

# Visual Analytics of Media Frames in Online News and Blogs

**Nicholas Diakopoulos**  
Columbia University &  
University of Bergen

**Amy X. Zhang**  
NewsCred

**Andrew Salway**  
Uni Research, Bergen

## ABSTRACT

Media frames define different perspectives or ways of communicating about issues and can be manifested through patterns of language use such as key words and their composition. Analytically there is interest in trying to *identify* new frames around issues, and to *compare* how types of frames vary across different news outlets, or over time. In this paper we consider these analytic needs in the context of two use-cases relating to news producers and news consumers, and describe the initial design of a visual analytics tool, the LingoScope, in terms of how it supports these use-cases.

## MEDIA FRAMING AND ITS VISUAL ANALYSIS

Framing refers to the ways in which an issue is presented in the media, including the various perspectives and conceptions that people communicate with respect to that issue. The way language used around an issue can subtly or not-so-subtly influence the way we think about it. Are you “pro choice”, or “pro life”? With the flip of a single word the abortion issue can be framed with values related to freedom of choice, or with respect to the value of human life – both moral propositions. These kinds of framing effects have implications for public perception and ultimately public opinion on issues of importance to society [2].

If framing is the process, and framing effects are the product, then frames are the objects – the discrete instantiations – of communication, operationalized through the choice and usage of words in conjunction with an issue. In his influential 1993 paper, Entman posits that there are different types of frames that correspond to: the definition of a problem, its causal interpretation, the moral evaluation of the issue, and how the issue or problem might be solved [4]. The issue of illegal immigration, for instance, has frames corresponding to politics, economics, and moral imperatives to name just a few. Each issue would of course have specific linguistic cues associated with each of these frame types.

One analytic task relating to frames calls for the use of visualization to support frame *reflection*, the goal being to identify and surface patterns of language use [1]. The linguistic manifestation of frames including particular keywords, stock phrases, or patterns of word use, suggests

an opportunity for visual analytics to identify and help surface these latent linguistic cues. Baumer’s work on this topic [1] seeks to encourage users to consider various linguistic patterns used in the discourse around the issue of cap and trade by visualizing relationships between words. In a study, they found that people who were less concerned with global warming benefitted from the visualization in terms of identifying issue-relevant frames.

The identification of frames of interest is one analytic task associated with frame analysis, but an even more prevalent task found in the literature is in differentiating categories of frames and how and with what frequency various communities or media outlets use different frames. Though there is much scholarly research in this domain, one example is that of Grundmann and Krishnamurthy, which looks at comparing media frames on the climate change issue across the US, UK, Germany, and France by assessing the frequency of co-occurrences of certain key terms. They find, for instance, that France and Germany use a predominantly moral frame on the issue, focusing on political perspectives, whereas the US discourse is dominated by a scientific frame [5]. This is a typical analytic task associated with frames in that it seeks to subdivide some corpus (in this case by national outlet) into analytic units and compare the use of frames among those units.

The use of visualization to provide better identification and reflection on issue frames in [1] is promising, but at the same time doesn’t address this other analytic need: *comparison* between sources, or *comparison over time* for a single or multiple sources. Indeed it’s arguable that visual comparison of texts is poorly supported by existing text visualization techniques such as Wordles [6], Word Trees [7], or even Double Trees [3] and that this needs to be a focus of development in text analytics. In this work we are specifically interested in addressing this issue by developing tools and visualizations to support the visual comparison of frames between different corpora and over time.

## SPECIFIC USE CASES

Our particular use cases combine the notions of frame *reflection and identification* with that of frame *comparison* and consider the related needs of (1) news *producers* and

(2) news *consumers*. The data in these scenarios is multidimensional in the sense of incorporating both the texts themselves (i.e. keyterms and their relationships) as well as a temporal component.

The first scenario we consider is that of a journalistic news outlet trying to assess its use and adherence to a style guide. A motivating example from earlier this year is how the Associated Press decided that its style guide should no longer allow for the use of the phrase “illegal immigrant” by reporters<sup>1,2</sup>. Related idioms like “illegal alien,” “an illegal,” or “undocumented immigrant” were also struck from the guide. Around the same time, the New York Times reported that it is also reconsidering the use of “illegal immigrant”<sup>3</sup>. Other media sources, such as Time, have written about the use of language around gay rights<sup>4</sup>, and of course there are many other examples of the media’s interest and self awareness of language use and codification in style guides. In all of these cases these outlets are clearly reflecting on how sensitive or political issues like immigration are framed in their written output. What can visual analytic tools offer them? What other words besides “undocumented” or “unauthorized” also appear together with the word “immigrant”, and how do such frames around the issue *compare* over time as the style guide adapts? A tool might help *identify* new or related frames around immigration to further enable the news outlet to consider its use of those frames. Or it might help a news outlet see to what extent its usage of that frame is adherent to the style guide. This could allow for more awareness of how the editorial voice of the outlet is developing over time.

In the news *consumption* use-case, the analytic tasks become perhaps less about identification of new frames, than about being able to compare frames across different corpora defined by outlet. For instance, as a consumer I may be interested in seeing a comparison of how the Wall Street Journal stacks up against the Washington Post, or the Arizona Republic in the use of the phrase “illegal immigrant”, both in volume and rate of use, as well as how that rate has evolved over time. In essence, visual analytics in this scenario is about empowering the end-user to monitor different media outlets and compare how an issue of interest to them is framed. This could impact their choice of media outlet, or just make them more aware of how a particular outlet’s view or bias on an issue might have

changed over time. If an end user is particularly interested in the editorial output and comparison of two outlets, the ability to hone in on sentiment valence, or the types of adjectives could further illuminate the comparison.

It should be clear that there is some overlap between these use cases in terms of their emphasis on comparison, and seeing how an outlet (or outlets) have evolved over time. Based on the extant literature on corpus linguistics and frame analysis we see that scholars have similar types of tasks in terms of comparing across and among corpora, as well as identifying new patterns of language use. For our own work, we are additionally interested in comparisons among not just corpora defined by outlet, but also along other dimensions of those outlets such as geography or magnitude of circulation, and among different types of media such as articles vs. comments, articles vs. blogs, and different groupings of blogs based on their network characteristics (such as using community detection algorithms to define subsets of blogs). Beyond the news outlet and news consumer use cases we have defined, we believe there is a rich opportunity to develop tools that also enable scholars to visualize and analyze media frames in many different ways.

## THE LINGOSCOPE

In this section we describe our initial work on designing a system, called the LingoScope, to address the use cases we have identified relating to media frame analysis. The system is shown in Figure 1. The user first selects a news outlet (or two outlets if seeking to compare them), and then enters a search term relating to an issue of concern (e.g., “immigration”). Then the user can define how to operationalize the frame, selecting among contexts defined by words preceding or succeeding the search term, or more loosely defined by words appearing in the same sentence or paragraph. Upon searching, the user will see a list of words, initially ranked by frequency, that correspond to words in the chosen context (in Figure 1 words that come before “immigration”). The user will see a visual comparison of the rate at which each of the corpora uses that word in the given context with the search term. Moreover, the user will be able to see, over time aggregated by month, how the rate of usage has changed and compare between the two corpora. Clicking on the timeline or bar chart allows the user to drill down and see details in the lower right panel, so that they can read through the use of that word in its original context. As the user proceeds, they may be interested in finding specific types of words, and we support this activity by providing filters based on part of speech (e.g. adjectives) or whether it is a sentiment laden word [8].

The choice of different frame operationalizations is important in our design so that news outlet users might *identify* different types of frames around the chosen issue. For instance, defining the word context as “before” or “after” essentially returns a list of bigrams that occur with

<sup>1</sup> <http://www.poynter.org/latest-news/mediawire/209045/ap-changes-style-on-illegal-immigrant/>

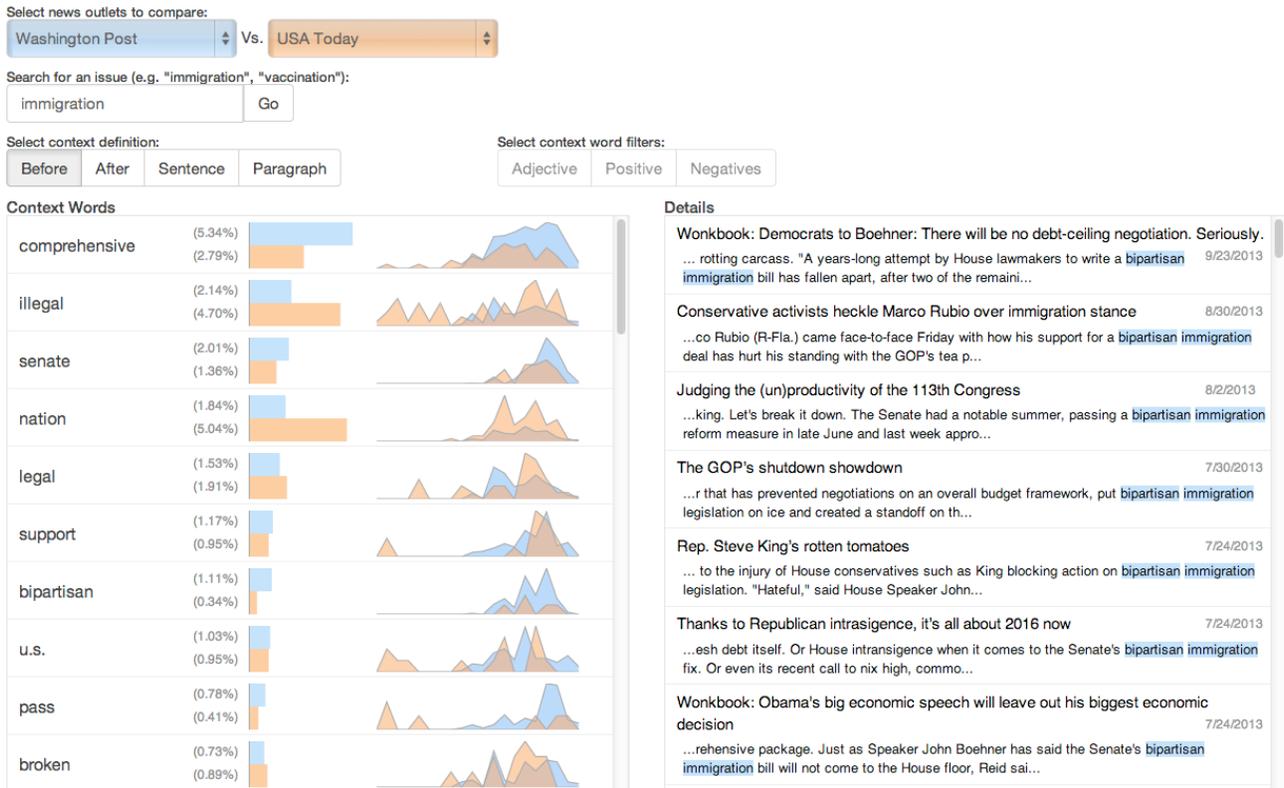
<sup>2</sup> <http://www.poynter.org/latest-news/top-stories/210053/what-data-algorithms-teach-us-about-the-language-news-organizations-use/>

<sup>3</sup> [http://publiceditor.blogs.nytimes.com/2013/04/02/the-times-too-is-reconsidering-the-term-illegal-immigrant/?\\_r=0](http://publiceditor.blogs.nytimes.com/2013/04/02/the-times-too-is-reconsidering-the-term-illegal-immigrant/?_r=0)

<sup>4</sup> <http://newsfeed.time.com/2013/03/27/seven-hang-ups-in-the-language-of-gay-rights/>

# LingoScope

Media frames are different perspectives on an issue which can manifest as patterns of language use and word choice. LingoScope helps you analyze and visualize these media frames by seeing and comparing how words are used around a given issue and across different news outlets.



**Figure 1. LingoScope prototype showing capabilities for selecting news outlets, searching for an anchor issue, and selecting different operationalizations of context or filters for context words.**

the issue word. But by defining the word context more broadly, such as based on co-occurrence in the same sentence or paragraph, we may uncover other words that often appear close together but not adjacently. Simple visual representations for comparing rate of word usage over time and between corpora, with straightforward color mappings, were chosen in order to provide a quick overview. The list nature of the context words also functions to create a vertical multiples display which enables quick comparison of patterns between words. The ability to drill into a word supports an analyst seeing and reading the word's exact usage in context, an important feature to the user in making sense of any nuances.

We are actively developing LingoScope using data provided by the NewsCred platform. NewsCred aggregates news content from thousands of sources and provides that content via an API. Drawing on the list of top 25 newspapers in the US<sup>5</sup> which we augmented with a few solely online sources, we selected an initial 14 news outlets

to focus on based on the criteria that they should have at least 10 months of full-text data to be interesting to compare over time. Across these 14 news outlets, NewsCred has almost 1.3 million full text articles covering a period stretching from 2010 to today, with new content continuously aggregating in real time. There is some variety in both volume of articles published and length of aggregation history by NewsCred across the different sources, and thus we normalize for this variance.

In addition to the NewsCred data for mainstream outlets, we are also integrating blog corpus data in order to compare media frames in use by different communities within the network defined by those blogs. We think this can also shed light on how different bloggers talk about sensitive issues, like immigration or climate change, as well as provide an additional comparison point between the less formal communication in blogs and the more formal communication present in the mainstream media. After completing the implementation of the system we hope to assess the two use cases we have identified by gathering feedback from newsrooms, and by making it available to

<sup>5</sup> <http://accessabc.wordpress.com/2012/10/30/the-top-u-s-newspapers-for-september-2012/>

end users interested in comparing and contrasting various news outlets online.

#### **ACKNOWLEDGMENTS**

Thanks to Mike Schäfer at the University of Zurich and Dag Elgesem at the University of Bergen for helpful discussions on media framing. This research was supported by a grant from the Norwegian Research Council's VERDIKT program and by NewsCred through their data and other computing resources.

#### **REFERENCES**

1. Eric Baumer, Francesca Polletta, et al. Developing Computational Supports for Frame Reflection. In Proc. iConference 2013.
2. Dennis Chong and James N. Druckman. Framing Theory. *Annual Review Political Science* 2007. 10:103–26
3. C. Culy. Double Tree: An advanced KWIC visualization for expert users. *Information Visualization (IV)* 2010.
4. Robert M. Entman. Framing: Towards Clarification of a Fractured Paradigm. *Journal of Communication* 43 (4), 1993, 51-58.
5. Reiner Grundmann and Ramesh Krishnamurthy. The Discourse of Climate Change: A Corpus-based Approach. *Critical Approaches to Discourse Analysis across Disciplines* 4 (2) 113-133. 2010
6. F. Viégas, M. Wattenberg, and J. Feinberg, "Participatory Visualization with Wordle," *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, no. 6, 2009.
7. M. Wattenberg and F. Viégas, "The Word Tree, an Interactive Visual Concordance," *IEEE Transactions on Visualization and Computer Graphics*, vol. 14, no. 6, 2008.
8. Theresa Wilson, Janyce Wiebe and Paul Hoffmann (2005). Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. *Proceedings of HLT/EMNLP*. 2005.