## ***Language Style Transfer between Speeches of US Presidents***

## **Final Project Proposal W266 Natural Language Processing**

Thong Bui, Jason Vantomme

**Overview**

Language style transfer is a focus area in machine learning research analogous to image style transfer; a technique made publicly accessible by applications like Prisma.1 Simply, language style transfer is the process of transferring a language style or set of styles onto another.2,3 Beyond the novelty of the test cases, the practical applications of language style transfer are real: author privacy through obfuscation, improved realism and variety in chatbot interactions, paraphrasing a text for a specific audience, or even providing character consistency across different writers in a canonical fictional universe (e.g. Star Wars).

We have chosen the speeches of President Trump and President Lincoln to serve as the style sources for transfer as these two speakers represent dramatically different styles and both exhibit unique characteristics in their choice of words. The text that will have these two styles transferred *onto* it will be the speeches of President Obama. These texts have the advantage of being in the same topic space, of the same general format and, like the speeches of Trump and Lincoln, are readily available.

Of particular challenge in language style transfer is the evaluation of transfer effectiveness. Even though language translation can be subjective, there are established evaluation metrics (e.g., BLEU4) to assess translation success. However, language style transfer is highly subjective, and while human evaluation is certainly an option (e.g., via Amazon Mechanical Turk), we will explore metrics already proposed in this area.4,5

**Datasets**

Primary sources for speech text from Presidents Trump and Obama will come from The American Presidency Project @ UC Santa Barbara (<https://goo.gl/rCJ8Gv>). This source is not only easy to parse, but is also comprehensive and kept up-to-date which will be important to ensure we capture as many speeches from President Trump as possible before project completion.

Primary sources of speech text for President Lincoln will come from one or more of the following sources:

* Abraham Lincoln Online (<https://goo.gl/J8AR6t>)
* The Abraham Lincoln Association @ U of Michigan (<https://goo.gl/SNAi0s>)
* *The American Presidency Project @ UC Santa Barbara:* (<http://www.presidency.ucsb.edu/>)

**Implementation**

Based on what we have learned to date, we will utilize an architecture including LSTM RNN’s and Seq2Seq. We will be running two instances of our system and, for each, we will train on a given speaker style (Trump or Lincoln), feed text to transfer a style “onto” (Obama), assess the results, tune and adjust, and repeat. Evaluation of the results for tuning and final assessment may include both human (e.g., Turk) and computed evaluations (e.g., BLEU, PINC).

**References**

1. Gatys, Leon A., Alexander S. Ecker, and Matthias Bethge. “A Neural Algorithm of Artistic Style,” August 26, 2015. http://arxiv.org/abs/1508.06576.
2. Jang, Se Won, Jesik Min, and Mark Kwon. “Writing Style Conversion Using Neural Machine Translation.” Accessed June 13, 2017. https://web.stanford.edu/class/cs224n/reports/2757511.pdf.
3. Kabbara, Jad, and Jackie C.K. Cheung. “Stylistic Transfer in Natural Language Generation Systems Using Recurrent Neural Networks.” In *Workshop on Uphill Battles in Language Processing: Scaling Early Achievements to Robust Methods*. Association for Computational Linguistics. Accessed June 13, 2017. https://aclweb.org/anthology/W/W16/W16-6010.pdf.
4. Papineni, Kishore, Salim Roukos, Todd Ward, and Wei-Jeng Zhu. “BLEU: A Method for Automatic Evaluation of Machine Translation.” In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, 311–18, 2002. http://www.aclweb.org/anthology/P02-1040.pdf.
5. Xu, Wei, Alan Ritter, William B. Dolan, Ralpa Grishman, and Colin Cherry. “Paraphrasing for Style.” In *Proceedings of COLING 2012: Technical Papers*, 2899–2914. Mumbai, India, 2012. https://aclweb.org/anthology/C/C12/C12-1177.pdf.