passing and contents.

    The bi-grams *the bill* and *the law* indicate that election coverage on Fox News involves discussions of legislation, particularly health care legislation. Often references to laws attest to a candidates voting record. Other times, legislation is discussed in relation to voter opinion. For example, instances of *the bill* often refer to voters' perceptions of legislation, as in *He got an earful from voters who said the bill will bankrupt American and shred the limited government intent of the U.S. constitution* (Lemmatized 256 Fox 95). Similarly, instances of *the bill* also directly link legislation to predicted voter behavior in the November elections, as in *Tea Party leaders…say their next step will be to turn from fighting the bill to running the people who voted for it out of Congress* (Lemmatized 251 Fox 90). As show in the two examples above, the legislation at issue is most commonly the health care. The same is true for instances of *the law*.

    In the MSNBC corpus, instances of the bill and the law also refer to health care legislations, but very few instances refer to a bill's reception by the public. Instead, instances of these differentials generally refer to the passing of a bill and its contents, for

example consider the following example containing the differential, *With that, Stupak and other anti-abortion democrats voted for the bill* (Lemmatized 62 MSN 7).

   Polls. Fox and MSNBC use poll results to support their articles with almost equal frequency. However, Fox has a distinctive way of incorporating them into its coverage, whereas MSNBC does not make polls a signature aspect of its reporting.

   Opinion polls figure very prominently in the Fox sub-corpus. Several of the differentials, *say they*, *margin of*, *the latest*, and *fox news*, are linked to reports of poll results. Most frequently, the bi-gram *the latest* refers to *the latest poll*, *the latest polling*, or *the latest survey*. And occasionally a poll is specified as a *Fox news poll*. Overall, coverage of these polls attempt to summarize voter opinion. For example, the differential *say they* often serves to interpret poll figures, as in *Thirty-five percent of American voters say they generally trust the federal government* (Lemmatized 148 Fox 235). In relation, the differential *margin of* serves to qualify the accuracy of those poll results by revealing the *margin of error*, as in *The poll had a margin of 5 percentage points* (Lemmatized 13 Fox 110).

   MSNBC often cites polls, but the lack of differentials relating to polls suggests that polls are not a distinctive trait of MSNBC's reporting. The word *poll* appears in 20% of MSNBC texts and 23.8% of Fox texts. However, Fox is the only corpus with differentials referring to polls. Therefore, polls are referred to nearly as often in each corpus, but Fox has a way of discussing polls that is indicative to its reporting.

   Political parties. The Fox sub-corpus often mentions Democrats as a group. The MSNBC sub-corpus frequently refers to the Democratic Party in differentials such as *his democratic*. However, the Fox sub-corpus often refers to members of the party using the

definite article in the differential *the Democrats*. This differential often refers to the party as a whole, particularly in coverage of election strategies, as in *It will strip the Democrats of their majority* (Lemmatized 189 Fox 33). Also, the use of the definite article groups members in a bloc, depicting them as a monolith, as in *Pollsters, analysts and politicians across party lines agree the Democrats have lost their grip on the issue* (Lemmatized 170 Fox 255).

The Democrats is a differential in the Fox corpus, but is used in a slightly larger percentage of the MSNBC corpus, 15.4%, than in the Fox corpus, 13.4%. The biggest difference lies between the network's mention of *the Republicans*, 12.3% on MSNBC and 5.7% on Fox. This difference suggests that MSNBC devoted more attention to Republicans than Fox devoted to Democrats.

Pan-corpora differentials. Differentials in this section appear in other Fox sub-corpora. Note that the differential *President Obama* is also a pan-corpora differential, but is analyzed in the "Politicians" sub-section found above.

Self-reference. As in all of the Fox sub-corpora, references to the network are among the top differentials. Most often, these differentials refer to interviews between officials or experts and the network. For example, instances of *Fox News* ($p < .001$) appear most often in combination with *told Fox* ($p < .001$). Likewise, *foxnews com* ($p = .004$) appears most often in the tri-grams *told Fox News*. On its own, *Fox News* refers most often to polls, as in the fragment *according to a Fox News poll released Friday* (Lemmatized Fox 1).

Populism.

As in the Fox sub-corpora on the economy, *the American* is a top differential. Similar to other corpora, *the American* is often, but not exclusively, found in quotes and part of the larger tri-gram *the American people*. In the context of elections, lawmakers use this tri-gram to position themselves on the side of the electorate, as in *"They want us to stop spending their money, and I think it's time for Congress to listen to the American people"* (Lemmatized 147 Fox 234).

Syntactic differentials. The differential *there is* is often followed negation, for example, *There is no doubt today that conservatives again feel compelled to protect constitutional liberty anew* (Lemmatized 191 Fox 35).

The differential *it will* is a frequent collocation, given the emphasis on the future in election cycles, for example, *If the GOP can run the table, it will strip the Democrats of their majority* (Lemmatized 189 Fox 33).

Of the instances of the bi-gram *have the*, the tri-gram *have the right* is the most common collocation, for example, *A New Jersey Appeals Court Ruled Tuesday that a conservative Tea Party group should have the right to try to throw a U.S. Senator out of office* (Lemmatized 246 Fox 86).

MSNBC. Differentials in the MSNBC sub-corpus are largely characteristic of election discourse in general. Some differentials rather neutrally refer to elections, whereas others are often used in a partisan way. Many neutrally used election differentials refer to political strategies in elections. For example, MSNBC discusses the political leanings of voting districts more than Fox. Furthermore, MSNBC discusses campaigns in personal terms, using the possessive pronoun *his*, more than Fox. MSNBC

also offers commentary on elections from professors more than Fox. As for partisan uses of differentials, MSNBC focuses on *the backing* of Republican candidates, whereas Fox gives more equal attention to *the backing* of both Republican and Democratic candidates. Both networks substitute *the GOP* for the Republican Party; however, MSNBC does it more than Fox, and disproportionately more when attributing a policy, person, or event to Republicans by using the possessive differential *the GOP's*.

Table 4

MSNBC Election Corpus Differentials

Differentials	Fox texts	MSNBC texts	Significance
the district	9	9	p = .003
seat and	1	5	p = .001
a moderate	9	8	p = .009
the incumbent	8	9	p = .002
and an	7	8	p = .003
and was	9	7	p = .023
who won	29	11	p = .208
the backing	8	7	p = .015
first-term	8	7	p = .015
governor's race	6	7	p = .006
his decision	9	7	p = .023
his democratic	5	7	p = .003

Table 4

MSNBC Election Corpus Differentials   (continued)

Differentials	Fox texts	MSNBC texts	Significance
lost his	6	6	p = .017
race is	9	6	p = .089
a senate	10	7	p = .054
close to	5	7	p = .003
I will	7	7	p = .010
including the	10	7	p = .054
is running	8	7	p = .015
money and	8	7	p = .015
people and	3	4	p = .031
role in	7	7	p = .010
who also	3	7	p < .001
former congressman	10	8	p = .013
his candidacy	10	8	p = .013
state house	9	6	p = .089
professor at	6	6	p = .017
to file	7	6	p = .027
challenge from	9	5	p = .165
the gop's	5	6	p = .010

Content-revealing differentials

General election discourse. Differentials in this sub-section are characteristic of election discourse in general, particularly election strategies and trends. Overall, they seem to be part of the election narrative regarding representation of parties in Congress. The distribution of seats in Congress was a key concern of Democrats, who were defending their majority in the 2010 mid-term elections. However, these differentials do not appear to refer proportionally more to one party than another, opposed to differentials in the Party Politics section below.

The top differential, *the district* ($p = .003$), refers directly to election strategies. Overall, discussions of the district serve to describe the district and its political leanings, for example *The district […] usually votes so strongly Democratic that the race is rated Democratic favored* (Lemmatized 2 MSN 11). Some instances of the district refer to the voting history of a district, as in *It was a complete reversal from 2004, when George W. Bush won the district (*Lemmatized 63 MSN 8). Others attempt to explain the voting histories through demographics, for instance *the district, which is 72 percent white, has sent a Republican to Congress for three decades* (Lemmatized 12 MSN 20).

The district is used similarly in the Fox corpus but appears only 3.4% of texts, opposed to 13.8% in the MSNBC corpus. This difference suggests that Fox devotes less attention to the effect districts' political leanings will have on elections.

Another differential, *challenge from*, indicates the main tensions in an election, generally primaries. For example, Lt. Gov. Bill Halter's entry into the Democratic primary for an Arkansas senate seat changed the dynamic of Sen. Blanche Lincoln's election campaign, as explained in the following sentence containing the differential, *But

the challenge from Halter, who has the backing of progressives, could force her to rethink her position (Lemmatized 59 MSN 63).

Instances of *challenge from* in the Fox corpus generally mention the tensions within a primary as well. The Arkansas Democratic primary featuring Lincoln and Halther is a major focus. Another major focus is the Arizona Republican primary where John McCain was challenged from J.D. Hayworth, as seen in *McCain is facing a primary challenge from the right in what is expected to be the toughest re-election campaign of his Senate career* (Lemmatized 233 Fox 74).

The differential *a senate* is generally related to specific races. The bi-gram appears most often in the tri-gram *a senate seat*, for example *The same goes for Louisiana Rep. Charlie Melancon, a Democrat in a GOP-leaning district who also is seeking a senate seat* (Lemmatized 64 MSN 9).

There are too few instances of *people and* ($p = .015$) in the corpus to identify patterns in its usage. However, it is interesting to compare with the thread of populism in the Fox corpus, evinced by the high frequency of the tri-gram *the American people*. In contrast, there are no instances of *American people and* in the MSNBC corpus.

Ownership. Several differentials from the MSNBC corpus feature possessive pronouns. The possessive pronoun is always the masculine *his*. Reoccurrence of these possessive pronouns emphasize a candidate's ownership of developments in his campaign. In this section on General Election discourse, the differentials featuring possessive pronouns are *his candidacy* and *lost his*. *His candidacy* is generally accompanied by some form of the verb *to announce*. The bi-gram *lost his* is often in the tri-grams *lost his bid* and *lost his seat*. Most of the instances of *lost his* pertain to primary

43

results, as in *Last week, another recruit, Jeff Reetz, lost his Kentucky primary to a Tea Party favorite* (Lemmatized 21 MSN 29).

These bi-grams appear in the Fox corpus, but less frequently. Together, his candidacy and lost his appear in only 6.1% of the Fox corpus, compared to 21.5% of the MSNBC corpus. This difference suggests that MSNBC devotes more attention to covering campaigns in personal terms.

Modifiers. In the MSNBC sub-corpus, political figures are often described in terms of their political positions and campaign activity. The differential *former Congressman* is an example of a bi-gram that defines candidates, as in *Dave Nagle, a Democratic political activist and former Congressman from Iowa* (Lemmatized_ MSNBC 1). Other differentials use relative pronouns to describe candidates, for example *who also* and *who wo*n, as in *Roxanne Conlin, who won the Democratic senate nomination in Iowa* (Lemmatized 23 MSN 30).

One modifier among the top differentials does not refer to politicians, *professor at* ($p = .017$), but rather to academics providing commentary, as in *Governors are among the most vulnerable of Democratic incumbents this year, said Seth Masket, a political science professor at the University of Denver* (Lemmatized 12 MSN 21). Fox similarly cites professors as sources, but less than MSNBC. Professors are mentioned in 5% of Fox texts but in 12.3% of MSNBC texts. This differences suggests a stronger reliance on the expertise of professors at MSNBC than at Fox.

State politics. Among the differentials, some refer to state politics specifically. The bi-gram *governor's race* points directly to this arena of election politics. The majority of the instances of the differential appear in articles about the Republican

primary for the Alabama governorship, as shown in *The Republican primary in the Alabama governor's race has provided some of the most provocative ads of 2010 so far* (Lemmatized 39 MSN 45). In addition, the differential *state house* refers to candidates' backgrounds in state politics, normally as *state house speakers*, before entering national races, as in *another well-established candidate, former state house speaker Marco Rubio, has stronger conservative credentials* (Lemmatized 24 MSN 31).

   Party politics. Differentials in this sub-section explicitly relate to partisan discourse or frequently coincide with mentions of one party more than another.

   MSNBC frequently refers to *the backing* of Republican supporters, where as Fox refers to *the backing* of both Republican and Democratic supporters. This difference suggests that MSNBC devotes more attention to explaining the support for Republican candidates than Democratic candidates.

   In the MSNBC sub-corpus, *the backing* ($p = .015$) of political candidates generally refers to conservative figures and groups supporting Republican candidates. Among other conservative organizations, the Tea Party is often mentioned as a backer of Republican candidates, such as *With Tea Party support and the backing of local GOP leaders* […]*, Brooks won Tuesday's primary with slightly more than 50 percent of the vote in a three-candidate field* (Lemmatized 49 MSN 54). One political figure in particular, Sarah Palin, is often cited as a supporter, as in *A state legislator with the backing of Tea Party activists and Sarah Palin, the Indian-American woman overtook the state's old-boy network to trounce Rep. Gresham Barrett, a four-term congressman* (Lemmatized 12 MSN 20).

In the Fox corpus, the backing similarly refers to the Tea Party and Sarah Palin, but it also frequently refers to Democratic supporters, for example, *Specter, who won the backing of the White House and state Democratic leaders for switching parties* (Lemmatized 84 Fox 176).

Both networks use the acronym GOP to refer to the Republican party. But differences in the use of *the GOP's* suggests that MSNBC uses the connotative acronym GOP when trying to associate the Republican Party with a policy, event, or person.

No revealing pattern exists around the differential *the GOP's* ($p = .010$), but the diction of the differential has implications for the assessment of bias. The GOP is a synonym for the Republican Party. The acronym GOP stands for the Grand Ole Party, which connotes a traditional, well-entrenched, and elitist political establishment. Therefore, the differential *the GOP's* indicates a more connotative way of referring to the Republican Party. In many instances, *the GOP's* can be easily replaced with *the Republican Party's*. And the effect is slightly more neutral. For example, consider this sample sentence, *Democrats hope to persuade voters to view the elections as a choice between their party's recovery efforts and what they call the GOP's preferences to reward corporations and wealthy taxpayers (Lemmatized 34 MSN 40).* Now, compare that example with this altered version, *Democrats hope to persuade voters to view the elections as a choice between their party's recovery efforts and what they call the Republican Party's preferences to reward corporations and wealthy taxpayers (Lemmatized 34 MSN 40).*

In the Fox corpus, the GOP, without the possessive morpheme, appears frequently, but less frequently than in the MSBC, 32.6% opposed to 44.6%. However, with possessive *s*, the GOP's appears in only 1.9% of Fox texts, compared to 9.2% of MSNBC texts.

In midterm elections, political parties are vying to control the largest number of seats in Congress. For this reason, instances of *seat and* ($p = .001$) often occurred in speculative discussions of which party would control currently Democratic seats. For example, Rep. Stupak's retirement mobilized the Democratic establishment, as seen in *He said he's committed to helping Democrats retain the seat and that his announcement gives other Democratic hopefuls time to organize* (Lemmatized 62 MSN 7). Likewise, Sen. Bayh's retirement made analysts reconsider the likelihood of changing majorities in Congress, as seen in *The announcement gives Republicans a strong chance of capturing his seat and makes it likelier that the 59 votes that give Democrats command of the 100-seat Senate will dwindle* (Lemmatized_ MSN 1).

In the Fox corpus, there is only one instance of seat and, in a quote from Sen. Arlen Specter, *"I will run a vigorous campaign to retain this U.S. Senate seat and help all Democrats up and down the ballot"* (Lemmatized 196 Fox 4).

Instances of *race is* coincide with analysis speculating which party will win a race, generally a Democratic candidate. For example, an instance regarding polls reveals the use of *race is* in terms of predicting election results, as shown in *About four in 10 GOP primary voters say they are undecided, indicating the race is still open* (Lemmatized 16 MSN 24). A sample sentence exhibiting the differential *the district* also

contains *race is*, *The district […] usually votes so strongly Democratic that the race is rated Democratic favored* (Lemmatized 2 MSN 11).

In the Fox corpus, *race is* normally coincides with an assessment of the unpredictability of the race, as in *The Race is tight* (Lemmatized 46 Fox 140).

Greene. Two of the differentials pertain to the anomaly of political outsider Alvin Greene's victory in the South Carolina Democratic senate primary. The bi-gram *money and* is often in the tri-gram *no money and*, which contributes to the explanations as to why Greene's victory was so surprising. *The move upholds the improbable win by Greene, who raised no money and didn't even have a campaign website* (Lemmatized 18 MSN 26). Similarly, the differential *to file* relates to the curiosity of how and why Greene entered the race, as seen in *A Washington-based watchdog group wants South Carolina's attorney general to investigate if someone had paid Greene to file for the office* (Lemmatized 10 MSN 19).

MSNBC covered Greene's election in more depth that Fox. Greene is mentioned in 7.7% of the MSNBC corpus, opposed to 4.6% of the Fox corpus.

Modifiers. The differential *a moderate* ($p = .009$) tends to qualify Democrats more than Republicans. Instances of a moderate often occur in discussions of candidates' electability, for example *He said a moderate Democrat would have a good change* (Lemmatized 62 MSN 7) and *He's actually a conservative Democrat, which makes him a moderate, whish is what most Hoosiers are—moderate to conservative* (Lemmatized 54 MSN 59).

In the Fox corpus, *a moderate* generally described Democrats as well, appearing in the tri-gram *a moderate Democrat* several times. This similarity suggests that both networks find it important to label Democrats who are closer to the political center. This label may be worth mentioning in this election cycle, in which Republicans overtook Democrats in the House and nearly matched them in the Senate.

Likewise, the differential *first-term* ($p = .015$) modifies Democrats more than Republicans. Instances of first-term generally refer to first-term Democratic congressmen whose re-elections were unsure in the 2010 midterm elections, for example Concerns about Obama's health care plan and the mounting federal debt could ensnare two first-term Florida Democrats (Lemmatized 64 MSN 9) Also, note the use of the verb *survived* in the sentence And in the 8th district, first-term Democratic Rep. Larry Kissell survived a primary challenge from on of his former campaign volunteers (Lemmatized 60 MSN 64).

In contrast, *first-term* refers to Democrats and Republicans equally in the Fox corpus.

Possessive Pronouns. Two differentials that include possessive pronouns, *his decision* ($p = .023$) and *his democratic* ($p = .003$), generally refer to Democratic candidates. *His decision* mostly coincides with statements regarding a candidates decision to run or not run for office, for example *Gov. Freudenthal denied any impropriety, and said the prospect of his wife serving on the federal court wasn't the determining factor in his decision not to run again* (Lemmatized 17 MSN 25). In addition, the differential *his democratic* is most frequently part of the tri-gram *his democratic opponent*, as in *It was a startling development that could play a role in shaping the outcome of the race in November by siphoning voted from Paul to the benefit*

of his Democratic opponent (Lemmatized 61 MSN 65). The differentials are used similarly in the Fox corpus.

 Syntactic differentials. Some differentials do not reveal content themselves, but consistently appear throughout the corpus to outline certain relationships among information.

 Two differentials featuring conjunctions, *and an* (p = .003) and *and was* (p = .023) are most often used in lists describing candidates. These descriptions most often include political offices and ideological leanings, as in *Pedro Espada Jr., the senate majority leader and a Bronx Democrat* (Lemmatized 43 MSN 49), and highlights in the candidate's political career, as in *Haley has had to fend off a barrage of attacks, including allegations of infidelity and a racial slur from a state senator* (Lemmatized 50 MSN 55). Similarly, the differential *and was* is also generally used to connect highlights in a candidate's career, for example *If anyone has a slight edge, it's Marchall, who got the most votes in the May primary and was endorsed by MoveOn.org* (Lemmatized 50 MSN 55).

 The differential *close to* (p = .003) is often used in statements explaining the relationship between two people or between a person and a thing in a figurative sense, opposed to the physical meaning of the preposition. Examples of close to describing the figurative proximity of two people include *A democratic official close to Lincoln* (Lemmatized 59 MSN 63) and *a person close to Paterson* (Lemmatized 3 MSN 12). And instances of close to outlining the relationship between a person and a thing include *an incumbent this close to the primary election* (Lemmatized 38 MSN 44) and *A democrat close to the situation* (Lemmatized 52 MSN 57).

Finally, future promises are common in election discourse, voiced in the MSNBC sub-corpus by candidates through the collocation *I will* ($p = .010$). Notice the emphasis on the future in the example, *"I will work with Gov. Perry and our fellow Republicans to keep Texas strong in the future"* (Lemmatized 55 MSN 6).

Immigration

The political event that most impacted the Immigration corpora is Arizona Senate Bill 1070, or the Support Our Law Enforcement and Safe Neighborhoods Act. Overall, the Fox corpus focuses on immigration policy, border security, and immigration law. In contrast, the MSNBC corpus emphasizes the immigration policy, the impact of immigration on the U.S., and the experience of immigrants.

Fox. The differentials from the Fox corpus indicate a focus on border security and legal jurisdictions in immigration matters. Differentials related to border security emphasize lawlessness at the border and the need for more manpower to combat it. Differentials related to politics generally deal with the passage of Arizona Senate Bill 1070 and the history of immigration policy. References to the previous administration's policies normally involve the phrase *the Bush administration*, however MSNBC more often opts for *George W. Bush* instead, drawing attention to the president's lineage. A number of differentials deal with the legal issues regarding illegal immigration. In particular, the differential *enforcement to* generally serves to clarify which government agency has which legal responsibilities, on Fox. However, the one instance of *enforcement to* in the MSNBC corpus is concerned with the posturing of the Obama

51

administration. The syntactic differentials *will not* and *if you* generally are concerned

with the prevention of racial profiling in Arizona.

Table 5

Fox Immigration Corpus Differentials

Differentials	Fox texts	MSNBC texts	Significance
President Obama	77	3	$p < .001$
Fox News	52	0	$p < .001$
foxnews. com	42	0	$p < .001$
National Guard	31	4	$p = .368$
told Fox	38	0	$p < .001$
will not	42	5	$p = .233$
San Diego	27	5	$p = .817$
well as	41	7	$p = .563$
and border	38	3	$p = .060$
is to	39	5	$p = .307$
the time	40	6	$p = .427$
if you	41	6	$p = .428$
Mexico border	39	6	$p = .544$
told Foxnews.	31	0	$p = .003$
after a	35	6	$p = .675$
on Tuesday	40	3	$p = .061$

Table 5

Fox Immigration Corpus Differentials (continued)

Differentials	Fox texts	MSNBC texts	Significance
Social Security	29	6	$p = 1.000$
which is	34	4	$p = .276$
on Thursday	32	5	$p = .516$
to send	32	5	$p = .516$
the homeland	31	5	$p = .655$
a group	32	7	$p = 1.000$
the house	35	7	$p = 1.000$
all the	31	6	$p = 1.000$
Arizona Gov	–	–	–
enforcement to	34	1	$p = .011$
S.- Mexico	33	5	$p = .515$
customs and	31	1	$p = .017$
the program	24	4	$p = .802$
the Bush	30	6	$p = 1.000$
border protection	28	2	$p = .140$

Content-revealing differentials. The content-revealing differentials in the Fox

sub-corpus focus on border security, politics regarding immigration legislation, and legal

issues.

Border security. One tri-gram, *U.S.- Mexico border,* refers to the border

explicitly. Mentions of localities such as *San Diego* illustrate the seemingly lawless

situation along the border. Other differentials, *National Guard* and *Customs and Border*

Protection, indicate that the discussion of the border is largely focused on means of

controlling activity on the border. Bi-grams such as *to send* and flexigrams including

National Guard and a form of *to deploy* show that the most frequently mentioned

response to the situation is to reinforce military capabilities on the border. The network's

differing uses of the differential, *National Guard*, indicates a small difference in their

coverage. Fox generally refers to the need to send more troops, where as MSNBC focuses

on the administration's commitment to sending more, suggesting that the administration

has responded adequately to the situation.

Use of the bi-gram *National Guard* on Fox emphasizes the need for more border

security, whereas on MSNBC it is often used to illustrate that President Obama is dealing

with the security situation.

Instances of *National Guard* often coincide with discussions of sending troops to

the border. The bi-gram is commonly found in tri-grams such as *National Guard patrol*

and *National Guard troops*. There are even several instances of the hexa-gram *national*

guard troops to the border, such as *Sen. John McCain, R-Ariz., proposed an amendment*

to send 6,000 national guard troops to the border, a move Republicans want to pay for

with unspent stimulus money (Lemmatized 73 Fox 164). There are also several instances

of the tri-gram *deploy National Guard troops*. However, there are even more instances of

a flexigram featuring the verb *to deploy*, as in *After the shooting death of a rancher, Sens.*

Jon Kyl and John McCain once again called Monday for National Guard troops to be

deployed along Arizona's border with Arizona (Lemmatized 266 Fox 338).

In contrast, the MSNBC sub-corpus features *National Guard*, but not the hexa-

gram *National Guard troops to the border* nor a flexigram involving a form of *to deploy*.

However, there are several instances of the quad-gram *1,200 National Guard troops* in

the corpus. This quad-gram often refers to Obama's response to the national debate on

immigration and the border, as explained in *President Barack Obama's plan to send as*

many as 1,200 National Guard troops back to the U.S.-Mexico border quelled demands

that he must do more to battle illegal immigration and drug smuggling, but advocates for

tougher enforcement say the troops needs authority to make detentions (Lemmatized 96

MSN 97).

The quad-gram *Customs and Border Protection* figures prominently in the Fox

sub-corpus, but there is only one instance of the quad-gram in the MSNBC sub-corpus.

This difference suggests that Fox's coverage of federal security mechanisms was more

thorough than MSNBC's. In the Fox corpus, the differentials *customs and*, *and border,*

and *border protection* most often co-occur in the quad-gram referring to the department

of *Customs and Border Protection*. The quad-gram often modifies *agents* or *officers,* as

in *One man was arrested by Customs and Border Protection officers* (Lemmatized 242

Fox 316). Or it is modifying a person, for example *Steven Cribby, a spokesman for U.S.*

Customs and Border Protection (Lemmatized 254 Fox 327).

The tri-gram *U.S.- Mexico border* frequently appears in the Fox sub-corpus, often accompanied by a mention of violence, opposed to the MSNBC sub-corpus. The two differentials *U.S.- Mexico* and *Mexico Border* most often co-occur in the tri-gram *U.S.- Mexico border*. Many of the instances of the tri-gram can be found in the quad-gram *along the U.S.- Mexico border*, which often appears in a flexigram with the word *violence*, for example *Homeland Security Secretary Janet Napolitano named a former federal prosecutor Wednesday to the new post of "border czar" to oversee efforts to end drug-cartel violence along the U.S.-Mexico border and slow the tide of people crossing illegally into the United States* (Lemmatized 382 Fox 50).

In contrast, there are only two instances of the quad-gram *along the U.S.- Mexico border* in the MSNBC sub-corpus. In addition, there is only one instance of the tri-gram *U.S.- Mexico border* pertaining to violence, *The killing of the third-generation rancher has become a flashpoint in the immigration debate as politicians cite the episode as further proof that the U.S. must do more to secure the violent U.S.-Mexico border* (Lemmatized 51 MSNBC 56).

The direct objects of the differential *to send* are often figurative *messages* and law enforcement to the border. For example, an activist said of Arizonan sheriff Arpaio, *"We want to send him a message, that the hatred he spews is not wanted here"* (Lemmatized 295 Fox 364). However, the largest group of direct objects are law enforcement, as seen in *President Barack Obama on Tuesday promised to send 1,200 guard troops to the border to support efforts to block drug trafficking and temporarily supplement border patrol agents until more can be trained* (Lemmatized 316 Fox 383). Likewise, half the

instances of *to send* in the MSNBC sub-corpus refer to the deployment of National Guard troops to the U.S.-Mexican border, suggesting both networks use the word similarly.

The city of *San Diego* is frequently mentioned in the Fox immigration sub-corpus and rarely in the MSNBC sub-corpus. The context of these instances often refers to the city's role as a gateway between the U.S. and Mexico. Many instances of the differential coincide with stories of illegal crossings into the U.S., such as *He headed up a border crackdown that discouraged illegal crossings in the San Diego area* (Lemmatized 152 Fox 235). In the Fox sub-corpus, some instances highlight immigration-related crime in the area, as in *Castro pleaded guilty in November to murdering a federal officer near Camp, a mountainous area about 60 miles east of San Diego* (Lemmatized 278 Fox 349).

Politics. The political event that most marked the immigration corpus was the passing of Arizona Senate Bill 1070. In the Fox corpus, *on Tuesday* indirectly refers to the passing of the law. The differential *a group* sometimes refers to political groups coalescing around issues sparked by the debate in Arizona, as well as to groups of illegal immigrants. To provide background on the situation, both networks refer to the immigration policies of the previous administration. In the Fox corpus, such historical references often involved the differential *the Bush*, whereas MSNBC generally specifies *W. Bush*, possibly to discredit him as part of a political "dynasty." an "heir" to the presidency.

In the Fox sub-corpus, the differential *on Tuesday* refers to the timing of several news events, few of which are referred to repeatedly. It appears frequently in the Fox sub-corpus, but only a few times in the MSNBC sub-corpus. One reoccurring event was the passing of Arizona's immigration law, as noted in *The Arizona House on Tuesday*

approved a bill to crack down on illegal immigration by making it against state law to be

in Arizona without legal immigration documents (Lemmatized 431 Fox 96). In addition,

President Obama made several remarks concerning the legislation on Tuesdays, for

example *President Barack Obama on Tuesday warned of harassment against Hispanics*

under Arizona's tough new immigration law (Lemmatized 16 Fox 112). In contrast, there

are only three instances of *on Tuesday* in the MSNBC sub-corpus, and only one refers to

the Arizona law.

      In both corpora, the differential *the house* refers both to the House of

Representatives and personal homes. Instances referring to Congress generally refer to a

specific committee, such as *The House Homeland Security Committee, led by Rep.*

Bennie Thompson, D-Miss., was to examine the so-called 287(g) program in a hearing

Wednesday (lemmatized 168 Fox 25). The few references to homes generally involve

violence, for example *"They say they're either afraid to leave the house or they're*

moving away," said Tara McCollum Plese (Lemmatized 356 Fox 419). In the MSNBC

sub-corpus, only one instance refers to a home and it involves violence, *Crime scene tape*

was stretched across the street about 20 yards from the house (Lemmatized 76 MSNBC

79).

      The differential *a group* is generally followed by *of*, *that,* or a past participle. The

most common kind of groups are politically-oriented, as in *A group called BuyArizona*

has launched a website calling on supporters not only to ignore the anti-Arizona rhetoric,

but to make those cities leading the charge pay for their political opposition

(Lemmatized 49 Fox 142). The most commonly mentioned group is *illegal immigrants*,

for example *Another let an informant smuggle in a group of illegal immigrants*

(Lemmatized 207 Fox 285). In contrast, the MSNBC sub-corpus features only one instance of *a group of illegal immigrants*.

The differential *the Bush* often coincides with discussions of immigration policy during the Bush administration. These statements refer to *the last years of the Bush administration* and *immigration enforcement during the Bush administration*, for example *Workplace raids involving the arrests of hundreds of illegal immigrants at a time became almost routine in the last years of the Bush administration* (Lemmatized 114 Fox 200).

The MSNBC sub-corpus features *the Bush*, but more often refers to the previous president as George W. The Bush appears in 6.2% of the texts, whereas George W. Bush appears in 11.3% of the texts. In contrast, Fox refers to the Bush administration in 6.9% of texts, but only refers to George W. Bush in 3.9% of texts. Frequently noting the president's middle initial serves to distinguish his administration from his father's. This may be a subtle way of discrediting George W. Bush by differentiating him from his father George Bush Sr., suggesting that he was an undeserving "heir" to the presidency.

Legal issues. Differentials referring to the *Homeland Security Department*, *the program* established under Section 287(g) of the Immigration and Nationality Act, and *Social Security* are generally used in similar ways on both Fox and MSNBC. However, the differential *enforcement to* generally refers to interpretations of immigration law on Fox, but on MSNBC it deals with the image of the Obama administration.

The differing uses of the bi-gram *enforcement to* illustrate Fox's focus on the implementation of immigration law and display MSNBC's sympathy with the Democratic administration.

Discussions featuring the differential *enforcement to* generally focus on which law enforcement has which responsibilities regarding immigration law. Some of the instances are part of the tri-gram *Immigration and Customs Enforcement*, for example *Dozens of state and local agencies have partnered with Immigration and Customs Enforcement to carry out immigrant laws, the domain of the federal government* (Lemmatized 168 Fox 25). Other instances deal with the powers of local and state law enforcement, as in *President Obama has also questioned the legal authority for Arizona to enforce federal law, arguing it would be a violation of civil rights for state law enforcement to question the legality of an individuals residency* (Lemmatized 17 Fox 113).

In opposition, the MSNBC sub-corpus features only one instance of enforcement to. Instead of dealing with the division of legal authority, it pertains to the image of the Obama administration, *The Obama administration has been trying to build its credibility on immigration enforcement to boost the changes of passing an immigration reform bill in Congress* (Lemmatized 15 MSNBC 23).

In both sub-corpora, the differential *the program* generally refers to a system of cooperation between federal and local agencies to enforce immigration laws. The program was established by Congress in Section 287(g) of the Immigration and Nationality Act in 1996. Instances of the differential tend to refer to the difficulties in

executing the program, as in *As more agencies have signed on to the program, allegation of racial profiling and civil rights violations have risen* (Lemmatized 179 Fox 26).

In both the sub-corpora, the bi-gram *the homeland* is almost exclusively found in the quad-gram *the Homeland Security Department*, and the quad-gram is normally used in the same way by both news outlets. In the Fox corpus, instances of the Homeland Security Department often appear in statements regarding Immigrations and Customs Enforcement, a branch of the department. These instances are often concerned with investigations of immigrants working illegally in the country as in *Homeland Security Secretary Janet Napolitano has said investigations will focus on businesses that knowingly hire immigrants who cannot legally work in the U.S.* (Lemmatized 396 Fox 63).

In both sub-corpora, *Social Security* often refers to *social security numbers* or *social security cards*. These are mentioned frequently as proof of American citizenship, as in *Florida has a "no match, no vote" law requiring driver's license and Social Security numbers on voter registration applications to be verified against state databases* (Lemmatized 112 Fox 2). In some cases, they are referred to in cases of faking citizenship, as in *The Supreme Court agreed Monday to decide whether people picked up on immigration violations also can face charges of identity theft if they use social security ad other identification numbers to belong to others* (Lemmatized 1 Fox 1).

Syntactic differentials. Some differentials mean very little in isolation. These I refer to as syntactic differentials, because they are useful in finding patterns in the way ideas are connected within the corpora. The differentials *will not* and *if you* often facilitate statements concerned with the prevention of racial discrimination.

A number of instances of *will not* can be attributed to a frequently quoted statement from Sen. Lindsey Graham. In the quote, Republican Sen. Graham criticizes Harry Reid for reprioritizing immigration legislation over climate change legislation, *"But I will not allow him to play one issue off of another, and neither will the American people. They expect us to do both, and they will not accept the notion that trying to act on one is an excuse for not acting on the other"* (Lemmatized 10 Fox 107). This frequently reproduced quote accounts for most of the instances of some other frequent tri-grams such as *I will not* and *will not allow*.

Otherwise, the bi-gram *will not* most often appeared in a flexigram including *be* and a past participle, notably *will not be tolerated.* Every instance of the quad-gram relates to racial profiling and is found in quotes from Arizona Gov. Jan Brewer of Arizona, for example, *"These new statements make it crystal clear and undeniable that racial profiling is illegal, and will not be tolerated in Arizona,"* she said in a statement (Lemmatized 30 Fox 125).

The differential *if you* is often employed to make readers consider the perspective of American Hispanics. In some cases, the perspective is more general, as in a quote from Obama, *"But you can imagine if you are a Hispanic American in Arizona, your great great grandparents may have been there before Arizona was even a state"* (Lemmatized 21 Fox 117). In other cases, the bi-gram is used in statements explaining the possible

impact of the Arizona immigration law, for example, *"If you look or sound foreign, you are going to be subjected to never-ending requests from police to confirm your identity and to confirm your citizenship"* (Lemmatized 12 Fox 113). The remaining instances of *if you* in the corpus follow typical uses of the second person to posit hypothetical statements regarding a wide range of issues.

One of the reoccurring instances of *is to* is in the tri-gram *goal is to*. It normally appears in quotes where the speaker articulates the objectives of a group, for example *"Our goal is to raise the banner of comprehensive immigration reform," Gutierrez said* (Lemmatized 157 Fox 24). *Goal is to* does not occur in the MSNBC sub-corpus.

Most of the instances of *the time* occur in tri-grams such as *at the time* and *by the time*. This trigram suggests that the time is mostly used in temporal prepositions and bare little semantic load, for example *Carrillo said his brother was home at the time* (Lemmatized 148 Fox 231). However, Bell (1991) notes the prevalence of such time expressions in news.

MSNBC. The differentials from the MSNBC sub-corpus suggest that the network's coverage of immigration issues focuses on the changing demographics of the country, the experience of immigrating, and immigration policy. Overall, MSNBC overwhelmingly contextualizes general discussions of immigration in terms of place. When covering the immigrant experience, MSNBC often mentions *a better life,* explaining the motivation of immigrants to come to the U.S., an issue that does not figure among the Fox differentials. The differential *work and* on MSNBC is generally accompanied by the rewards of work, whereas it is generally accompanied by discussions of immigrants' contributions to the state on Fox. There are also differences in the way the

two networks discuss immigrants' *family members*. When covering immigration policy,

MSNBC prefers to refer to the previous president as *George W. Bush*, whereas Fox

prefers the metonymy of *the Bush administration*. Also, MSNBC portrays *ICE agents*

negatively, in contrast to Fox, which occasionally describes them as effective.

Table 6

MSNBC Immigration Corpus Differentials

Differentials	Fox texts	MSNBC texts	Significance
her husband	12	11	$p < .001$
look at	10	9	$p = .004$
and its	13	11	$p = .001$
director for	12	10	$p = .002$
their home	13	8	$p = .037$
a better	11	9	$p = .004$
the community	14	10	$p = .005$
ICE agents	10	8	$p = .008$
learn English	12	8	$p = .017$
and new	12	8	$p = .017$
in America	6	8	$p = .001$
of two	12	9	$p = .007$
town of	11	8	$p = .012$
W. Bush	17	11	$p = .009$

Table 6

MSNBC Immigration Corpus Differentials (continued)

Differentials	Fox texts	MSNBC texts	Significance
already in	13	8	$p = .037$
family members	9	8	$p = .005$
supporters of	9	8	$p = .005$
will continue	10	9	$p = .003$
her case	8	7	$p = .010$
the fact	7	8	$p = .002$
460, 000	6	8	$p = .001$
for legal	13	8	$p = .037$
has an	10	8	$p = .008$
immigration in	10	7	$p = .022$
legislation that	12	9	$p = .007$
of race	7	8	$p = .002$
see if	11	8	$p = .012$
system is	11	8	$p = .012$
there's no	9	8	$p = .005$
work and	9	6	$p = .039$
year in	7	6	$p = .018$

Content-revealing differentials. The content-revealing differentials include those that refer to the impact of immigration, the immigrant experience, and legal issues regarding immigration.

Immigration in America. A theme among the MSNBC differentials is the subject of immigration's impact on the country. Differentials like *immigration in* ($p = .022$) and *in American* ($p = .001$) address the general issue and relate it to a place, leading to more comprehensive discussions of immigration. And a number of differentials like *the community* ($p = .005$) and *town of* ($p = .012$) refer to localities affected by immigration. Overall, the MSNBC sub-corpus focuses on the changing demographics engendered by immigration in terms of place.

The more frequent use of *immigration in* suggests that MSNBC pays more attention to the entire issue of immigration, instead of honing in on specific issues such as border security like on Fox News.

Immigration in ($p = .022$) is often found in a flexigram where the differential is followed by a locative noun, particularly *America* and *Arizona*. Immigration is a large topic and discussing the expansive issue in terms of a place lends itself to global discussions. In the flexigram, *immigration* is occasionally preceded by *illegal*, for example *Yet many are wary of negative reactions prompted by heightened attention to illegal immigration in Congress and the presidential campaign trail, and they increasingly cite discrimination as a problem* (Lemmatized 19 MSN 27).

The differential also appears in the Fox corpus and is found in the same kind of flexigram. However, *immigration in* occurs in 8.2% of the MSNBC texts, where as the differential appears in only 2.3% of the Fox News texts. This difference suggests a lack of coverage of the issue in terms of place.

In America (p = .001) is often preceded by the word *stay* or references to illegal immigrants. Instances of *stay in America* are generally in quotes, such as *Wong said Onyanho is hopeful the judge will rule in her favor because "she really wants to stay in America"* (Lemmatized 8 MSN 17). The second use of the bi-gram *in America* is a flexigram where the bi-gram is preceded by references to immigration, such as *illegal immigrant*, *immigration*, and *undocumented Mexican workers*, for example *After the 9-11 attacks, in the name of national security, the Bush administration announced it wanted to detain, and then deport, every illegal immigrant in America* (Lemmatized 66 MSNBC 7).

In the Fox corpus, *in America* occurs in 1.4% of texts, opposed to 10.3% of texts in the MSNBC corpus. In those texts, there is only one example of *stay in America*, *She will do whatever she can do to fight for the privilege to stay in America," she said* (Lemmatized 2 Fox 10). The other instances do not follow the patterns of the MSNBC corpus. Some instances refer to immigrants and others refer to events with a national scope, as in this example about the census *It's required by the Constitution that every person residing in America be included* (Lemmatized 154 Fox 237). These examples suggest that the Fox corpus refrains from drawing a direct connection between illegal immigrants and their host country. As seen above, Fox tends to first associate illegal immigrants with the border.

Local impact. Several differentials from the MSNBC sub-corpus localize the discussion of immigration. This localized discussion tends to be more descriptive than Fox's, which tends to emphasize issues of legality and security.

The community refers to the presence of immigrants on MSNBC, and Fox's usage of the differential focuses on the illegal nature of their presence. In addition, both networks use *the community* in discussions of policy, but Fox emphasizes policies' impact on security.

In the MSNBC corpus, instances of *the community* ($p = .005$) generally are part of discussions of how a locality has been affected by immigration or immigration policy. Some instances regard the influx of immigrants, such as *The community about 35 miles Northwest of Omaha has seen its Hispanic population surge in the past two decades* (Lemmatized 47 MSNBC 52). In the Fox corpus, some instances refer to the presence of immigrants, but Fox tends to emphasize that their presence is clandestine, as in *Immigrant advocates say many illegal workers who avoided the sweep remain in the community but lead underground lives* (Lemmatized 161 Fox 243).

In the MSNBC corpus, some instances of *the community* concern measures designed to prevent illegal immigration, as in *Bizzell helped Dole broker the expansion, and said at the time the purpose wasn't to begin a wholesale roundup of illegal immigrants but to help the community* (Lemmatized 69 MSNBC 72). In the Fox corpus, instances similarly refer to policies, but emphasize the policies as deterrents for violence. The emphasis on violence can be seen in the example *Stephens says he wants to protect the community from "criminal elements" and not "people who want to earn an honest buck"* (Lemmatized 337 Fox 401).

The bi-grams *and new* (*p* = .017) and *town of* (*p* = .012) both refer to places in the MSNBC and Fox corpora.

In the MSNBC corpus, *and new* is often part of a list of geographic places ending with *New York*, *New Jersey*, or *New Mexico*, for example *In 2008, California had the most illegal immigrants at 2.7 million, double its 1990 number, followed by Texas, Florida, New York and New Jersey* (Lemmatized 21 MSNBC 29).

Town of is most often followed by the name of the town or its population. Many of these instances are preceded by a determiner followed by a locative noun, for example *the Western Arizona town of Kingman* (Lemmatized 55 MSNBC 6). Examples of the bi-gram introducing population counts include *a town of 90,000* (Lemmatized 28 MSNBC 35) and *a town of 18,000* (Lemmatized 66 MSNBC 7).

The subject of the differential *has an* (*p* = .008) is normally a locative noun. However, it is most likely a differential because of the recycling of a phrase *Arizona has an estimated 460,000 illegal immigrants*, which appears in several texts such as Lemmatized 46 MSNBC 51 and Lemmatized 23 MSNBC 30. In the Fox sub-corpus, the phrase also appears several times, suggesting that this article came from a wire service such as the Associated Press.

The differential *460, 000* (*p* = .001) is probably a differential for the same reason. The figure is an estimate of Arizona's population of illegal immigrants, as explained in *The state is home to an estimated 460,000 illegal immigrants—a population larger than that of entire cities such as Cleveland, St. Louis and New Orleans* (Lemmatized 57 MSNBC 61).

Immigration experience. Another theme among the differentials is the

immigration experience. Some of these differentials such as *a better* (p = .004) refer to

the reasons people choose to immigrate. The differential *work and* (p = .039) is often

associated with the fruits of labor in the MSNBC corpus, whereas it is often discussed in

terms of immigrants' contributions to the government in the Fox corpus. Others such as

her case (p = .010) deal with the legal issues immigrants face once in the country.

Finally, there are differentials like *family members* (p = .005) associated with the familial

dynamics involved in immigrating.

    Motivation. In the MSNBC sub-corpus, the differential *a better* (p = .004) is

generally found in the tri-gram *a better life*, but is rarely used in the Fox sub-corpus.

Instances of the tri-gram normally are part of anecdotes highlighting a specific immigrant

family's motivations for immigrating, for example *"Mexico doesn't have a lot of*

opportunities," she says. "Here, we work honestly, and we have a better life"

(Lemmatized 72 MSNBC 75). Note the mention of children in the following example,

She said she was happy living in the United States—a place that offered a better life for

her and her three sons (Lemmatized 58 MSN 62). Overall, *a better life* appears in 5.2%

of the MSNBC sub-corpus, opposed to 0.5% of the Fox sub-corpus.

    In the MSNBC sub-corpus, *work and* (p = .039) often appears in flexigrams

accompanied by the fruits of labor, whereas it is often in flexigrams accompanied by

contributions to the state in the Fox sub-corpus. Generally the flexigram appears in

discussions of why immigrants want to be in the U.S., for example, *We're going to focus*

on those people who choose to pursue a life of crime in the United States rather than

pursue the American dream of education, hard work and success (Lemmatized 4

MSNBC 13). In contrast, instances of the differential in the Fox sub-corpus often appear in flexigrams where the bi-gram *work and* is accompanied by some form of contribution to the country, for example a quote where a woman refers to *people at gas stations who are looking for work and not paying taxes* (Lemmatized 173 Fox 254).

   Legal issues. In both sub-corpora, *her case* ($p = .010$) normally appears in personal interest stories concerning a woman in a legally sensitive immigration situation, in particular a case involving Obama's aunt. In the MSNBC sub-corpus, several instances deal with the case of Obama's aunt seeking asylum in the U.S., as shown in *In November, Onyango said she never asked Obama to intervene in her case and didn't tell him about her immigration difficulties* (Lemmatized 8 MSNBC 17). In the Fox sub-corpus, many of the instances of her case are also in articles about Obama's bid for asylum.

   In both sub-corpora, *already in* ($p = .037$) is generally found in flexigrams where the differential is preceded by references to immigrants and followed by a reference to the U.S. In the MSNBC sub-corpus, references to immigrants can be general, such as *immigrants*, or specific, like *Haitians*. References to the U.S. include *the United States*, *the U.S.*, and *the country*. In general, these instances coincide with the legality of immigrants' residency or plans for their possible naturalization, for example *65 percent support allowing undocumented immigrants who are already in the country to pay a fine, learn English, and go to the back of the line for the opportunity to become U.S. citizens* (Lemmatized 18 MSNBC 26). Similar examples can be found in the Fox sub-corpus, like *And 65 percent said enforcing the border was paramount to legalizing the status of illegal immigrants already in the U.S.* (Lemmatized 427 Fox 92).

Family. The top differential in the corpus is *her husband* ($p < .001$). It is the top differential because it appears repeatedly in a single article, therefore, it is an anomalie. However, the articles does comment on the situation of immigrants. It is about foreign woman with American husbands who die before they get married. As a result, the women are denied citizenship. The situation is summarized in the example sentence, *Had they done it then, Irina would have been married more than two years by the time of her husband's death and allowed to stay in the United States–or at least allowed to file for residency on her own behalf* (Lemmatized 74 MSNCS 77). There are many instances of her husband in the Fox sub-corpus, but they refer to a variety of personal interest stories.

Family members ($p = .005$) generally refers to relatives wanting to immigrate to the U.S. and the help provided by relatives. In both sub-corpora, the differential is used to refer to specific people's relatives, such as *Krentz's family members declined requests for comment* (Lemmatized 430 Fox 95), but this use is unrevealing in regard to assessing bias. In both sub-corpora, some instances discus family wanting to join relatives who are already nationalized, for example *The provisions concerning same-sex couples are part of legislation that would increase the number of visas provided to family members of people already in the U.S. legally* (Lemmatized 44 MSNBC 5). In the MSNBC corpus, other instances describe relatives as parts of a support network, such as *They're just calling different family members in different states and asking around about work* (Lemmatized 38 MSNBC 44). However, the Fox sub-corpus rarely uses the differential in this way, suggesting that Fox focuses more on immigrant's familial ties as advantageous when applying for citizenship.

Immigration policy. Many of the differentials pertain to immigration policy. Some refer to current policies, with an emphasis on government entities such as Immigration and Customs Enforcement. Others refer to new or proposed legislation. Proposed legislation tends to focus on comprehensive immigration reforms. And the legislation with the highest impact on the corpus is the Arizona immigration law.

Past and present policies. Both networks examine the problems in current immigration policy. Both say that the *system is* inadequate. For example, both make an issue of the behavior of *ICE agents*. MSNBC generally portrays then negatively, whereas Fox sometimes depicts the agency as effective. Both explain the current state of immigration policy by referring to the previous administration. *W. Bush* is a differential in the MSNBC corpus, whereas *the Bush administration* is a differential in the Fox corpus, attributing the president different levels of credibility.

In the MSNBC sub-corpus, instances of *system is* (*p* = .012) refer to the system of verifying immigrants status and the system of naturalizing immigrants. In both sub-corpora, most instances provide negative assessments of the system. Frequently the differential appears in the quad-gram *the immigration system is*, as in *The immigration system is undeniably broken, he said* (Lemmatized 33 MSNBC 4). In the FOX sub-corpus reveals *system is* often appears in a flexigram involving the word *broken*, as in *She says the immigration system is severely broken* (Lemmatized 173 Fox 254).

In the MSNBC sub-corpus, *ICE agents* (*p* = .008) are often portrayed as aggressive and as overstepping their legal bounds; they are portrayed similarly in the Fox sub-corpus, but are occasionally depicted as effective law enforcement. ICE is the acronym for the Immigrations and Customs Enforcement.

In the MSNBC corpus, the authority of ICE agents is the subject surrounding several instances of the differential, for example, *Since then, there have been constant disputes over what authority the agreement gives ICE agents* (Lemmatized 25 MSNBC 32). Other instances portray the agents as poorly performing their duties or abusing their authority, as in a *Pentecostal minister who was forcibly drugged by ICE agents after he asked to contact his wife* (Lemmatized 67 MSNBC 70) and *ICE agents have blundered badly in their dealings with informants and other sources, covering up crimes and even interfering in a police investigation into whether one informant killed another* (Lemmatized 9 MSNBC 18).

In the Fox sub-corpus, there are many instances of ICE agents being portrayed negatively. However, there are a few instances that presume that the agency functions properly, as in *They must check in regularly with ICE agents and were given background checks, she said* (Lemmatized 367 Fox 43).

Instances of *W. Bush* ($p = .009$) are always in the tri-gram *George W. Bush* and correspond with explanations of the history of immigration policy, for example *Immigration prosecutions rose to record levels in 2009 as the Obama administration kept up aggressive enforcement that began under president George W. Bush* (Lemmatized 2 MSNBC 11). The differential corresponds to similar discussions in the Fox sub-corpus. However, Fox refers to the Bush administration more, preferring not to highlight that he was the second president George Bush.

The differential *their home* ($p = .037$) is normally found in either the tri-gram *their home country* or *their home countries*. Most of the instances of these tri-grams involve deportation, as in *"Those who come to the United States to prey upon our*

74

neighbors and communities will be prosecuted for their crimes and ultimately returned to their home countries," said John Morton (Lemmatized 61 MSNBC 65). Another example is *Instead of turning juveniles over to immigration officers, San Francisco was housing them or flying them back to their home countries at city expense* (Lemmatized 89 MSNBC 90). The differential is used similarly in the Fox sub-corpus.

The bi-gram *for legal* (*p* = .037) is typically in the tri-gram *for legal immigrants* and in the quad-gram *Americans for Legal Immigration*. An example regarding the Arizona law is *It also makes it a state crime for legal immigrants to not carry their immigration documents and bans day laborers and people who seek heir services from clocking traffic on streets* (Lemmatized 94 MSNBC 95). The latter is a North Carolina-based political group, *the Americans for Legal Immigration political action committee, which favors tighter immigration controls* (Lemmatized 95 MSNBC 96). The differential is used similarly in the Fox sub-corpus, including the reoccurring references to Americans for Legal Immigration.

The differential *director for* (*p* = .002) is generally followed by the name of an organization and preceded by details of the job title, for example *the assistant field office director for ICE's Detention and Removal operations* (Lemmatized4 MSNBC 13). The differential appears in the same way in the Fox sub-corpus.

Proposed policies. In the MSNBC sub-corpus, *legislation that* (*p* = .007) is often followed by *would*, indicating that most discussions involving the differentials are about hypothetical legislation, for example *Obama is pressing lawmakers to take up legislation that would deal with border security, employment and citizenship* (Lemmatized 56 MSNBC 60). The differential is often used similarly in the Fox sub-corpus.

There's no (p = .005) often appears in a flexigram along with *way*. These instances generally are part of discussions of political possibilities, for example *"They want the immigrants removed—but how should they be sorted out from the approximately 40 million Hispanics who are U.S. citizens? There's no easy way," said Roger Clegg* (Lemmatized 63 MSNBC 67). In the Fox sub-corpus, the differential is not found in a flexigram.

Nationally. In both sub-corpora, *Learn English* (p = .017) is used in slightly different ways. In the MSNBC sub-corpus, it appears almost exclusively in lists of steps illegal immigrants would have to complete in order to become citizens under proposed immigration legislation. The bi-gram is often accompanied by the tri-gram pay a fine, as in *He said they must pay a fine and taxes, learn English and "go to the back of the line" of people trying to enter the United States from their home country* (Lemmatized 5 MSNBC 14).

In the Fox sub-corpus, Learn English similarly appears as a requirement in a proposed method for naturalization. However, there are also several instances of learn English in the context of immigrant children trying to excel in school, as in *"It's my job to make sure they're taught English in the most rigorous possible way so they can learn English quickly, can compete with their peers, and succeess academically," Home told Fox News* (Lemmatized 69 FOX 160).

Arizona. Some differentials referred most often to the Arizona law. In both sub-corpora, *supporters of* (p = .005) often appears in the quad-gram *supporters of the law*. This law is generally Arizona's immigration law, for example, *Supporters of the law say*

it is a commendable effort to combat what is fast becoming a scourge in the U.S. gateway

for illegals (Lemmatized 57 MSNBC 61).

In both sub-corpora, there are several instances of *of race* ($p = .002$) that deal with

the issue of racial profiling, particularly in any laws directed at identifying illegal

immigrants, for example *Supporters have dismissed concerns of racial profiling saying*

the law prohibits the use of race or nationality as the sole basis for an immigration check

(Lemmatized 29 MSNBC 36).

See if ($p = .012$) is probably a differential because it is featured in a phrase that

was recycled in several texts. The phrase is *Obama called the Arizona bill "misguided"*

and instructed the Justice Department to examine it to see if it's legal. It appears in the

texts Lemmatized 23 MSNBC 30, Lemmatized 29 MSNBC 36, and Lemmatized 46

MSNBC 51. In the Fox sub-corpus, there are four instances of the same phrase and

several instances that use the differential in a similar way.

Syntactic differentials. The two in the bi-gram *of two* ($p = .007$) generally

functions as a determiner with an object and occasionally as a determiner with an ellipsis.

Examples of the determiner function are *same sex couples of two nationalities*

(Lemmatized 44 MSNBC 5)), *a period of two months* (Lemmatized 42 MSNBC 48), and

enforcement of two state laws Lemmatized 24 MSNBC 31). However, two can stand

alone when followed by an ellipsis, as in *a soft-spoken mother of two* (Lemmatized 62

MSNBC 66), compared with the determiner function of *a father of two daughters*

(Lemmatized 70). Note that several instances of the differential refer to family. In the Fox

sub-corpus, the differential is rarely followed by a relative.

The differential *and its* (p = .001) is generally immediately precedent by its

antecedent. These referents tend to be groups or places, for example, *The new war is*

between the country's white majority and its burgeoning Hispanic population

(Lemmatized 75 MSNBC 78).

Look at (p = .004) is overrepresented because two texts are basically the same,

save one sentence that is different. For this reason the two texts were not identified during

cleansing.

Discussion

This study assesses bias in terms of selection bias and biased language. Both

methods use indicative linguistic features, or differentials, as the main unit of analysis.

However, it is important to remember that these differentials are not necessarily valid,

since the corpora are unequal and inconsistent. Nevertheless, they offer insight into the

reporting of elections and immigration policy from Fox and MSNBC, because they are

the most frequently used collocations in the news cycles covered by the corpora.

Selection Bias

A news outlet exhibits selection bias when it publishes articles that coincide with

its ideologies. In this way, news outlets prioritize news events, giving preference to some

stories over others. Therefore, selection bias has the effect of impacting what the

audience considers important. In van Dijk's (1988) framework for discourse analysis,

selection bias creates a macroproposition, because it sends a global message, over a

number of texts, that an issue is more important than another.

In the case of political news, the news outlet has the power to influence what

political issues deserve attention. It also has the power to familiarize an audience with

one side of a debate more than another, making readers sympathize with the news outlet's stance on the issue.

This study limits its sample of articles to a set timeframe in order to assess selection bias. I can identify the events each network deemed newsworthy by measuring how frequently they discuss certain events. Normally, this type of analysis is restricted to the article level; however, I can measure overall mentions of an issue at the sentence level by analyzing linguistic features, instead of entire texts.

Elections

Fox. Fox chose articles that create the macroproposition that Democratic incumbents will pay for their associations with the Obama administration and its legislative agenda, which the voters disagree with.

Frequent mentions of *President Obama* attest to the focus on the administration in Fox's election coverage. Likewise, frequent mentions of *Harry Reid* remind readers of the bills that moved through the 111th Congress. At the same time, *Joe Sestak* and *Blanche Lincoln* are examples of candidates whose electoral fates were to be determined by their associations with the administration and its agenda. The tri-gram *the former Rep*. most often modifies the candidates J.D. Hayworth, Pat Toomey, and Tom Campbell, who were all in tight races. Moreover, extensive coverage of *Sarah Palin*'s support for challengers to incumbents gives voice to the resistance against the Democratic Congress and White House that lead to the Republican gains in the House of Representatives after the midterm elections in 2010.

The network associates this voter backlash to frustration with legislation passed during the 111th Congress. *The bill* and *the law* are used nearly synonymously in the

corpus, and both are differentials, indicating that legislation was a major focal point of Fox's election coverage. The network articulates voter's perceptions of the bills through polls, which are cited so frequently that three related bi-grams figured among the differentials, *say they*, *margin of*, and *the latest*.

Together, these differentials manifest Fox's conservative selection bias, because they suggest that Fox chose to publish articles dealing with contentious legislation passed by the Democratic Congress, voter's frustration with the legislation, and races that could be affected by these sentiments.

MSNBC. Instead of focusing on the motivation behind voter behavior, MSNBC chose articles focused on the election strategies of candidates, often sympathizing with Democrats. For example, frequent use of *the district*, *challenge from*, and *a senate seat* deal with political maneuvers and considerations that affect campaigns. Attentnio is further put on politicians through the frequent use of personal pronouns in the differentials *his decision*, *his democratic, lost his,* and *his candidacy*. This focus on election factors reflect a concern for Democratic incumbents, because the Democrats were defending their majorities in Congress in the 2010 midterm elections.

The liberal bias of MSNBC's article selection was more evident in differentials concerning the Republican Party. For example, *the backing* refers almost exclusively to the support of Republican candidates. Likewise, MSNBC also emphasizes *the GOP's* ownership of issues through frequent use of that synonym for the Republican Party along with a possessive morpheme. In addition, instances of *seat and* and *race is* generally dealt with the future of seats held then by Democrats.

Immigration

Fox. Fox chose articles that create the macroproposition that more bureaucratic, military, and legal enforcement is needed to secure the border. Coupled with the differential *U.S.-Mexico border*, frequent mention of the *National Guard*, *Customs and Border Protection, the Homeland Security Department, and program 287(g)* focus on the federal tools for securing the border. The differential *to send* indicates that Fox gave coverage to calls for deploying more law *enforcement to* the border. Mentions of *the Bush administration* refer to previous policies to addressing border issues. And *on Tuesday* refers to the Arizona Senate Bill 1070, an example of policies passed during this administration to deal with the border.

MSNBC. MSNBC chose articles that create the macroproposition that immigrants form a substantial part of the population and should be sympathized with; therefore, policies to deal with immigration, both legal and illegal, should be reformed. Frequent mention of *immigration in America* emphasizes the number of illegal immigrants *already in* the country. Coverage of the issue is localized through mentions of *the community* and a *town of* growing immigrant population. MSNBC also chooses articles that explain the lives of immigrants, such as their motivations for *a better* life in the U.S., where they can find *work and* other opportunities. MSNBC further encourages readers to sympathize with immigrants by choosing articles that refer to their *family members*. Overall, the message is that the immigration *system is* inadequate and gives law enforcement, such as *ICE agents,* too much latitude. To fix the deficient policies put in place by previous administrations, including that of *George W. Bush*, the country needs *legislation that* is

81

fair *for legal immigrants* as well as illegal immigrants, who could be called upon to

follow steps to naturalization such as *learn English*.

Biased Language

To assess bias, this study compares differing ways the networks refer to the same

issues and investigates linguistic opportunities for bias outlined by theorists. As explained

by van Dijk (1988), discourse inherently involves incompelte representations of reality.

This study examines the networks' ways of representing the reality of current events with

differing word choices. Bell (1991) describes some of the characteristics of news

discourse, such as quotes, sources, and setting, that open reporting to bias.

Differing uses of words. As van Dijk (1988) notes, reality must be interpreted

through words in order to form discourse. Interpretations are inherently subjective. And

the way the two networks interpret reality differently is evident in the different words

they choose for the same concepts and the different concepts they associate with the same

words.

Similar differentials. Differing bi-grams for the same concepts include *President

Obama* and *the White House*, *the Bush administration* and *George W. Bush*, and *the

Republicans* and *the GOP's*.

President Obama figures prominently in both Fox corpora, whereas he is not a

differential in either MSNBC corpora. Furthermore, he is hardly mentioned at all in the

MSNBC election corpus.

President Obama is a pan-corpora differential in the Fox corpus, appearing in

both the election and immigration corpora. In the election corpus, *President Obama's* is a

differential as well as *President Obama*. Together the two are used to emphasize the

administration's ownership of the contentions policies behind the mid-term election debates. Overall, Fox seems to depict the administration as the origin of divisive policy maneuvers.

In the MSNBC corpus, the President is oddly absent from the election corpus. There are only two instances of *President Obama* and two instances of *the Obama administration*, together present in only 6.2% of texts. However, the administration is far more frequently referred to as the *White House*. Referring to the president through metonymy serves to remove his name from the discourse. This word choice suggests that MSNBC considered associations with the administration detrimental to Democrats' re-election bids and thus refrained from explicitly linking Obama's name with particular policies or elections.

In coverage of the immigration debate, Fox refers to the previous administration most frequently with the tri-gram *the Bush administration*, whereas MSNBC most frequently mentions *George W. Bush*. Fox prefers the Bush adminstration, which is more respectful and official, since it recognizes that Bush was the head of a government. On the other hand, *George W. Bush* is technically correct, but emphasizes the president's junior status, compared to his father. This may be a subtle way of discrediting George W. Bush by differentiating him from his father George Bush Sr., suggesting that he was an undeserving "heir" in a political "dynasty."

References to the opposing party also manifest the networks' biases. In the election corpus, Fox often refers to *the Democrats* as a monolithic bloc, ignoring differences within the party. Fox and MSNBC mention *the Democrats* almost equally, in 13.4% and 15.4% of the corpora respectively. But MSNBC refers to *the Republicans* in

12.3% of texts, much more than Fox, 5.7%. This difference suggests that MSNBC devoted more attention to Republicans than Fox devoted to Democrats. Furthermore, when emphasizing the Republican's ownership of something, MSNBC prefers the differential *the GOP's*, opting for a more connotative synonym for the party.

   Issues covered differently. In election coverage, Fox and MSNBC both referred to *the backing, the bill* and *the law, and Sen. Blanche Lincoln*, but the networks used these terms in ways that reinforce their own political positions. For example, discussions of *the bill* and *the law* tend to refer to the voting public's perceptions of contentious legislation, whereas MSNBC focuses on bills' passage and contents, downplaying the controversial behavior of the Democratic Congress. Showing more divergence, MSNBC frequently refers to *the backing* of Republican supporters, whereas Fox refers to *the backing* of both Republican and Democratic supporters. This difference suggests that MSNBC devotes more attention to explaining the support for Republican candidates than Democratic candidates. One candidate in particular received differing coverage: *Sen. Blanche Lincoln*, whose blue-dog Democrat credentials were emphasized on Fox, and who was depicted as a natural target of anti-incumbent sentiment on MSNBC.

   In coverage of the immigration debate, Fox and MSNBC both use the phrases *enforcement to*, *National Guard*, *ICE agents, immigration in, in America,* and *the community,* but in ways that coincide with their political ideologies.

   The differing uses of the bi-gram *enforcement to* illustrate Fox's focus on the implementation of immigration law and display MSNBC's sympathy with the Democratic administration. Likewise, use of the bi-gram *National Guard* on Fox emphasizes the need for more border security, whereas on MSNBC it is often used to

illustrate that President Obama is dealing with the security situation. Both networks are critical of the Immigration and Customs Enforcement, but MSNBC almost exclusively portrays *ICE agents* as aggressive and as overstepping their legal bounds; however in the Fox corpus they are occasionally depicted as effective law enforcement.

In contrast, MSNBC's more frequent use of *immigration in* suggests that MSNBC coverage of immigration is more comprehensive, instead of honing in on specific issues such as border security like on Fox News. MSNBC frequently addresses the issues regarding life *for illegal* immigrants *in America*, especially their desire to stay in America. In contrast, Fox rarely uses *in America*, refraining from drawing a direct connection between illegal immigrants and their host country. Overall, the differences in coverage of immigration is exemplified by the differential *the community*, which refers to the presence of immigrants on MSNBC, and the illegal nature of their presence on Fox.

Characteristic Features of News Discourse

As Bell (1991) notes, news discourse has certain characteristic features. Many of these features, like citing sources, quoting sources, and providing statistics, are typical of news, because they substantiate the claims and credibility of the news report. He also suggests that these features are opportunities to inflate the credibility, factuality, and objectivity of the report's contents. Both Fox and MSNBC manipulate these features to serve their own political agendas.

Sources. Bell explains that the quality of an article's sources influences the credibility of the article. He gives professors as examples of credible sources, whose commentary can clarify and illuminate situations. MSNBC often cites professors, as

indicated by the differential *professor at*. This dependence on professors for commentary indicates an effort to give a credible air to commentary in MSNBC's reporting.

Bell also notes that politicians and officials are often quoted as sources, even when they are not experts on an issue. However, they are frequently quoted because they are high status sources, and high status sources can give legitimacy to an article. After many quotes, Fox often explains that politicians *told Fox News* or *told FoxNews.com* their comments. This seems like an attempt to boost the credibility of not only the article, but the network as a whole.

Quotes. Bell explains that quotes are an opportunity for journalists to introduce explicit opinion and uncharacteristic features into news discourse. In the Fox corpus, *Harry Reid*, *if you*, and *I will* appear most often in quotes. Mentions of *Harry Reid* are often in quotes that criticize him in an overt way, indirectly voicing Fox's conservative bias. *I will* introduces the uncharacteristic first person into news discourse; it also gives politicians the opportunity to emphasize their campaign promises in more powerful, emotional language than is permitted in news discourse. In the Fox immigration corpus, *if you* is often employed to make readers consider the perspective of American Hispanics, especially regarding the opportunities for racial profiling in Arizona State Bill 1070.

Also, one of the two instances of *President Obama* in the MSNBC corpus is in a quote. The fact that there are only two mentions of President Obama in the MSNBC election corpus indicates that references to the president in this way are uncharacteristic of MSNBC's election discourse. One out of the two instances appearing in a quote indicates that the bi-gram is highly uncharacteristic of the discourse. The effect of this relative absence of the president's name is discussed above.

86

Numbers. According to Bell, figures are often used for rhetorical effect in news articles. Facts sometimes do not need extra support, and readers are rarely capable of putting the numbers into context. Therefore, numbers are often used to exaggerate the effect of facts or give an air of objectivity to story. This often seems to be the case in the Fox corpus, given its frequent mention of poll numbers. In fact, polls are mentioned so often, that a distinct way of talking about them emerged in the Fox corpus, as illustrated by three differentials related to polling, *say they*, *margin of*, and *the latest*. Polling is cited in the MSNBC corpus, but no distinct way of reporting the numbers emerged.

Place. Bell notes that journalists mention place in order to situate the reader. He also says that places can elicit stereotypes associated with that place. *San Diego* appears so often in the Fox Corpus that it is a differential. And instances of *San Diego* normally highlight the border town as a gateway for illegal immigration and crime, thereby building the macroproposition that even affluent, established border cities are susceptible to the consequences of lax border security.

In the MSNBC corpus, place plays a major role in coverage of immigration, eliciting grand notions of America and quaint images of small towns. *Immigration in* is often found in a flexigram where the differential is followed by a locative noun, particularly *America* and *Arizona*. *In America* is often preceded by the word *stay* or references to illegal immigrants, highlighting America as a place of hope for illegal immigrants. *The community* and *town of* often serve to ground the discussion of illegal immigrants in small town America, making immigrants less abstract and more like neighbors.

Time. Bell explains that journalists must constantly give indications of time in article, because they must draw connections between events in a limited amount of space. Therefore, the use of time phrases is to relate events to each other. In the Fox corpus, on Tuesday generally refers to the passage of Arizona State Bill 1070. The fact that the Arizona law was present in the differentials in the form of a time marker suggests that Fox was concerned with explaining that this bill was the catalyst of the ensuing national debate in the summer of 2010.

Conclusion

The hard news articles from Fox and MSNBC are biased in two ways. To cover the 2010 mid-term elections and the 2010 immigration debate, both networks chose articles that highlighted topics that coincide with their conservative and liberal stances, respectively. Fox emphasized the unpopularity of the 111th Congress and border security, whereas MSNBC focused on the re-election of Democrats and the injustice of immigration policy. Furthermore, the networks systematically used particular language that indicates their biases. For example, Fox makes a constant issue of *President Obama*, and MSNBC makes an issue of *the GOP's* policies and maneuvers. *Therefore, the results suggest that my hypothesis is true, because Fox and MSNBC emphasized conservative and liberal issues respectively, while using language that systematically criticizes the opposite side and defends their own positions.*

References

Aday, S. (2010). Chasing the bad news: An analysis of 2005 Iraq and Afghanistan war coverage on NBC and Fox News Channel. *Journal of Communication*, *60*(1), 144-164.

Aday, S., Livingston, S., & Hebert, M. (2005). Embedding the truth: A cross-cultural analysis of objectivity and television coverage of the Iraq War. *Press/Politics, 10*(1), 3-21.

Ansolabehere, S., Snowberg, E. C., & Snyder, J. M. (2005). Unrepresentative information: The case of newspaper reporting on campaign finance. *Public Opinion Quarterly, 69*(2), 213-231.

Baker, P. (2010). Representations of Islam in British broadsheet and tabloid newspapers 1999-2005. *Journal of Language and Politics, 9*(2), 310-338.

Bell, A. (1991). *The language of news media*. Oxford: Basil Blackwell Ltd.

Britto, S., & Dabney, D. A. (2010). "Fair and Balanced?" Justice Issues on Political Talk Shows. *American Journal of Criminal Justice*, *35*(4), 198-218.

Bruno, R. (2009). Evidence of bias in the *Chicago Tribune* coverage of organized labor: A quantitative study from 1991 to 2001. *Labor Studies Journal*, *34*(3), 385-407.

Chen, L. (2005). Transitivity in media texts: Negative verbal process sub-functions and narrator bias. *International Review of Applied Linguistics in Language Teaching, 43*(1), 33-51.

Clark, C. (2007). A War of Words: A Linguistic Analysis of BBC Embedded Reports During the Iraq Conflict. In N. Fairclough, G. Cortese, & P. Ardizzone (Eds.), *Discourse and Contemporary Social Change* (119-40). Bern: Peter Lang.

Conway, M., Grabe, M. E., & Grieves, K. (2007). Villains, victims and the virtuous in
Bill O'Reilly's "No-Spin Zone." *Journalism Studies*, *8*(2), 197-223.

Fico, F., Zeldes, G. A., Carpenter, S., & Diddi, S. (2008). Broadcast and cable networks
news coverage of the 2004 Presidential election: An assessment of partisan and
structural imbalance. *Mass Communication and Society, 11*, 319-339.

Gries, S. T. (2009). What is corpus linguistics? *Language and Linguistics Compass, 3*, 1-
17.

Groeling, T. (2008). Who's the fairest of the all? An empirical test for partisan bias on
ABC, CBS, NBC, and Fox News. *Presidential Studies Quarterly*, *38*(4), 631-657.

Hayes, D. (2008). Party reputations, journalistic expectations: How issue ownership
influences elections news. *Political Communication, 25*, 377-400.

Holtzman, N. S., Schott, J. P., Jones, M. N., Balota, D. A., & Yarkoni, T. in press.
Exploring media bias with semantic analysis tools: Validation of the contrast
analysis of semantic similarity (CASS). *Behavior Research Methods*.

Johnstone, B. (2008). *Discourse analysis* (2nd ed.). Malden, MA: Blackwell Publishing.

Kent, A., & McCarthy, P. M. (in press). The Gramulator: A tool for the identification of
indicative linguistic features. In P. M. McCarthy & C. Boonthum (Eds.), *Applied
natural language processing and content analysis: Identification, investigation,
and resolution*. Hershey, PA: IGI Global.

McCarthy, P. M., & Boonthum, C. (Eds.), (in press). *Applied natural language
processing and content analysis: Identification, investigation, and resolution*.
Hershey, PA: IGI Global.

McCarthy, P.M., Watanabe S., & Lamkin, T. (in press). The Gramulator: A tool for the

identification of indicative linguistic features. In P. M. McCarthy & C. Boonthum

(Eds.), *Applied natural language processing and content analysis: Identification,*

investigation, and resolution. Hershey, PA: IGI Global.

Niven, D. (2003). Objective evidence on media bias: Newspaper coverage of

congressional party switchers. *Journalism & Mass Communication Quarterly,*

80(2), 311-326.

Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007). *Linguistic Inquiry and Word*

Count (LIWC2007): A text analysis program. Austin, TX: LIWC.net.

Schiffer, A. J. (2006). Assessing partisan bias in political news: The case(s) of local

Senate election coverage. *Political Communication, 23*, 23-39.

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

Inc., Publishers.

Weatherly, J. N., Petros, T. V., Christopherson, K. M., & Haugen, E. N. 2007.

Perceptions of political bias in the headlines of two major news organizations.

Harvard International Journal of Press/Politics, *12*(2), 91-104.