Journal of Elections, Public Opinion and Parties
Vol. 19, No. 4, 359–376, November 2009

Campaign News and Vote Intentions

STUART SOROKA, MARC ANDRÉ BODET, LORI YOUNG & BLAKE ANDREW
McGill University, Montreal, Canada

ABSTRACT *This paper explores the relationship between campaign news and vote intentions, drawing on manual content analyses from the 2004 and 2006 Canadian federal elections. The content analysis is designed to capture, among other items, the 'tone' of coverage for parties and leaders. Combining time series of 'tone' and commercial polling results, econometric methods are then used to build a model of vote intentions. Media content appears to explain a good deal of the over-time variance in vote intentions. Results are discussed as they pertain to two versions of the media-opinion relationship: (1) media content captures and arranges in a readily quantifiable form the evolving mood of the campaign, or (2) media do not simply reflect, but affect vote intentions.*

That mass media are central to modern representative democracy requires little discussion. Mass media play an important role in producing an informed (at least moderately informed) public. They are critical to the dissemination of information about, for instance, national conditions, government activities, and public policy issues. Media content is accordingly strongly connected with public opinion and politics. Indeed, there are burgeoning literatures detailing the relationship between media content and, for instance, public attentiveness to issues, policymakers' framing of policy matters, and public attitudes about public policy (on attentiveness see, e.g., McCombs & Shaw, 1972; Behr & Iyengar, 1985; Soroka, 2002; on framing see, e.g., Iyengar, 1996; Baumgartner & Jones, 1993; on policy attitudes see, e.g., Jamieson & Cappella, 1998; Fan & Norem, 1992; Soroka, 2003).

Whether media leads or follows public opinion is of course in many cases not clear; it is likely that at any given time it does a bit of both. Media content can be regarded in two often empirically inseparable ways: (1) it can reflect the issues, themes and actors that are currently prominent in public debate, and (2) it can be a potential driver of public opinion and policy. In the former case, mass media act simply as a mirror. Media content in this view is a useful summary indication of the more general public sphere. In the latter case, mass media are not mirroring but affecting. Media content differs from what citizens or politicians currently think, and has the potential to affect these actors' attitudes.

Correspondence Address: Stuart Soroka, Department of Political Science, McGill University, 855 Sherbrooke St West, Montreal QC, H3A 2T7, Canada. Email: stuart.soroka@mcgill.ca

ISSN 1745-7289 Print/1745-7297 Online/09/040359-18 © 2009 Elections, Public Opinion & Parties
DOI: 10.1080/17457280903275030

Both of these connections between media content and public opinion are particularly strong during election campaigns. These are periods of heightened attentiveness to political issues for journalists and for many citizens. The modern election campaign is hugely dependent on media – on advertising partly, but on news content as well. Indeed, election campaigns are almost by definition media campaigns: for most citizens, TV, newspapers, and internet are the principal if not the only sources of information about parties, candidates, policies, and, of course, the horserace.[1] Journalists are highly attentive to the campaign as well as to the state of public opinion, particularly vote intentions. And many citizens are probably, compared to other periods at least, more attentive to media content. Given this heightened mutual attentiveness, we might expect the link between media and the public during an election campaign to be especially strong.

This paper explores the relationship between campaign-period media content and public opinion, focusing on two recent Canadian federal election campaigns. Specifically, we examine the relationship between the tone of media coverage and vote intentions for the major parties during the 2004 and 2006 Canadian federal election campaigns. Results suggest that this relationship is rather strong – that, due to a number of possible mechanisms, media content and pubic opinion are positively related. We also explore a more specific (and perhaps more ambitious) possibility, however: with a good measure of tone in media content, is it possible to actually predict vote intentions?

We believe we can. This suggests, at a minimum, the potential importance of media content (and particularly the valence of that content) in understanding election campaigns. It may also demonstrate a rather striking effect of media content on vote intentions during election campaigns. And note that while demonstrating a link between media content and public opinion during campaigns is by no means an original contribution,[2] the fact that media content analytic data can actually be used as a (powerful) leading indicator of trends in opinion is original.

Data and results are presented in two sections below. The first describes the data for 2004 and 2006. The second explores the strength of these data in predicting vote intentions during these two campaigns. Overall, results speak to a predictive relationship between media tone and opinion during election campaigns. Vote intentions appear to have been strongly influenced by parties' shares of "good press "and "bad press "during the formal campaign period.

Data

Our analyses rely on two bodies of data. The first is relatively simple: a database of all commercial polls reported during the 2004 and 2006 Canadian federal election campaigns. In 2004, we make use of five polling firms and 29 polls conducted during the campaign. In 2006, there were eight pollsters regularly in the field, with 40 reported polls and two rolling-cross sections of a few hundred respondents per day conducted by SES and the Strategic Counsel.[3]

The other database is a body of content analytic data, manually coded by teams of trained undergraduate and graduate student coders during the 2004 and 2006 federal campaigns.[4] The 2004 and 2006 studies were conducted separately, but are directly comparable methodologically speaking. Each tracks all campaign content – news, editorials, and opinion pieces – published in seven major daily newspapers across Canada (five English-language, two French-language): *Vancouver Sun*, *Calgary Herald*, *Toronto Star*, *Le Devoir*, *La Presse*, *National Post* and *The Globe and Mail*. Daily coding for the 2004 election began one week before the writ dropped, lasting for six weeks until the day of the election, 28 June 2004. The 2006 campaign – among the longest in Canadian history – accounts for eight weeks of coding material before the election on 23 January 2006. Practically speaking, coders read through about 7–10 full-text newspaper stories each campaign day. In total, 6,694 articles are included in this dataset: 4,280 news stories and 2,414 editorial and opinion pieces. In both 2004 and 2006, coders surveyed the main news sections of the major Canadian dailies for the duration of the campaign. There were about a dozen coders for each campaign, introduced to the study through formal training sessions that included a series of practice coding exercises and a guide for our online data entry system. Coding happened daily, as the campaign progressed. Coders were responsible for a different newspaper each week, in order to test for (and avoid) any coder effects or bias. Stories were also randomly selected for double coding throughout the campaign to check inter-coder reliability – the consistency with which different coders come up with identical codes.[5]

Coding captured a large body of relatively objective data, including mentions of issues, parties, and party leaders. Most critical for the forthcoming analyses, coding also included one set of subjective codes for tone – positive, negative or neutral – for major parties and their leaders.[6] The major parties and leaders received a –1, 0, or +1 tone for each article they appeared in; a –1 tone score for the Liberal Party represented a bad news story, whereas +1 represented a good news story for them. Hence, "net tone", measured by the percentage of positive stories minus the percentage of negative stories, indicates the relative weight of good news over bad news for parties and leaders during the campaign period in the Canadian press.[7]

The specific instructions for coding the tone of media content were as follows: the default for all mentions is neutral; a leader or party mention has to be clearly good press or bad press to be coded as such (see also the coding strategy developed by Robinson and Sheehan, 1983; Brady & Johnston, 1987). Put another way, unless the story was obviously and intentionally positive or negative, a mention of a leader or party is neutral. This is what you might call latent rather than manifest measurement of election news content – it captures tone evident in the reporting of or commentary on a given event, rather than negativity or positivity of the event itself (see Andrew, 2007; Riffe et al., 2005). For that reason, careful attention was paid to training and to reliability analyses for this indicator.

Note that the tone code does not simply reflect reports of leaders and parties criticizing policy platforms and records of competitors. Normative statements made by leaders, candidates, and partisans are excluded from this version of campaign tone.

Instead, our measure of an article's tone reflects critical and positive commentary of the main leaders and parties from sources other than the main protagonists of the campaign itself. For instance, reporting a Harper speech in which the Conservative leader objected to or attacked something about Liberal Party Leader Paul Martin was considered neutral – just reporting the news. Reporting that speech and using it as evidence for Martin's failings was considered negative, however. Our coders also noted tone when, for example, an economist issues an endorsement for a party's tax policy proposal, but not when another party leader attacked (or endorsed) it. This aspect of our tone measure closely reflects the strategy developed by Robinson and Sheehan (1983) to assess media coverage of the 1980 US Presidential campaign.

Our measure of media tone differs from theirs in one key respect, however: tone includes assessments of leaders' and parties' performance in the campaign and in public opinion polls. Specifically, reports of a party "surging ahead" or "gaining ground" in the polls were noted as a positive. Conversely, stories noting that parties were "falling behind" and "uninspiring" or "gaffe-prone" campaign performances were duly recorded as negative press. Stories that simply reported a party's or leader's standing in an opinion poll were coded as neutral. In short, media tone for coverage of opinion polls was contingent on a clear expression of journalistic interpretation, not on the numbers themselves.

Note that there are both advantages and disadvantages to including articles dealing with opinion polls. The disadvantage is clear: doing so means that our media data incorporate to some degree polling trends. The consequence is that the predictive power of our media measure is likely strengthened by information that is essentially a lagged value of the dependent variable (future polling results). That said, the proportion of articles dealing with polling results directly is relatively small: roughly 9% of articles in 2004 and 2006 combined,[8] only some of which – given our conservative measure of tone – affect our measure of tone. Excluding polling articles thus makes very little difference to the overall results. There are, we believe, also substantive advantages to including polling articles. In particular, using all articles means that we are using the "media signal" with which we might predict vote shares during a campaign.[9]

The overall result of our approach to coding was that mentions of parties and leaders in news stories were predominantly neutral, and mentions in editorial and opinion pieces were mainly negative or positive. Table 1 provides some basic descriptive results. Our data include 3,766 articles from the 2006 election and 2,957 from 2004. Combined, roughly 64% of our sample is news stories; 28% opinion pieces; and 8% editorials. The table also shows the percentage of each article type mentioning the two major parties, and the percentage of those mentions that were coded as positive or negative. In 2006, for instance, 75% of news stories mentioned either the Conservative Party or their leader Stephen Harper. Of those mentions, 5% were negative and 6% were positive. The remaining 89% of mentions were neutral, then. Looking to opinion and editorial pieces, a much greater proportion of mentions were non-neutral. The table also shows the divergence in Liberal and Conservative coverage: the Liberals got consistently poor coverage in both elections, while the

Table 1. Descriptives: sample size and party tone by article type

	Election year	
	2006	2004
Total articles	3,766	2,957
News	65% (2,441)	63% (1,866)
CPC/Harper mentions	75%	74%
negative mentions	5%	13%
positive mentions	6%	4%
LPC/Martin mentions	85%	86%
negative mentions	20%	11%
positive mentions	3%	2%
Opinion	27% (1,029)	29% (854)
CPC/Harper mentions	80%	74%
negative mentions	15%	31%
positive mentions	15%	14%
LPC/Martin mentions	87%	85%
negative mentions	49%	37%
positive mentions	4%	6%
Editorial	8% (296)	8% (237)
CPC/Harper mentions	74%	67%
negative mentions	19%	38%
positive mentions	22%	23%
LPC/Martin mentions	85%	81%
negative mentions	51%	51%
positive mentions	11%	6%

Conservatives receive somewhat less negative coverage in 2004, and marginally positive overall coverage in 2006 (for more detailed descriptives for all variables in the media datasets, see Andrew et al., 2006; Goodyear-Grant et al., 2004).

Our data thus reflect the traditional wall of separation between hard news and the op-ed section of newspapers. The editorial and opinion sections convey much of the clearly positive and negative tone about the contestants of these campaigns. To be sure, the coders have disregarded some of the subtle tone conveyed by articles, but then so do everyday citizens. In short, our measure of tone has been designed with the intention to capture relatively clear (and reliable) endorsement cues in campaign news.

Analysis

This combination of polling and media data allow for a wide range of analyses on electoral campaigns and political communications. We focus here on one relatively

narrow question, however: how well can we predict vote intentions using media content?

We explore this question using a simple Ordinary Least Squares (OLS) regression[10] model of the current vote share at time *t* for each of the major two parties, and two matrices of lagged independent variables. The first is a series of dummy variables for Canadian polling firms in our sample, including SES, the Strategic Counsel, Ekos, Léger Marketing, Decima, Ipsos Reid, Environics, Zogby Poll, Pollara, and Compas. These variables are equal to 1 if a given firm has a poll in the field on a given day and 0 otherwise. They are intended to capture "house "or "pollster" effects – the tendency for different firms to capture slightly but systematically different distributions of vote intention, due to methodological decisions relating to, for instance, the partition of undecided voters and "don't knows" (see, e.g., Erikson & Wlezien, 1999; Jackman, 2005; Pickup & Johnston, 2007; McDermott & Frankovic, 2003).

Pollster effects can be viewed as one of several different types of error in polling data, alongside what classical statistics labels "random error", a function of sample size, clerical errors, and so forth. The error resulting from pollster effects is dealt with by using the dummy variables outlined here.[11] In any case, we do not interpret pollster effects dummies below, but include them only as controls.

The second matrix of independent variables includes four net tone measures, for (1) the Liberal Party, (2) the Conservative Party, (3) Paul Martin, and (4) Stephen Harper. Recall that our measure of media tone is simply the number of positive mentions for a leader or party minus the number of negative mentions on a given day.[12]

Put more formally, our prediction model is as follows,

$$Vote_{p,t} = \alpha + \Sigma(\beta_f * Pollster_{f,t}) + \Sigma(\omega_\eta * Tone_{\eta,t-k}) \quad [+\omega_\eta * Vote_{p,t-k}] + \varepsilon_t,$$

where polled vote intentions for party (*p*) at some time (*t*) is a function of a set of dummy variables capturing pollster effects for each firm (*f*) at that time and a set of net tone measures for each major party and leader (η) lagged by *k* time periods. We also include in square brackets above the dependent variable, $Vote_p$, also lagged *k* time periods. On the one hand, while a model's predictive capacity can be limited by not taking into account a strong autoregressive (AR) process (the tendency for a value at *t* to be strongly related to a value at *t-k*), excluding the AR process provides a good opportunity to evaluate the predictive capacity of media content variables without the strong input of the lagged dependent variable.

On the other hand, any effort to produce estimates of some variable some time in the future would certainly include, if available, recent or current values of that variable. Since we do have current polling results at any given (*t–k*) point in the campaign, there is nothing preventing us from including the party's vote share in the prediction. Indeed, including both media and lagged vote share together provides a strong test of the degree to which media content improves the prediction, above and beyond what we would know using just the current vote share; and a model using just lagged vote share provides a good baseline against which to

measure the contribution of media variables. We accordingly estimate each model three ways below: (1) vote share at *t–k*, (2) media content at *t–k*, and (3) both vote share and media content at *t–k*.

The choice of lag – that is, the value of *k* – is driven by a combination of pragmatic and statistical considerations. Pragmatically speaking, the further back the lags are (the greater the value of *k*), the further forward we are able to predict. A model using media at *t–1* would allow us to predict only one day forward; a model using media at *t–6* would allow us to predict six days ahead. At the same time, with a limited election period high order lags are costly in terms of degrees of freedom – each additional lag means one less data point. We would thus ideally select lags over some kind of middle period, not too proximate (so we can predict), but also not too distant (to preserve sample size).

That said, the data also speak for themselves. If there is a relationship between media content and vote intentions, the strength and timing of that relationship should be relatively clear in preliminary models and in simple cross-correlations, calculated between media content and vote intentions at various lags and leads. Our preliminary tests (not shown here) suggested that correlations were typically strongest at lags *t–4* through *t–6*.[13] This finding was roughly consistent for both Liberal and Conservative vote intentions, in both the 2004 and 2006 elections, and using any combination of the net tone measures. The lags also have the happy coincidence of allowing us to predict four days in advance, and not losing too many degrees of freedom. We accordingly include lags *t–4* through *t–6* of media content in all our estimations of equation 1. For vote share, where included, we use just *t–4* – the most proximate polling data we would have if we were mid-campaign, using today's media content to project vote share four days ahead.[14]

Results from these estimations are included in Tables 2 and 3.[15] With three lags of four different variables, each of which is correlated to the others, the model is vastly over-specified. Coefficients can be very difficult to interpret, given that there will be several coefficients capturing what is, substantively speaking, a single effect. Standard errors will also be inflated by multicollinearity. This is of course common for prediction models that seek to explain as much variance as possible, and place little emphasis on interpreting individual coefficients. It presents no particular problem for our work here, but it does mean that we should not place too much weight on the individual coefficients.

In an effort to make results more readily interpretable, the first row of Tables 2 and 3 shows the summed coefficients (and related standard errors) for all lags of tone relating to both the Conservative Party and their leader Stephen Harper. The same is done for the Liberal Party and Paul Martin in the second row. Each provides an omnibus test of the direction and magnitude of the effects of tone for the two major parties and leaders combined. The rather obvious expectation is that the Conservative's news tone will be positively related to Conservative vote intentions and negatively related to Liberal vote intentions, while Liberal tone will be positively related to Liberal vote intentions and negatively related to Conservative vote intentions.[16]

Table 2. 2006 vote intentions

	Conservatives			Liberals		
	1	2	3	1	2	3
Σ $CPC_{t-(4,5,6)}$		35.434	29.599		–32.586	–35.492
		(10.984)	(7.513)		(11.010)	(6.007)
Σ $LPC_{t-(4,5,6)}$		–0.152	3.094		30.533	16.372
		(14.723)	(10.249)		(14.785)	(8.233)
DV_{t-4}	0.709		0.621	0.698		0.703
	(0.119)		(0.110)	(0.123)		(0.088)
Strategic Counsel	–0.410	–0.638	–0.153	–1.580	–1.236	–2.206
	(0.775)	(1.108)	(0.755)	(0.835)	(1.111)	(0.617)
SES	0.393	2.462	1.568	0.633	–0.797	–0.120
	(0.979)	(1.435)	(0.985)	(1.051)	(1.439)	(0.788)
Decima	1.150	3.003	1.304	–3.381	–4.551	–2.286
	(1.417)	(2.039)	(1.414)	(1.474)	(2.044)	(1.149)
Ekos	2.080	5.533	2.284	–2.006	–5.341	–1.172
	(1.098)	(1.212)	(1.004)	(1.188)	(1.215)	(0.845)
Environics	1.08	7.737	2.776	–4.305	–9.845	–5.534
	(2.307)	(3.217)	(2.328)	(2.426)	(3.225)	(1.838)
Ipsos-Reid	–0.741	–1.610	–1.562	0.424	–0.146	0.281
	(0.92)	(1.408)	(0.954)	(0.981)	(1.411)	(0.770)
Léger	–1.992	–1.076	–1.853	0.154	0.453	–0.125
	(1.108)	(1.450)	(0.992)	(1.183)	(1.454)	(0.795)
Pollara	–0.444	–2.015	–0.300	1.956	3.274	0.829
	(1.746)	(2.453)	(1.689)	(1.846)	(2.459)	(1.374)
Intercept	9.888	28.755	11.718	10.554	40.869	13.944
	(3.750)	(2.171)	(3.367)	(4.559)	(2.176)	(3.601)
Statistics						
Adjusted R^2	0.657	0.706	0.761	0.745	0.576	0.874
N	47	47	47	47	47	47
Accuracy						
MAE	1.548	1.586	1.060	1.654	1.625	0.810
	(1.097)	(1.149)	(0.768)	(1.166)	(1.101)	(0.666)

Cells contain OLS coefficients with standard errors in parentheses.

Let us begin with results for the Conservatives in 2006, the first three models in Table 2. Column 1 shows results using just lagged vote share and pollster effects. The adjusted R-squared is very high, 0.657 – already a good deal of the variance in vote share at *t* is explained using just vote share at *t–4*. But note that while the R-squared provides a summary of the proportion of variance in *y* explained by *x*, it does not provide the piece of information we are most interested in where prediction

Table 3. 2004 vote intentions

	Conservatives			Liberals		
	1	2	3	1	2	3
Σ CPC$_{t-(4,5,6)}$		7.027	38.543		15.405	6.132
		(14.070)	(18.868)		(10.758)	(12.880)
Σ LPC $_{t-(4,5,6)}$		17.410	24.473		0.664	4.645
		(15.325)	(13.988)		(11.718)	(11.927)
DV $_{t-4}$	0.451		0.616	0.497		0.437
	(0.115)		(0.276)	(0.170)		(0.348)
SES	2.135	3.009	2.589	0.058	–0.263	–0.127
	(0.723)	(0.997)	(0.906)	(0.634)	(0.762)	(0.756)
Ekos	2.073	0.397	0.908	–0.512	–0.238	–0.192
	(1.188)	(1.913)	(1.716)	(0.975)	(1.463)	(1.436)
Ipsos-Reid	–1.443	–1.409	–1.823	–1.028	–0.526	–0.505
	(0.832)	(1.235)	(1.113)	(0.666)	(0.944)	(0.926)
Léger	–0.582	0.843	–0.037	0.833	0.218	0.562
	(1.165)	(1.488)	(1.381)	(0.938)	(1.138)	(1.150)
Intercept	16.718	34.769	19.055	16.886	34.344	19.845
	(3.43)	(3.973)	(7.879)	(5.845)	(3.038)	(11.918)
Statistics						
Adjusted R^2	0.548	0.632	0.399	0.208	0.098	0.131
N	34	32	32	34	32	32
Accuracy						
MAE	1.365	1.145	1.027	1.112	0.867	0.843
	(0.970)	(0.851)	(0.666)	(0.763)	(0.662)	(0.600)

Cells contain OLS coefficients with standard errors in parentheses.

is concerned – exactly how close are the predictions to the future values of *y*? To better assess this critical piece of information, we rely here on the Mean Absolute Error (MAE), which captures the average gap between the prediction and the actual vote intentions.[17] For this first model, the MAE is 1.548. (MAE is in the same unit as the dependent variable, so we are talking here about a prediction that on average misses the Conservative vote share by about one and half percentage points.)

Column 2 of Table 2 shows results for the 2006 Conservative vote share, this time using just lagged media variables. First, note that Conservative net tone is positively related to Conservative vote share; Liberal net tone is essentially zero. The R-squared improves somewhat (to 0.71), though the MAE remains almost the same, at 1.586. We should not lose sight of the fact that this is a model that uses just media content, however. And the R-squared is at its highest, and the MAE at its lowest, in the third column, which combines lagged media and vote shares. Importantly, this last model also shows that media coverage of the Conservative Party still has a significant effect, even when lagged vote intentions are included in the model.

Media effects are clearer still in the model for Liberal vote share in 2006. The MAE is considerably lower for the combined model than for the model using just vote share, 0.810 versus 1.654. Using media content in the prediction of Liberal vote share cuts the average error in half. And in both columns 5 and 6, the effects of Conservative and Liberal tone are correctly signed: Liberal vote intentions during the 2006 campaign appear to have been positively influenced by the tone of Liberal coverage, and (powerfully) negatively influenced by the tone of media coverage for their main opponent.

Table 3 presents roughly similar results for 2004, though in this case results are not nearly as robust. The MAE does indeed improve with the inclusion of media variables. Indeed, the effect of Conservative media on that party's vote share is very powerful once lagged vote intentions are included (in column 3). But the omnibus tests for the tone coefficients are in all other cases positive (regardless of party) and imprecise. Again, though, there is evidence that Conservatives' vote intentions were driven mainly by coverage they received in Canadian dailies, and not by coverage of the Liberal Party's campaign.

The key point is that predictions improve with the inclusion of media content. This is strong evidence of the effect of campaign period media, we believe. The degree to which media both reflect current vote shares and affect future trends is particularly evidenced by the strength of models including just media content. Looking across Tables 2 and 3, it is striking that we can explain as much as 75% of the variance in major party vote intentions using just lagged media.

The predictive power of media content is further illustrated in Figures 1 and 2, which show both polls and (lowess-smoothed) predictions for the two major parties over the 2004 and 2006 elections, relying on the models that include media content only. This is the first glimpse of the campaigns themselves, at least in the present paper, so it is worth noting some aspects of the campaign that are driving the dynamics apparent in these figures. In reaction to an ongoing scandal in which money from a "sponsorship" program intended to promote the Federal Government in Quebec appeared to have been funneled to Liberal-friendly agencies, the first weeks of the 2004 campaign show a general shift away from the Liberals and towards the Conservatives. The drop in Conservative fortunes and corresponding increase in Liberal vote shares later in the campaign began around the period of the leaders' debates (14 and 15 June). This point in the campaign was also marked by a high-profile Conservative gaffe, and intense attacks by the Liberals suggesting that the Conservative Party was too socially conservative (on both same-sex marriage and abortion), and was concealing a "hidden agenda" on healthcare from the electorate. Preferences in 2006 show a much more prolonged shift towards the Conservative Party, though here too there are blips along the way. Vote intentions are relatively flat for the pre-Christmas part of the campaign. This period was characterized by a good deal of policy discussion – mainly about a series of Conservative policy proposals – but not the kind of intense partisan attacks (particularly surrounding a Royal Canadian Mounted Police (RCMP) investigation of a senior Liberal Minister's office, announced on 28 December) and horserace

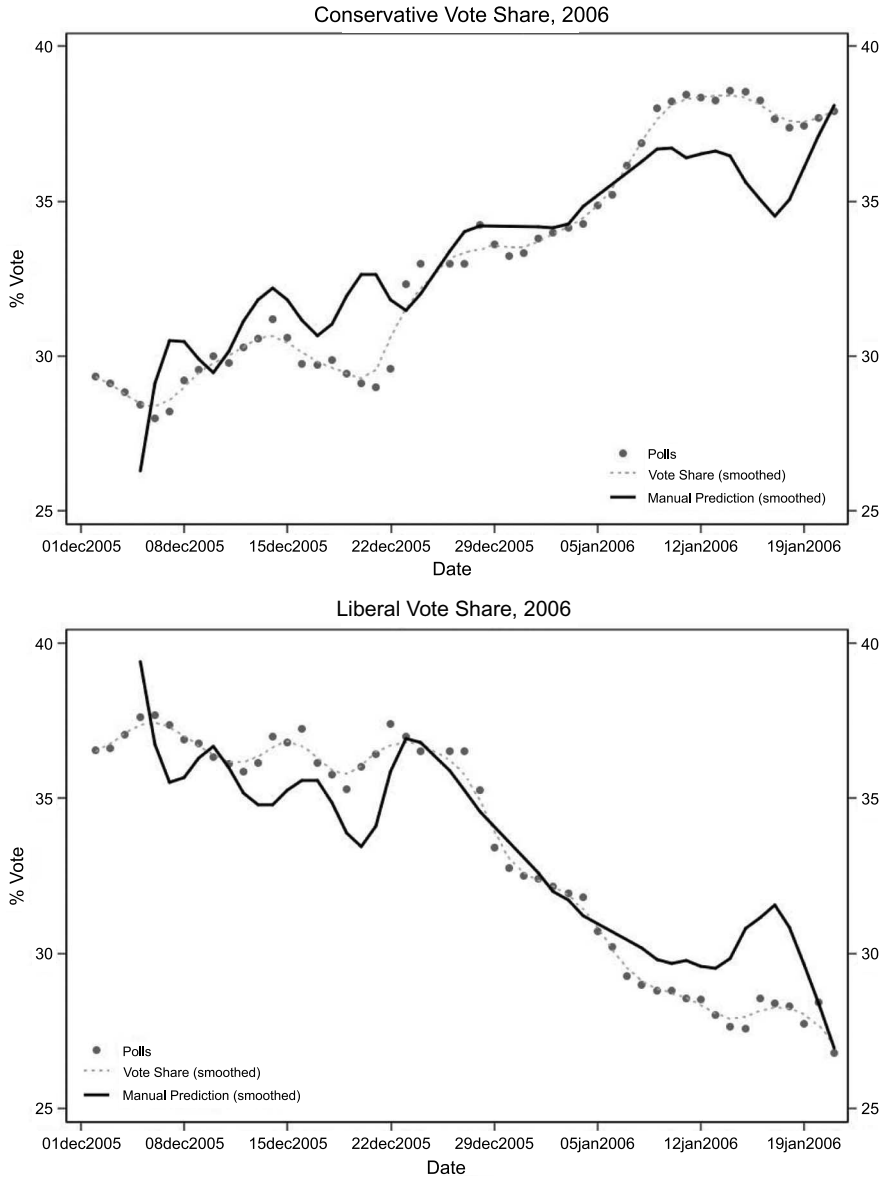

Figure 1. 2006 predictions, media content only.

coverage that characterized the second part of the campaign. Debates were on 15 and 16 December, and again on 9 and 10 January.

In both elections, predictions track vote intentions rather well. There are some interesting exceptions, including the beginning of the last week of the 2006

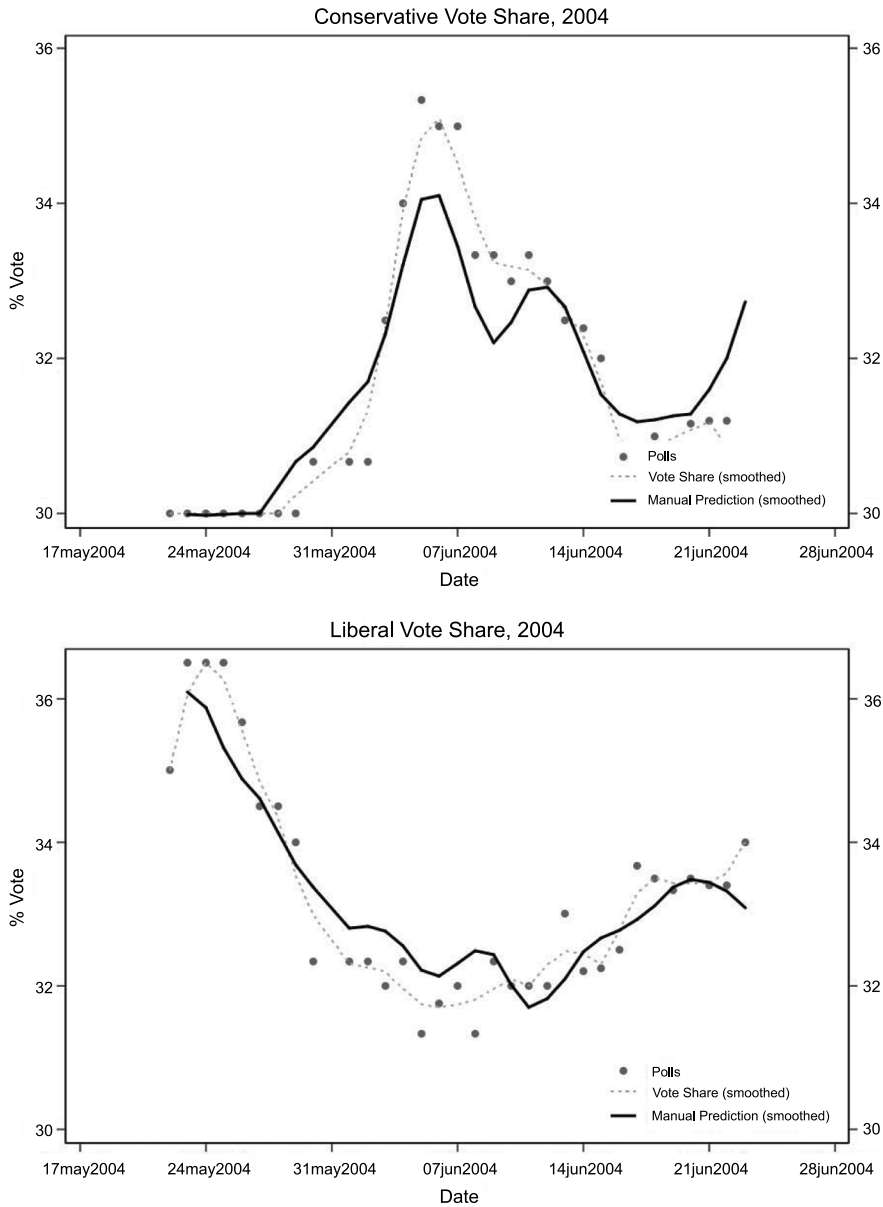

Figure 2. 2004 predictions, media content only.

campaign and last few days of the 2004 campaign. In 2006, media tone underestimated those leaning toward a Conservative vote, and overestimated those favoring the Liberals. The reverse happened in the final days of the 2004 campaign. We can only speculate that this may have represented the point in these campaigns when

vote intentions solidified, while media content did not. Voters may have heard enough after more than five weeks of daily campaign coverage. Media tone, in the final days of this campaign, appears to play a less prominent role in the calculus of Canadian voters. But, on the whole, the media's version of how well the parties' campaigns were going predicted vote intentions to a considerable degree in both of these elections. The closeness with which the prediction – which relies on media data only – tracks actual vote shares is rather striking.

Conclusions

This series of analyses has pointed to the potential for media content analysis in campaign-period vote predictions. We have found a good degree of predictive power using relatively simple models of lagged media content (and a series of dummies for pollster effects). The performance of these models is rather impressive, particularly given that they include only pollster effects and media content measures. And, importantly, even when a lagged dependent variable is included, media content improves the accuracy of predictions.

There is certainly some room for advancement where the content analysis is concerned, particularly the potential to automate the coding of party and leader tone. This possibility would allow us to expand this test to include a much bigger sample of election campaigns. There is a vast literature in computational linguistics suggesting that, while difficult, it may be possible to capture election tone reliably using precise software. Indeed, it may suffice to build a dictionary of positive and negative words,[18] and then look at the proximity of these words to party and leader names.[19] Natural language processing techniques may have even more potential.

Even with our more labor-intensive manual coding process, however, preceding results make clear that the general tone of major newspapers in Canada can help predict shifts in vote intentions.[20] Whether this is a media effect is perhaps not clear. The argument against media effects is that, rather than lead opinion, media content happens to capture and arrange in a readily quantifiable form the evolving mood of the campaign. Journalists are highly attentive to both the campaign and their audience. They react quickly, and measurably, to shifts in the mood of the campaign. They also likely react quickly to shifts in public opinion – not just public opinion generally, but to "opinion leaders".[21] Media content may as a consequence not lead so much as mirror, albeit mirror very efficiently. The public opinion captured in polls is in contrast rather slow and clumsy. So media content leads opinion, but perhaps only in a statistical sense. (Another related possibility is that journalists are not just interpreting the campaign but trying to forecast it. Media coverage may incorporate not just current information, but expectations as well.)

The argument for media effects is nevertheless strong. We begin with a basic fact: most information about the campaign that citizens receive comes from mass media; it follows that almost all movement over a campaign is a media effect.[22] This is a relatively broad definition of media effects, admittedly. It does not

distinguish between mass media acting simply as a conduit for information coming from parties and mass media playing a more active role in selecting and defining the campaign. But note that the kind of information we are extracting from news stories – affect-laden vocabulary – is likely to capture the part of media content relating to description and interpretation. That our media measure is likely to capture evaluative language may make more likely the possibility that the media–opinion link discovered here is indeed a causal one.

We clearly lean towards the media effects story, then, though we cannot entirely refute the possibility that media simply reflect evolving trends. For now, knowing that there is a strong connection between the tone of media content and vote intentions may have to suffice. Note, however, that the strength of that connection is great enough that vote predictions based on media data are clearly possible. That this is true speaks, in a methodological sense, to the value of media content analysis in campaign prediction and, in a substantive way, to the importance of media in modern election campaigns.

Acknowledgements

A previous version of this paper was presented at the Annual Meeting of the Canadian Political Science Association, Vancouver BC, 4–6 June 2008. We are grateful to Mark Pickup for providing the dataset of Canadian election polling results, to John Galbraith for comments on an early draft, and to the *Journal*'s anonymous reviewers. This work was funded in part by the Fonds québécois de la recherche sur la société et la culture, the Donner Canadian Foundation, the McGill-Max Bell Strategic Initiative, and the McGill Institute for the Study of Canada.

Notes

1. Horserace coverage has received a particularly large amount of attention in the literature (see, e.g., Craig, 2000; Fletcher, 1981, 1991; Graber, 1976; Jamieson, 1992; Mendelsohn, 1993; Patterson, 1993; Wilson, 1980).
2. There are of course vast literatures on the link between media and opinion in campaigns (for work in Canada see, e.g., Blais & Boyer, 1996; Mendelsohn, 1994, 1996; Mendelsohn & Nadeau, 1997; Johnston et al., 1992; Wagenberg et al., 1988; for work elsewhere see, e.g., Brians & Wattenberg, 1996; Druckman, 2004; Krosnick & Kinder, 1990; for a more thorough review of the earlier US literature, see Weaver, 1996).
3. All polls are of course in the field for several days, so we have indexed polls for this dataset in a way that captures as accurately as possible the timing of shifts in opinion. For polls in the field over a three-day period (most are), we use the middle day. For polls in the field for two or four days, we use the second or third day, respectively.
4. These data are distributed by the Media Observatory at the McGill Institute for the Study of Canada, available online at <http://www.mcgill.ca/misc/research/media-observatory>.
5. All measures included in this analysis achieved an appropriate level of reliability. Detailed methodological information is available at the website for the Media Observatory at the McGill Institute for the Study of Canada.
6. We consider, for the purposes of this article, only stories mentioning at least one of the two major parties and their leaders. This includes the incumbent Liberal Party (Paul Martin), and the Conservatives

(Stephen Harper), the largest opposition party represented in the House of Commons prior to each election campaign kickoff.

7. This is not the only conceivable measure of campaign's tone or mood, of course (see Ridout & Franz, 2008, for an overview of campaign tone measurement in the media and politics literature.
8. The manually coded data used here capture horserace coverage generally, but actually do not record direct mentions of polls. That said, the proportion of coverage dealing with polls is relatively easy to assess using a simple automated search. In this case, we rely on an existing dataset of all election-related stories in the five English-language newspapers (available through the Media Observatory at the McGill Institute for the Study of Canada), and a search identifying all articles mentioning the word "poll" more than once.
9. And note that lagged poll results are included in the final prediction model (discussed in the Analysis section). If lagged media content matters to vote share predictions, above and beyond lagged vote shares themselves, then there is strong evidence that there is some other component of media content (that is, not just polling reports) that matters.
10. We have also conducted a series of analyses making use of Bayesian statistics, similar to recent work by Pickup and Johnston (2007) and Jackman (2005). Using vague priors centered at zero with a low level of precision, the performance of our predictions is for the most part unaffected.
11. Note also that there is another aspect of error in polling results that is rarely discussed but that is often implicitly accepted in public opinion research. Because respondents may be in a better position to accurately express vote intention later in the campaign, survey responses may be more reliable closer to the election date. As a consequence, time series of vote intentions may be prone to temporal heteroskedasticity – random error variance may not be constant across time. This violates a critical assumption of OLS, but we do not address it here. We simply assume temporal homoskedasticity.
12. Note that for "net tone" we lump together all articles from all newspapers – we do not give newspapers different weights based on audience reach, nor do we distinguish between the potentially different content in different newspapers. While newspapers differ in levels of tone for different parties (Soroka & Andrew, 2010), however, they follow very similar trends over the campaign. It is not clear that there is much to gain by looking at newspapers separately.
13. This choice was based on a series of cross-correlograms exploring the relationships between (a) the two party vote share (dependent) variables, and (b) each of the four leader and party tone (independent) variables, in both 2004 and 2006. That said, the structure of our prediction models present some difficulties for diagnostics such as bivariate cross-correlograms, since in each model a single party's vote share is regressed on four independent variables: lagged values of both party and leader tone, for both parties. Our decision to use lags *4–6* was thus also based in large part on tests of various alternative forms of the final prediction models, e.g., lags *3–5*, lags *5–7*, a series of models using just two lags, and so on.
14. Preliminary tests confirmed that using a single lag for vote intentions was all that was required – once vote intentions at *t–4* are included, vote intentions at *t–5* and *t–6* have no substantive effect.
15. Note that these results cannot rule out the possibility that media content is at least partly affected by prior opinion (see also the discussion of news stories about polls in the *Data* section above). Our focus on prediction makes this somewhat of a secondary issue. That said, relatively simple Granger causality tests in which each series is regressed on lagged values of itself and the other series suggest that media does in most cases "Granger-cause" opinion. And the fact that media matter, above and beyond lagged opinion (see results below), also suggest a largely uni-directional causal effect.
16. Note that much of the campaign tone during these campaigns was negative. Bad news typically outweighed good news stories for all the major parties and leaders, as we would expect. The negativity bias in election reporting does not change our expectation. But, in practical terms, it does means that we expect that *less negative* Conservative tone is positively related to Conservative vote intentions and vice versa.
17. On the value of the MAE and SEE (Standard Error of the Estimate) as goodness of fit measures in prediction and forecasting, see Krueger & Lewis-Beck, 2005.

18. There exist many such dictionaries (see, e.g., Stone et al., 1966; Strapparava & Valitutti, 2004; Mergenthaler, 1996, 2008; Pennebaker et al., 2001; Martindale, 1975, 1990; Whissell, 1989; Bradley & Lang, 1999).
19. On proximity-based automated assessments of tone, see, e.g., Pang et al., 2002; Mullen & Collier, 2004.
20. It is worth stressing that this finding is based on newspaper-supplied election news. It is reasonable to expect that our results will actually strengthen as we incorporate news from other sources.
21. On "opinion leaders", see work on the two-step flow in political communications (esp. Lazarsfeld et al., 1944 and Katz & Lazarsfeld, 1955).
22. Assuming the movement is not just random, of course; that is, assuming that movement has something to do with the campaign.

References

Andrew, Blake (2007) Above the fold and behind the link: a comparative analysis of election shortcuts in newspapers and on the web. Paper presented at the 4th European Consortium for Political Research (ECPR) General Conference. Pisa, Italy.

Andrew, Blake C., Maioni, Antonia & Soroka, Stuart (2006) Just when you thought it was out, policy is pulled back in. *Policy Options*, 27(3), pp. 74–79.

Baumgartner, F. R. & Jones, B. D. (1993) *Agendas and Instability in American Politics* (Chicago, IL: University of Chicago Press).

Behr, Roy L. & Iyengar, Shanto (1985) Television news, real-world cues, and changes in the public agenda. *Public Opinion Quarterly*, 49(1), pp. 38–57.

Blais, Andre & Boyer, M. Martin (1996) Assessing the Impact of televised debates: the case of the 1988 Canadian election. *British Journal of Political Science*, 26(2), pp. 143–164.

Brady, Henry E. & Johnston, Richard (1987) What's the primary message: horse race or issue journalism?, in: Gary R. Orren & Nelson W. Polsby (eds) *Media and Momentum: The New Hampshire Primary and Nomination Politics* (Chatham, NJ: Chatham House), pp. 127–186.

Bradley, M. M. & Lang, P. J. (1999) *Affective Norms for English Words (ANEW): Stimuli, Instruction Manual and Affective Ratings.* Technical report C-1, Gainesville, FL: The Center for Research in Psychophysiology, University of Florida.

Brians, Craig Leonard & Wattenberg, Martin P. (1996) Campaign issue knowledge and salience: comparing reception from TV commercials, TV news and newspapers. *American Journal of Political Science*, 40(1), pp. 172–193.

Craig, Richard (2000) Expectations and elections: how television defines campaign news. *Critical Studies in Mass Communication*, 17(1), pp. 28–44.

Druckman, James N. (2004) Priming the vote: campaign effects in a U.S. senate election. *Political Psychology*, 25(4), pp. 577–559.

Erikson, Robert S. & Wlezien, Christopher (1999) Presidential polls as a time series: the case of 1996. *Public Opinion Quarterly*, 63(2), pp. 163–177.

Fan, David P. & Norem, Lois (1992) The media and the fate of the Medicare Catastrophic Extension Act. *Journal of Health Politics Policy and Law*, 17(1), pp. 37–70.

Fletcher, Frederick (1981) Playing the game: the mass media in the 1979 campaign, in: Howard R. Penniman (ed.) *Canada at the Polls, 1979 and 1980: A Study of the Genera Elections* (Washington, DC: American Enterprise Institute for Public Policy Research).

Fletcher, Frederick (1991) *Reporting the Campaign: Election Coverage in Canada* (Toronto: Dundurn Press).

Goodyear-Grant, Elizabeth, Soroka, Stuart & Maioni, Antonia (2004) The role of the media: a campaign saved by a horserace. *Policy Options*, 25(8), pp. 86–91.

Graber, Doris A. (1976) Press and TV as opinion resources in presidential campaigns. *Public Opinion Quarterly*, 40(3), pp. 285–303.

Iyengar, Shanto (1996) Framing responsibility for political issues. *Annals of the American Academy of Political and Social Science*, 546(1), pp. 59–70.

Jackman, Simon (2005) Pooling the polls over an election campaign. *Australian Journal of Political Science*, 40(4), pp. 499–517.

Jamieson, Kathleen Hall (1992) *Dirty Politics: Deception, Distraction, and Democracy* (New York: Oxford University Press).

Jamieson, Kathleen Hall & Cappella, Joseph N. (1998) The role of the press in the health care reform debate of 1993–1994, in: Doris A. Graber, Dennis McQuail & Pippa Norris (eds) *The Politics of News: The News of Politics* (Washington, DC: CQ Press), pp. 110–131.

Johnston, Richard C.G., Blais, André, Brady, Henry E. & Ciête, Jean (1992) *Letting the People Decide: Dynamics of a Canadian Election* (Montreal, QC: McGill-Queen's University Press).

Katz, Elihu & Lazarsfeld, Paul F. (1955) *Personal Influence: The Part Played by People in the Flow of Mass Communication* (New York: Columbia University Bureau of Applied Social Research).

Krosnick, Jon A. & Kinder, Donald R. (1990) Altering the foundations of support for the president through priming. *The American Political Science Review*, 84(2), pp. 497–512.

Krueger, James S. & Lewis-Beck, Michael (2005) The place of prediction in politics. Paper presented at the Annual Meeting of the American Political Science Association, Washington, DC.

Lazarsfeld, Paul F., Berelson, Bernard & Gaudet, Hazel (1944) *The People's Choice* (New York: Columbia University Press).

Martindale, C. (1975) *Romantic Progression: The Psychology of Literary History* (Washington, DC: Hemisphere).

Martindale, C. (1990) *The Clockwork Muse: The Predictability of Artistic Change* (New York: Basic Books).

McCombs, Maxwell E. & Shaw, Donald L. (1972) The agenda-setting function of the mass media. *Public Opinion Quarterly*, 36(2), pp. 176–187.

McDermott Monika L. & Frankovic, Kathleen A. (2003) Horserace polling and the survey method effects: an analysis of the 2000 campaign. *Public Opinion Quarterly*, 67(2), pp. 244–264.

Mendelsohn, Matthew (1993) Television's frames in the 1988 Canadian election. *Canadian Journal of Communication*, 18(2), pp. 149–171.

Mendelsohn, Matthew (1994) The media's persuasive effects: the priming of leadership in the 1988 Canadian election. *Canadian Journal of Political Science*, 27(1), pp. 81–97.

Mendelsohn, Matthew (1996) The media and interpersonal communications: the priming of issues, leaders, and party identification. *The Journal of Politics*, 58(1), pp. 112–125.

Mendelsohn, Matthew & Nadeau, Richard (1997) The religious cleavage and the media in Canada. *Canadian Journal of Political Science*, 30(1), pp. 129–146.

Mergenthaler, E. (1996) Emotion-abstraction patterns in verbatim protocols: a new way of describing psychotherapeutic processes. *Journal of Consulting and Clinical Psychology*, 64(6), pp. 1306–1315.

Mergenthaler, E. (2008) Resonating minds: a school-independent theoretical conception and its empirical application to psychotherapeutic processes. *Psychotherapy Research*, 18(2), pp. 109–126.

Mullen, A. & Collier, N. (2004) Incorporating topic information into sentiment analysis models. In Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP 2004), Barcelona, Spain.

Pang, Bo, Lee, Lillian & Vaithyanathan, Shivakumar (2002) Thumbs up? Sentiment classification using machine learning techniques. In Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP-02), Philadelphia, PA.

Patterson, Thomas E. (1993) *Out of Order* (New York: Knopf).

Pennebaker, James W., Francis, Martha E. & Booth, Roger J. (2001) *Linguistic Inquiry and Word Count: LIWC 2001*. Mahway, NJ: Erlbaum Publishers.

Pickup, Mark & Johnston, Richard (2007) Campaign trial heats as electoral information: evidence from the 2004 and 2006 Canadian federal election. *Electoral Studies*, 26(2), pp. 460–476.

Ridout, Travis N. & Franz, Michael (2008) Evaluating measures of campaign tone. *Political Communication*, 25(2), pp. 158–179.

Riffe, Daniel, Lacy, Stephen & Fico, Frederick (2005) *Analyzing Media Messages: Using Quantitative Content Analysis in Research.* 2nd ed. Mahwah, NJ: Lawrence Erlbaum.

Robinson, Michael J. & Sheehan, Margaret A. (1983) *Over the Wire and on TV: CBS and UPI in Campaign '80* (New York: Russell Sage Foundation).

Stone, P. J., Dumphy, D. C. & Ogilvie, D. M. (1966) *The General Inquirer: A Computer Approach to Content Analysis* (Cambridge, MA: MIT Press).

Soroka, Stuart (2002) *Agenda-Setting Dynamics in Canada* (Vancouver, BC: University of British Columbia Press).

Soroka, Stuart (2003) Media, public opinion, and foreign policy. *Harvard International Journal of Press and Politics*, 8(1), pp. 27–48.

Soroka, Stuart & Andrew, Blake (2010) Media coverage of Canadian elections: horserace coverage and negativity in election campaigns, in: Linda Trimble & Shannon Sampert (eds) *Mediating Canadian Politics* (Toronto: Pearson), pp. 113–128.

Strapparava, Carlo & Valitutti, Alessandro (2004) WordNet-affect: an affective extension of WordNet. In Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004), Lisbon, IT.

Wilson, R. Jeremy (1980) Media coverage of Canadian election campaigns: horserace journalism and the meta-campaign. *Journal of Canadian Studies*, 15(4), pp. 56–68.

Wagenberg, Ronald H., Soderlund, Walter C., Romanow, Walter I. & Briggs, E. Donald. (1988) Campaigns, images and polls: mass media coverage of the 1984 Canadian election. *Canadian Journal of Political Science*, 21(1), pp. 117–129.

Weaver, David H. (1996) What voters learn from media. *Annals of the American Academy of Political and Social Science*, 546, pp. 34–47.

Whissell, Cynthia (1989) The dictionary of affect in language, in: R. Plutchnik & H. Kellerman (eds) *Emotion: Theory and Research* (New York: Harcourt Brace), pp. 113–131.