

Advising Meeting

Mykel Loren Brinkerhoff

Outline

- Recap of what I did over the summer
- Working analysis of the data
- Projection of where next

Summer Recap

- This project is concerned with NegShifting and whether or not pronominal NegShifting is subject to prosodic factors as mentioned by Penka (2011) similar to pronominal OS in Scandinavian languages
- Took a long time to find a corpus that I could use for my paper
 - Most of the corpora that I found were very narrow and did not contain the breadth that I needed for a good statistical analysis
 - I finally came across the Swedish Culturomics Gigaword Corpus (Eide, Tahmasebi & Borin 2016)
 - * This was very useful because it consisted of five different genres from the 1950s to early 2010s
 - * This enables me to try and find genres that might correspond to different registers of Swedish. There is some discussion by Engels (2012) that there are differences in acceptability of NegShift based on register.
 - * I chose three different genres with the help of my intern
 1. **Social Media:** We assumed that this would be our best chance at getting results that would be of the lowest register and potentially showing more dialectal influence
 2. **Fiction:** We assumed that this second genre would represent a more neutral register. One that is neither too formal or too informal.
 3. **Government:** We assumed that this would be our best chance to see a high register or most formal variety. The other possibility was scientific writings. However, most of these were Wikipedia articles from 2015 and we felt that we would get more consistent and clearer evidence from the government genre.
- SIP
 - Worked with two interns. One left the program due to medical problems

- I focused on teaching them basics of syntax and how to write computer programs
- Both of my interns had very limited to no experience with programming. This took a good portion of our time in the program because I tried to limit the amount of time we were in meetings to avoid Zoom fatigue
- By the end of the program my intern that remained was able to write code to extract and organize lines from the corpus for NIs
- Presented at AMP 2020 with Eirik Tengesdal on our analysis of OS.
 - Received a lot of good feedback from several attendees.
 - Continuing with an analysis of more complex syntactic structures
 - This helps with understanding how weak pronouns interact with syntax and prosody and helps with understanding what things I need to look for in a possible prosodic answer to NegShifting.

Analysis

- A simple look at the distribution shows that there are more instances of what appears to be bear pronouns instead of complex DPs in the tokens that I extracted.
 - See Table 1 below
 - I am still working on the government files because there was an issue with extracting the genre from the corpus into a POS-tagged format for extracting NIs.
- This is making me think about how I need to actually going analyzing the results of the data.
 - I might be best to try and do a multiple part analysis. With comparing all NIs first for shifting. Followed by looking at the differences in percentages between the number of pronouns that shifting and the number of complex NIs that shifted.
 - I am planning on talking to several people that I know in the Stats department to see if this is the correct way to approach the data and to perform a statistical analysis.
 - This is making me think that it might be good to have Amanda or Matt on my QP committee as well to help with this portion of the analysis.

Table 1: Number of lines for each genre and whether the line contained a NI pronominal

Genre	Pronominals	Complex NIs
Fiction	10580	5174
Social	10889	7106
Government	?	?

Where next

- Extract the Government files and get the NIs from it.
- See if there is a way that I can write a program that will help with classifying whether or not a NI that I see in the line is shifted or not.
 - This will help speed up the process of annotating which means that I should then only need to pull a random sample and then see if the program correctly sorted the random sample or not.
- Continue to synthesis what has been written concerning NegShifting.
- Start to Outline the paper so I can work on adding parts as I go.

References

- Eide, Stian Rødven, Nina Tahmasebi & Lars Borin. 2016. The Swedish Culturomics Gigaword Corpus: A One Billion Word Swedish Reference Dataset for NLP. 5.
- Engels, Eva. 2012. Scandinavian Negative Indefinites and Cyclic Linearization: Scandinavian Negative Indefinites and Cyclic Linearization. *Syntax* 15(2). 109–141. <https://doi.org/10.1111/j.1467-9612.2011.00161.x>.
- Penka, Doris. 2011. *Negative indefinites* (Oxford Studies in Theoretical Linguistics no. 32). Oxford ; New York: Oxford University Press. 264 pp. <https://doi.org/10.1093/acprof:oso/9780199567263.001.0001>.