The word genre can be used and interpreted differently, and while some analysts we’ve discussed have been interested in redefining, solidifying, or otherwise reintroducing scholars to what it means, I’m curious to see if the principles that they’ve incorporated into their software can be used on a smaller scale to discover whether political leanings could function similarly to genre in the realm of journalism. The question I would like to attempt to answer with an MFW approach is:

Can the political leaning of a political news article’s host site be determined by the frequency of certain (common or uncommon) words or phrases?

The first step is to identify websites whose political stories consistently and firmly align them with a particular end of our dual political spectrum. The internet has already done most of my legwork for me here, as most websites that report political news make little effort to hide their propagandistic motivations. I plan to have a pool of 10 websites to draw articles from, 5 from each political side.

The next step in selecting the corpus is to choose which topics/issues I’ll be drawing articles from, and then which specific events or moments in time. Here are a few possibilities:

1. Abortion (also possibly ACA): The week following the Hobby Lobby ruling
2. Gun control – The week following Wayne Lapierre’s post-Sandy Hook speech
3. Same-sex marriage – The week following last month’s Supreme Court decision

In collecting the data other information will be collected that could prove informative as well as disruptive to the process, such as disproportionate amounts of writing on a topic from certain sites, conflated issues (as noted in example 1), and “genres” of article (editorial, report, etc.). By the end I hope to have a reliably representative corpus that will be used for the MFW search itself, but also other supplemental information that could demonstrate trends or habits among particular sites or political leanings.

Past that, it might be fruitful to note any consistencies that act as identifiers and see how they apply to less overtly biased websites.