

Supplemental Materials for “Validating Estimates of Latent Traits From Textual Data Using Human Judgment as a Benchmark

September 24, 2012

1 Results before cleaning the Gormley Text

Our first attempt to scale the 14 budget speeches uses the unmodified budget speeches in their entirety, based on the reasonable assumption that each speech followed the guidelines set for the debate and spoke only on the key issue at hand: whether to argue for passing Ireland’s most austere budget in history, or whether to oppose the government and its budgetary response to the financial crisis and argue against it.

In this estimation prior to editing the speech of Green party leader John Gormley, his speech was scaled by the Poisson model as being the most supportive of the budget, even more strongly than Finance minister Lenihan and Prime minister Cowen. Given the acknowledged reluctance of the Greens in supporting the budget, this result is difficult to square with either the human judgments or any general understanding of Irish politics of the publicly stated positions of the Greens or Gormley specifically.

Given the inconsistent placement of Gormley’s speech by the Poisson scaling model, we immediately suspect something may be qualitatively different about his speech. With a little investigation, the reasons for this become clear. It turns out that this is exactly the case. John Gormley was not only the Green party leader, but also was the Minister for the Environment, Heritage and Local Government. As Environment minister and Green party leader, Gormley not only was charged with maintaining a faltering image of his party as having joined the government in order to effect more environmentally responsible policies, but also faced mounting pressure to bring Ireland in line with EU water quality standards.

In his full speech of 2,998 words, Green Party leader suddenly ceased to express a position on the budget, and instead began to speak on his department’s implementation of household water meters, as part of his Ministerial remit. After introducing his speech with “As leader of the Green Party I want to take this opportunity to set out my party’s position on Budget 2010”, Gormley went on to comment extensively about Ireland’s challenges in meeting EU water quality standards, the planned introduction of household water meters, and policies for flood protection. This additional, off-topic text added 2,130 additional words, including 568 new word types.¹ When we removed this text and scaled Gormley’s speech without the middle section discussing water quality, his position moved to a position consistent with his fellow Green party members as well as the the human placements, as indicated in Figure 1b. Directly comparing the results from the Poisson scaling both with

¹This text, which was unconnected to the budget debate, was not included in the human placement exercise.

and without this speech, we see that this cleaning up of the texts moves the positional estimate of his speech (on the *y*-axis) to a position ordinally equivalent to that from the human estimates and consistent with his cabinet responsibility as the leader of the junior coalition party.

In this exercise, we had both a great deal of substantive political knowledge about each speaker, as well as human placements from the qualitative exercise to indicate that the scaled position of Gormley’s unmodified speech was out of place. In many examples scaling large quantities of texts, however, this is unlikely to be the case, such as Proksch and Slapin (2010)’s scaling of all text in the European Parliament. In any case, the lesson is clear: one dimensional scaling models produce invalid placements when the text concerns more than one topic or dimension. The Poisson scaling model is prone to being overly sensitive to differences in observed word frequencies, even when those differences do not reflect the dimension of difference sought by the analyst. Statistical scaling models such as the Poisson model claim as an advantage that they are “unsupervised”, in the sense of not requiring reference texts as in Laver, Benoit and Garry (2003), but as the Gormley examples demonstrates, careful supervision and selection of the texts remains necessary to ensure that the dimension of difference being scaled is indeed the dimension of difference we wish to measure.

2 Detecting coder effects

Finally, in order to check for any effects of prior knowledge, we also compared the sample of those reporting prior knowledge of Irish politics to those without, as well any “experience” effect that might come from having coded political text—in any sort of exercise—before. Of our sample of 19 human raters, 10 reported knowledge of Irish politics, typically the readers from Trinity College Dublin who studied this area, although several of the non-Irish students at Trinity reported knowing essentially nothing about Irish politics. Of the combined sample, only three readers reported having any experience with coding or content analyzing text before. Testing the differences in mean placements for each of 12 texts that humans placed, we detected not a single statistically significant difference attributable to either prior political knowledge or prior coding experience. (These tests applied a standard *t*-test assuming unequal variances. Full results are contained in the replication materials.) The human readers seemed to be taking their cues from the texts exclusively, without being influenced by external or prior substantive or methodological knowledge.

A Appendix: Human Validation Experiment Instructions

SUMMARY CODING INSTRUCTIONS – READ CAREFULLY

The following texts consist of 14 speeches from the Irish legislature during the debate over the adoption of the 2010 budget. These speeches took place on December 10, 2009 following the introduction of the budget by Finance Minister Brian Lenihan on December 9.

In this coding exercise, you will be asked not to code text units, but rather to read each speech and then answer a brief questionnaire about the text as a whole. The questionnaire is designed to get two key pieces of information:

1. how pro- versus anti-budget you feel was the speech overall;

2. whether and how confident you feel in distinguishing selected pairs of speeches from one another.

The first speech you read will be an extract of the speech of the Finance Minister Brian Lenihan who introduced the budget.

The second speech comes from Taoiseach (Prime Minister) Brian Cowen, who (not surprisingly) advocated passage of the budget. It is not necessary for you to score this speech (see instructions below) since we will take for granted that it expresses the most pro-budget position possible.

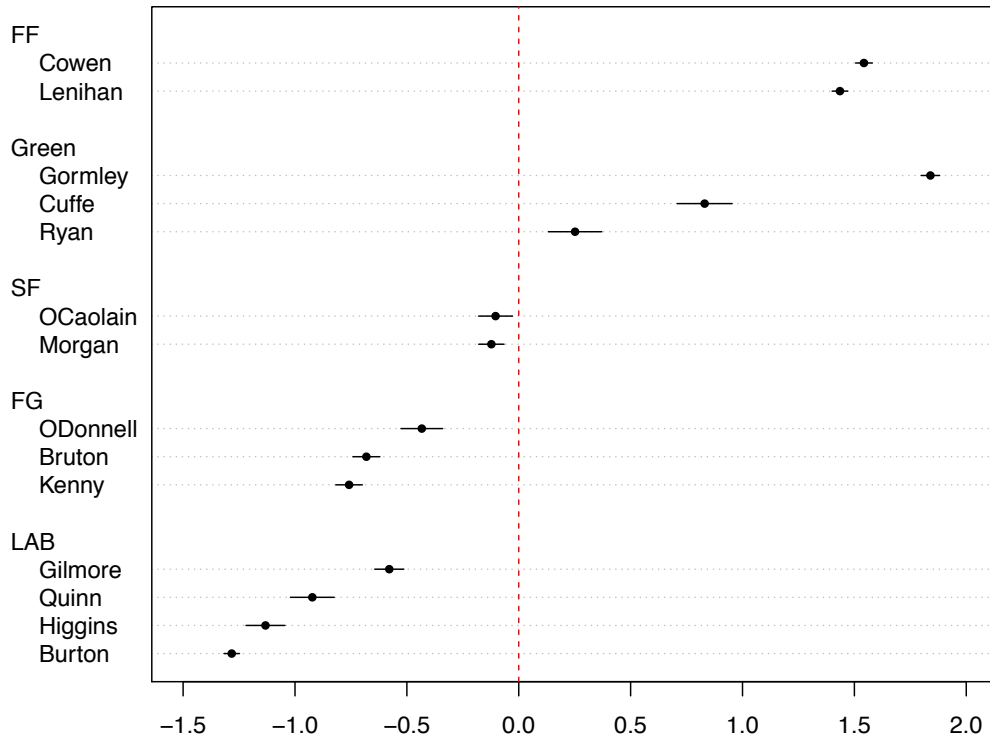
The remaining twelve speeches you will be asked to read and to score using the short questionnaire that follows each text. Please complete this in the spaces provided (following each text) using pen or pencil.

As you read each speech, you are encouraged to make marginal notes, use a highlighter pen, or underline passages of the text that might provide a clue as to why you have made your decision. Any notes you make will be read with great interest when we receive your completed coding kit.

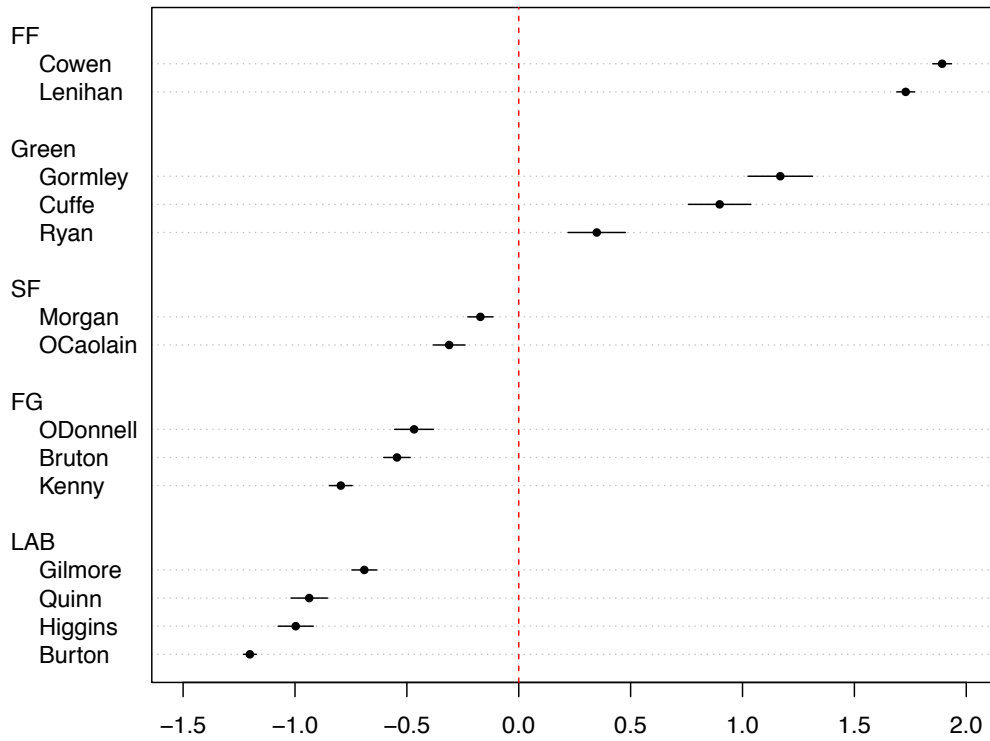
References

Laver, Michael, Kenneth Benoit and John Garry. 2003. "Estimating the policy positions of political actors using words as data." *American Political Science Review* 97(2):311–331.

Proksch, S.-O. and J. B. Slapin. 2010. "Position taking in the European Parliament speeches." *British Journal of Political Science* 40(3):587–611. forthcoming.



(a) Poisson Scaling Results Using Unmodified Speeches



(b) Poisson Scaling Results With Gormley Water Meter Text Removed

Figure 1: *Poisson scaling results on (a) Full Gormley text and (b) Gormley text with water meter text removed.*

QUESTIONS

23a. Comparing Speech 13 to Speech 12, was either speech more strongly pro-budget than the other?

- Could not detect a difference
- Speech 13 more pro-budget
- Speech 12 more pro-budget

23b. If you answered that one was more pro-budget than the other, how confident are you that there was a detectable difference?

1
2
3
4
5
6
7
8
9
10

Not at all confident
Very confident

Figure 2: Pairwise comparison question example from human coding experiment.

QUESTIONS: PLACEMENTS

Now consider that each speech can exhibit a range of support or opposition for the budget, where 0 is completely in support of the budget (a total lack of opposition) and 100 is completely opposed. The speeches by Finance Minister Lenihan and Prime Minister Cowen are both fixed at 0. Please place the other speeches on a 0-100 point scale, by drawing the score on the approximate place on the scale. If you are unsure, you may indicate a range by drawing a line across a range of values, and you may indicate a most likely position if you feel this is possible.

EXAMPLES

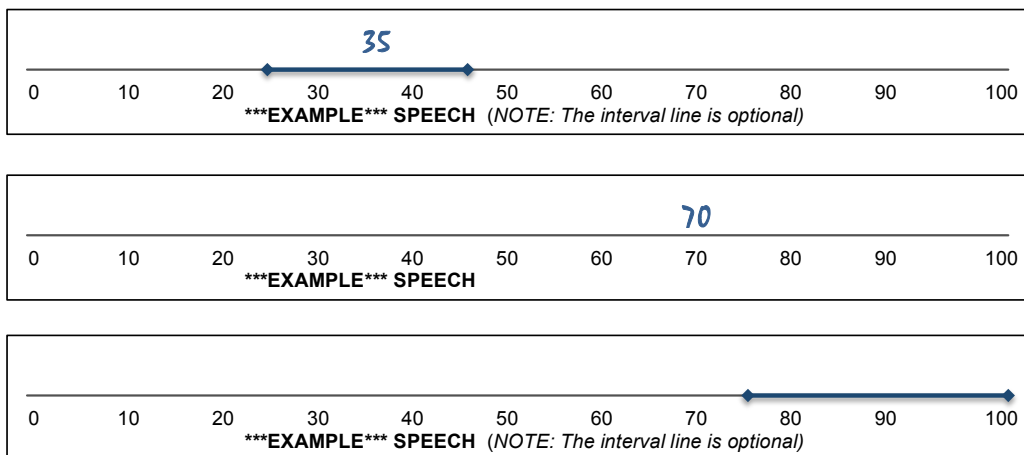


Figure 3: Scale question example from human coding experiment.

	Lenihan FF	Cowen FF	Gilmore LAB	Kenny FG	Bruton FG	Quinn LAB	Higgins LAB	Burton LAB	ODonnell FG	Morgan SF	O'Caolain SF	Gormley Green	Cuffe Green	Ryan Green
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Lenihan FF	1													
Cowen FF	2	i												
Gilmore LAB	3		L											
Kenny FG	4		L	L,e										
Bruton FG	5			e	e,i									
Quinn LAB	6					x								
Higgins LAB	7						e,i							
Burton LAB	8					x	i	e,i						
ODonnell FG	9								x					
Morgan SF	10			x										
O'Caolain SF	11		L	L	L					e,i				
Gormley Green	12		L,e	L	L						L			
Cuffe Green	13											e,i		
Ryan Green	14									e		i	e,i	

Guide: e Edge comparison
i Internal party comparison
L Party leader comparison

Figure 4: *Pairwise combinations in the qualitative questionnaire.*