Instructions for Reviewing and Annotating Different Perspectives on an Issue

Objective

This project uses a computerized method to identify different political perspectives on an issue. The perspectives are represented using words strongly associated with them ("top words") or words characteristic of that perspective, as determined by the computational method. You will be reviewing the output of this computational model to help determine:

1. If the top words for a given perspective are clearly connected with the given issue label, i.e., are the words shown representative of what could be said on that particular issue?
2. If the two perspectives shown represent ideologically polarized ways to talk about the given issue.
3. If each of the two perspectives, represent a liberal/conservative perspective on the issue.

This task will be carried out in two different spreadsheets, corresponding to perspectives from two different sources of information (Twitter and Congressional speeches) but the instructions will apply the same way to both files.

This analysis requires an understanding of American politics and partisan polarization.

Background

The computational model works in two stages. In the first stage, it automatically discovers issues, by analyzing a dataset of texts to produce "topics" or categories corresponding to those issues within the collection of text. These automatically discovered topics have then been curated by people with subject matter expertise in political science

In the second stage, the model automatically discovers different ways of talking about the same issue, selecting texts that prioritize some set of words over other words that are salient to that issue. This often yields groups of words that are related to different perspectives on the same issue. A systemic differentiation in perspectives might exist in political texts like congressional floor speeches, social media posts by legislators, etc. because of the potential for liberals and conservatives to talk about the same issue in a different way. This technique therefore has the potential to discover politically polarized perspectives on an issue, however such capability

needs to be validated. Human annotators, like you, are helping us to validate that the proposed sets of perspective-related words do, in fact, correspond to different perspectives.

Validation includes determining whether the two perspectives are related to a specified issue, as well as which perspective (that is, which of the two sets of perspective-related words) corresponds to which ideology, liberal versus conservative. Your rating the the polarization or ideology of these perspectives will also help us determine whether a clear ideological difference exists between the discovered perspectives on an issue.

The materials you'll be working with

Along with these instructions, you should have received one zip folder which will have a PDF copy of these instructions and two .xlsx files: **annotation_file_speeches.xlsx** and **annotation_file_tweets.xlsx**. Both these .xlsx files will be structured in exactly the same way, and the same instructions below should be followed for both files, one after the other.

How much time this will take

The length of time required for the labeling process is expected to be about 2 hours per file for a total of 4 hours. It is completely fine if the task ends up taking considerably less time.

Please keep track of the time you've spent working on the whole process (apart from reading these instructions) and report that to us along with your results. In the instructions below, you are asked to set a timer for each instance of annotation - this should help ensure not too much time is taken deliberating on your judgements.

Instructions

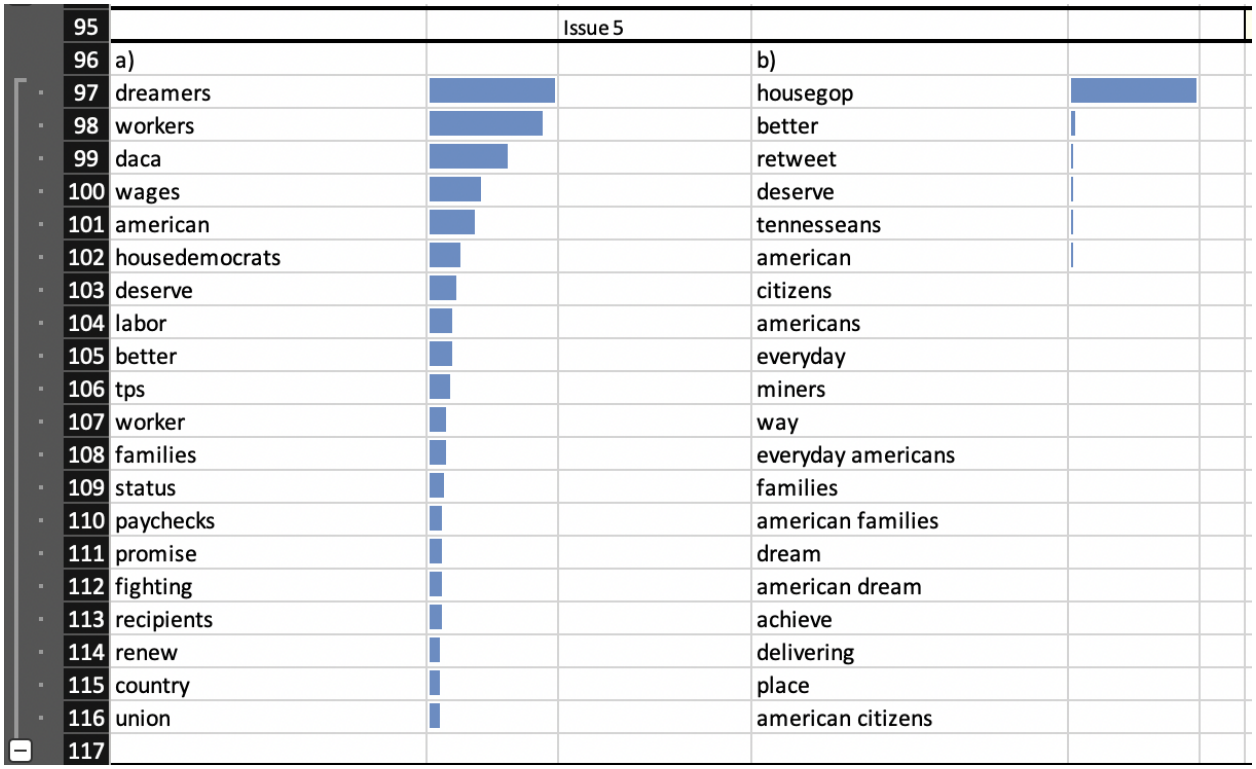
(Note that the steps will be followed twice - once for **annotation_file_speeches.xlsx**, and once for **annotation_file_tweets.xlsx**.)

NOTE: The below example is meant to be just an example – your labeling or any other interpretation does not need to mirror this example.

1. Open one of the two excel files.
2. Column C ("Issue") has the Issue number, e.g. "Issue 4". You should proceed issue by issue. Each issue has two perspectives - marked as "a)" and "b)".
3. Set a (new) timer for 3 minutes for each issue. You will follow Steps 4 through 8 for every issue, with a new 3 minutes for each issue.
4. To start annotating for an issue, first click the icon to expand rows, located towards the left and below the cell with "a)" under every issue number.

72			Issue 4			
73	a)			b)		
94						
95			Issue 5			
96	a)			b)		
117						
118			Issue 6			
119	a)			b)		
140						
141			Issue 7			
142	a)			b)		
163						
164			Issue 8			
165	a)			b)		
186						
187			Issue 9			
188	a)			b)		
209						
210			Issue 10			
211	a)			b)		
232						

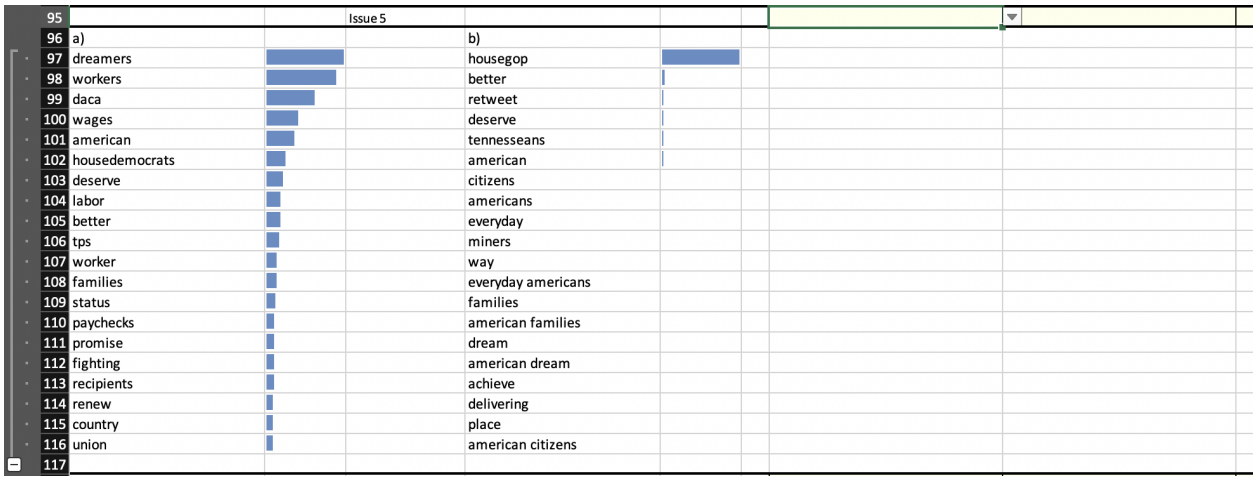
5. For example, we click the icon below a) in row 117 of the above image for “Issue 5”
-



95			Issue 5			
96	a)			b)		
97	dreamers			housegop		
98	workers			better		
99	daca			retweet		
100	wages			deserve		
101	american			tennesseans		
102	housedemocrats			american		
103	deserve			citizens		
104	labor			americans		
105	better			everyday		
106	tps			miners		
107	worker			way		
108	families			everyday americans		
109	status			families		
110	paychecks			american families		
111	promise			dream		
112	fighting			american dream		
113	recipients			achieve		
114	renew			delivering		
115	country			place		
116	union			american citizens		
117						

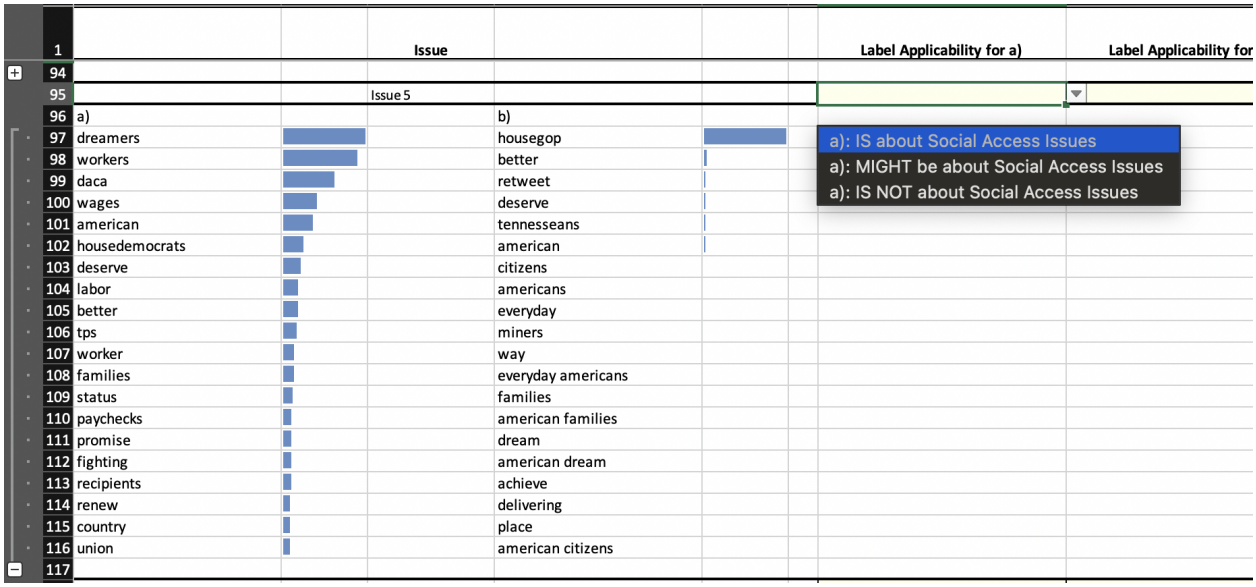
This expands hidden rows that will show the top characteristic words associated with perspectives a) and b) for Issue 5. Take a look at these top words (first before moving on to annotate for the issue you are currently at. The bars show the strength of the word for that perspective.

6. In this next step, you will decide whether a given issue label is applicable for each perspective – that is, is this group of perspective-related words about that issue? For each of the Columns G to K, the yellow highlighted cells have drop-down menus. Follow this example to access the drop-down menu in each case - e.g., for Column G ("Label Applicability for a)" - select the cell in column G for the row with the Issue number in it. You will then see the drop-down menu icon appear once you select the cell (G95 below) -



95		Issue 5				
96	a)		b)			
97	dreamers		housegop			
98	workers		better			
99	daca		retweet			
100	wages		deserve			
101	american		tennesseans			
102	housedemocrats		american			
103	deserve		citizens			
104	labor		americans			
105	better		everyday			
106	tps		miners			
107	worker		way			
108	families		everyday americans			
109	status		families			
110	paychecks		american families			
111	promise		dream			
112	fighting		american dream			
113	recipients		achieve			
114	renew		delivering			
115	country		place			
116	union		american citizens			
117						

Click the drop-down menu and select one of the options that appear -



1		Issue			Label Applicability for a)	Label Applicability for
94						
95		Issue 5				
96	a)		b)			
97	dreamers		housegop			
98	workers		better			
99	daca		retweet			
100	wages		deserve			
101	american		tennesseans			
102	housedemocrats		american			
103	deserve		citizens			
104	labor		americans			
105	better		everyday			
106	tps		miners			
107	worker		way			
108	families		everyday americans			
109	status		families			
110	paychecks		american families			
111	promise		dream			
112	fighting		american dream			
113	recipients		achieve			
114	renew		delivering			
115	country		place			
116	union		american citizens			
117						

7. Now proceed to select one of the options for each of the five yellow-highlighted the cells for the issue in the following order -
    a. Column G asks you to rate the "Label Applicability for a)". The drop down menu has three options - i. "a): IS about <Issue Label>", ii. "a): MIGHT be about <Issue Label>", iii. "a): IS NOT about <Issue Label>". Looking at the top words shown for the perspective a) of a particular issue number (columns A and B under the issue number), select one of the three options based on your judgment of whether the

Issue Label in the drop down menu is a suitable label for the underlying *concept* or category you think the words shown for “a)” represent. Put differently, is the shown <Issue Label> applicable to the words shown for the words under “a)” for that issue?

b. Column H asks for the same rating as G above, but for the perspective characterized by the top words under “b)” for that <Issue Label>. So, follow the same procedure as in 6a, but for perspective b) of the issue (columns D and E under the issue number).
c. Column I is titled “Ideological Polarization” and asks you to rate the ideological polarization represented by the two perspectives a) and b) for that issue. There are four options to choose from -
    i. *“a) and b) represent polarized perspectives on the issue: <Issue Label>”*
    Select this option if you think one of a) or b) represents a liberal perspective on the issue named by the issue label and the other one represents a conservative perspective on the issue. A liberal/conservative perspective here means that you think the words in a) or b) are how US liberal or conservative politicians would talk when speaking on that particular issue. Note that you should only even consider selecting this option if in the previous step, you said that <Issue Label> applies or might apply to a) and b) in the first place (i.e., you selected option i. or ii. for both the corresponding cells in G and H columns).
    ii. *“a) and b) represent SOMEWHAT polarized perspectives on the above issue”*
    Select this option if you think that liberal/conservative polarized perspectives *are* being captured in a) and b), *but* feel it is only being somewhat captured. In other words, if there is some ideological difference but it is not a lot and not as clearly ideologically apart, choose this option. This would usually mean that the two perspectives are only somewhat apart on the ideological spectrum of US politics according to you. By its very nature, this task is subjective and relies on your knowledge of US political rhetoric. Note that you should only consider selecting this option if <Issue Label> applies or might apply to a) and b) in the first place (i.e., you selected option i. or ii. for both the corresponding cells in G and H columns).
    iii. *“a) and b) DO NOT represent polarized perspectives on above issue”*
    Select this option if you think that a) and b) are not showing different ways US liberal and conservative politicians talk about the <Issue Label> according to you. This can happen in multiple ways: the <Issue Label> could be applicable to both a) and b) but there is no substantive ideological difference between the two. Or if you said that the <Issue Label> is NOT applicable for either of a) or b) (i.e., you selected option iii, which says “... IS NOT about <Issue Label>” for corresponding cell of *either* column G or column H), then a) and b) cannot represent polarized perspectives on that issue by default!

iv. *"Unsure"*
This option is to be selected when <Issue Label> applies or might apply to a) and b) in the first place (i.e., you selected option i. or ii. for both the corresponding cells in G and H columns), but you are simply not sure if a) and b) represent ideologically polarized perspectives on the issue or not. They could or could not be ideologically polarized but you simply cannot say for sure either way.

d. Column J is titled "Ideological Position expressed in a)", and asks you to judge the ideological position of the perspective a) shown with the top characteristic words for that particular issue. You will select one of the four options from the drop-down menu as described below:
   i. *"Liberal"*
   If the top characteristic words under a) on that <Issue Label> represent a liberal perspective to you (i.e., you think liberal politicians in the US would be likely to use those words when they are saying things pertaining to that particular issue). Note that this only applies if you selected option i. or ii. for corresponding Column G cell - the <Issue Label> applies or might apply to "a)".
   ii. *"Unsure"*
   If you selected option i. or ii. for corresponding Column G cell - the <Issue Label> applies or might apply to "a)" - but you are unsure if it represents a liberal or conservative perspective. This could happen for multiple reasons: maybe you feel that one of the two perspectives (a or b) is neither how liberal or conservatives speak on the issue, or you feel that both liberals and conservatives could be using those words when speaking on that issue, etc.
   iii. *"Conservative"*
   If the top characteristic words under a) on that <Issue Label> represent a conservative perspective to you, i.e., you think conservative politicians in the US would be likely to use those words when they are saying things pertaining to that particular issue. Note that this only applies if you selected option i. or ii. for corresponding Column G cell - the <Issue Label> applies or might apply to "a)".
   iv. *"N/A" - Not applicable*
   Select this by default if you selected option iii. for corresponding Column G cell - the <Issue Label> does not apply to what is shown under "a)" for the issue.

e. Column K is titled "Ideological Position expressed in b)", for the perspective b) shown with the top characteristic words for that particular issue, with the same four options as above but for "b)". Follow what is laid out in 6d above but replace "a)" with "b) and Column G with Column H.

8. Last, you can note any internal thoughts, comments, conflicts, etc. you had when doing the five annotations for that issue, in the corresponding cell of Column L titled "Notes/Comments". While this is optional to fill out, in cases where you felt it was not

easy or clear-cut when you selected an answer for cells in column G through to column K, you can write out why you made certain choices. Please feel free to scribble in this column during the option selection above - you do not need to fill this out only after going through columns G to K.

9. You are now done with the particular issue and can move on to the next one. First select the icon to the left and below the last word under “a)” of that issue to re-hide the rows containing words for the perspectives a) and b). Stop the timer for the particular issue and note the amount of time that was taken (to help have an estimate of the overall time it took for the full task - we do not need per-instance time taken).

95		Issue 5			
96	a)			b)	
97	dreamers			housegop	
98	workers			better	
99	daca			retweet	
100	wages			deserve	
101	american			tennesseans	
102	housedemocrats			american	
103	deserve			citizens	
104	labor			americans	
105	better			everyday	
106	tps			miners	
107	worker			way	
108	families			everyday americans	
109	status			families	
110	paychecks			american families	
111	promise			dream	
112	fighting			american dream	
113	recipients			achieve	
114	renew			delivering	
115	country			place	
116	union			american citizens	
117					

10. Repeat instructions 3 through 9 for the next issue.
11. Once you are done with all the issues in a file, review that you have selected an option for each of the yellow highlighted cells in the sheet, save your final completed file, and move to the other excel file and follow the same instructions above again.

Expected OUTPUT -

12. **Two completed, filled out** .xlsx files - **annotation_file_speeches.xlsx** and **annotation_file_tweets.xlsx**. In addition, in your email sending these annotated files, please **report the total time you spent working on this task (excluding reading instructions or taking breaks).**

Thank you for your help on this project!