Appendix A: Additional Detail on Speech Codification

This appendix describes some additional details on the method used to collect speeches from the Canadian House of Commons.

Computer code written in the Python language processed all interventions made in the Canadian House of Commons between 2006 and 2011. The code counts the number of sentences in each speech, the number of words, the number of characters and the number of syllables per word. The syllable counting regime is as follows. First, each word is looked-up in Python's Natural Language (Carnegie Mellon) dictionary. If the word is in the dictionary, the dictionary pronunciation guide is used to estimate syllables. If the word is not contained in the dictionary, it is separated into vowel clusters. As an example, "turtle" has two vowel clusters – "u" and "e" – and therefore two syllables. Syllables in "turtle" would be counted properly. How- ever, there are exceptions such as "delicious". In this case, the third vowel cluster is actually two syllables, but the algorithm estimates three syllables when there would actually be four ("dee" + "lish" + "ee" + "us"). A cursory inspection led us to believe that the number of exceptions – i.e., words not contained in a dictionary *and* those not correctly coded according to vowel clusters – is small and would require significant manual processing to rectify. Therefore, we maintain the count created by the algorithm. These sentence, word, character and syllable count statistics are sufficient to construct various measures of text complexity. Simple measures, for example, are the number of words per sentence or the number of syllables per word. Slightly more complex statistics, like those used in the main text and in Appendix B, include the Flesh-Kincaid grade level score. These readability scores combine the components of a speech into an index of speech complexity. The Python code will be included with the data.

Finally two additional dummy variables are created. First, an indicator distinguishes between a speech made in Question Period versus a committee meeting or alternative debate. Further, it is straightforward to determine whether the speaker is a member of the Government or Opposition.