Supervised project – Self-assessment report
TALECO

Maxime Méloux

June 2022

Group: Omar Cherif, Dalila Ladli, Maxime Méloux
Subject: Tracking Theoretical Shifts over Time: A Natural Language Processing Analysis of Cass R. Sunstein

I believe that the TALECO project was quite different from other ones, due to the nature of our collaboration with professor Ferey and the gap between the field of NLP and that of economics and political theory. As such, our relationship was a customer–consultant one rather than a student–supervisor one. However, I believe this was extremely beneficial to our team in several ways. Not only did it teach us about a field that was largely unknown to us, thus preparing us for the interdisciplinary work that is common in NLP, it was also a good example of a research scenario, with little existing literature. Most of the ideas had to come either from previous literature, from professor Ferey's intuitions, or from us directly.

At the end of the year, I believe I have a good understanding of the topic. Most of the analysis conducted in our project relied on our ideas rather than it did on any previous work. In addition, the time spent cleaning the corpus, as well as the resources given to us by professor Ferey, gave me a relatively good picture of the topic from the perspective of economics and political theory.

I believe our approach was quite similar to that of lab research. Despite having identified the relevant literature in the first semester, unforeseen issues kept occurring, forcing us to find new solutions. Furthermore, we constantly had to evaluate what to do at the current step, how to model the situation, what tools to use, and how to evaluate our results.

While we were able to work well in our group and to split the tasks in a sensible manner, I personally wish there had been more group work earlier in the year. I often felt that the workload wasn't balanced very well and it was difficult to find time for group meetings before the end of May.

Globally, I believe that we achieved what professor Ferey set out to do, and that our results are satisfactory. However, I regret that we didn't have more time

to work on the project before the end of May, and I believe that we could have found more robust and smarter ways to model the problem. While I understand the importance of having a good corpus, I also wish we could have spent less time on it and more on the modeling of the problem.

As for skills from the master, I believe they were all used during this project.

- *Analyze a problem that can be addressed by digital means and know how to collect the relevant data, adapting to the context (oral and/or written sources).* We estimate that about half of the time spent of the project was spent on corpus acquisition and preprocessing. We had to make use of several methods taught in the master's degree (scraping, regular expressions, use of language models) and this phase was a central part of our project.
- *Model a phenomenon or a problem, in particular with computer tools associated with mathematical tools or Linguistics.* Modeling was an important part of our project. We used several tools for it, both mathematical (sigmoid function, TF-IDF), computational (topic modeling) and linguistics (semantic networks).
- *Know how to collect, structure and represent data (sound, text, image, etc.).* We only had to deal with text in our project. Collection was done using semi-automated methods (scraping, virtual browsers) and the text were stored as plain text files, so not much was required in that regard.
- *Synergize multidisciplinary knowledge within innovative solutions (computer science, linguistic, artificial intelligence and mathematics).* This skill was crucial to our project. The problem in our project was not found in the existing literature, and some of our solutions were completely innovative and not relying on previous work. We made use of all four mentioned fields, as well as that of economics and political theory.
- *Implement algorithmic techniques, linguistic analysis, statistics and knowledge processing.* While we did not have to implement complex algorithmic techniques or linguistic analysis (thanks to existing libraries as well as Tropes), we had to implement a number of algorithms for statistics (sigmoid curve fit, error measurement). Knowledge processing also relied on the use of tools such as pandas.
- *Synthesize and present your work: explain and document the implementation of a technical solution.* Presenting all of our work and results in only 30 pages, a 20-minute presentation and a poster was immensely challenging, especially with the assumption that the reader has little to no technical knowledge. We believe that we applied this skill successfully.
- *Conduct a project independently or in a team.* Our work was mostly conducted individually, and we shared ideas during group meetings with professor Ferey as well as on our own.

- *Develop an argument with a critical mind.* Due to the lack of automated evaluation of our results (we could only rely on professor Ferey for checking the coherence and pertinence of our results), constant reassessment of our models was mandatory. We discussed a few key points of criticism in the report, and we are largely aware of the shortcomings and limitations of our approach.
- *Demonstrate autonomy and initiative.* As mentioned before, we had a consultant – customer relationship with professor Ferey, which led (most of the time) to us proposing solutions and professor Ferey vetting them. We therefore had to show autonomy and initiative by coming up with intelligent ways to model the problem.

Finally, I would like to re-iterate that I have found this interdisciplinary project to be extremely fruitful, and I would highly recommend to have one or two such projects (where the supervisor is not a computer scientist) every year.