**Assignment 1b – Due before class on February 10**

Learning Objectives:

1. Gain experience with qualitative discourse analysis
2. Gain more insight into what LDA variants are doing with text data

Description:

In this assignment you continue working with a data set where groups of 3 9th grade students worked together on an assignment related to the concept of diffusion. In the first part of the assignment, you evaluated the models in comparison with role labels, negotiation labels, or heteroglossia labels. In this assignment, instead of taking a top-down approach where you assume a set of hand assigned, pre-defined labels are an external standard for reasonableness, you will take a more bottom-up approach where you use Jim Gee’s methodology to let the text speak to you about what is happening socially in the text.

The foundation for your analysis will be a Jim Gee style analysis in the form you see in My Analysis.xls, which you will find in this assignment folder. Select two different chat episodes for your analysis. You can tell where a new chat episode is starting in the annotated data file because the parent id is -1. You should use the analysis you did for assignment 1a as a guide for selecting two chat episodes. Do so with the following method:

1. Based on the quantitative analysis you did, you should be able to select which topic model (LDA, Block HMM, or M4) showed up in your validation as best. If you did more than one prediction task, pick the task that had the best overall results.
2. Using that model, you should be able to rank episodes based on how well the model performed on them. Pick the episode the model performed best on and the episode the model performed worst on.

Now with the two chat episodes you picked, do the Jim Gee style analysis. Based on your analysis, what is different in what is happening within the two episodes? Can you see any explanation for why one fit the model better than the other one?

In class and in the discussion blog we discussed briefly why the Jim Gee style analysis leverages associations between language features that “feel” similar to what LDA (in one of the three forms) is doing on a conceptual level. Based on your Jim Gee analysis and where you see the computational models fitting or not what is happening in the discussions, what is your opinion now about whether there is a connection between Jim Gee’s approach and what LDA is doing?

**Deliverable:** What you will turn in is your write up where you will describe how you selected the 2 episodes, your Jim Gee style analysis for each episode, and your reflections about the differences between the two episodes and how this changes (or not) your understanding of LDA and what kinds of patterns it finds in texts. As in the previous assignment, you will turn this assignment in to the Turnin link on Blackboard.