Using LSTMs to predict code-switching

Group Members: Adam Goodkind, Nicole Mirea

Our project aims to use neural language models to better understand code-switching in speech production. Code-switching is a phenomenon in multilingual language production where a speaker switches between multiple languages within the same conversation, using terms or phrases from a secondary language while producing the bulk of language in a primary language. These deviations from the primary language are thought to be systematic on a syntactic or phonological level (Sankoff, 1998; Bhatt & Bolonyai, 2011), and thus provide the possibility of being predicted via computational modeling—specifically, via a recurrent neural network architecture that uses the previous model state in order to predict the next word.

However, previous attempts to do so have either used only simple recurrent neural networks (Adel, Vu, & Schultz, 2013); to our knowledge, long short-term memory (LSTM) units have not been applied to the problem of predicting upcoming code-switches, despite the considerable improvements that LSTM networks achieve in other areas of language modeling. There are also a few unresolved questions about the cognitive architecture that underlies code-switching, as well as the motivation behind it, that we believe can be answered through this project. Namely:

- **Does code-switching involve many subprocesses?** If the neural network architecture requires many hidden layers, the answer might be yes.
- **Is code-switching conditional on long-distance dependencies?** If using LSTMs to predict upcoming code-switches yields an improvement over existing methods using simple RNNs, it would seem so, since the main advantage of LSTMs is that they are capable of encoding long-distance dependencies.

The two papers from class that are most relevant to our project are:

1. **Merity, Keskar, & Socher, 2017: Regularizing and Optimizing LSTM Language Models**
   Our project will be an LSTM language model. This paper is useful because it describes state-of-the-art methods for regularizing and optimizing such models.
2. **Luong, Pham, & Manning, 2015: Effective Approaches to Attention-based Neural Machine Translation**
   It may be the case that code-switching is essentially translation, on a smaller scale—translation from a primary language into an idiolect, with the to-be-translated parts affected by some sort of attention mask. This paper describes an implementation of attentional mechanisms which may be useful, if we conceive of the project in this way.

Since code-switching is thought to be systematic, we believe that code-switching instances can be reasonably modeled. Several freely-available software packages exist for LSTM implementation, including a PyTorch implementation ("Sequence Models and Long-Short Term Memory Networks," 2017). Moreover, corpora of spontaneous code-switching dialogues exist (Lyu, Tan, Chng, & Li, 2015), which provides an opportunity to evaluate our model and test our hypotheses.

Our project will be deemed a success if we are able to more accurately predict when a speaker deviates from her primary language and uses words or phrases from another language. We would like to predict when this occurs within speech and then predict what the alternative word choices will be. The ability to answer the theoretical questions above is another condition for success.

Bibliography

Adel, H., Vu, N. T., & Schultz, T. (2013). Combination of Recurrent Neural Networks and Factored Language Models for Code-Switching Language Modeling. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics* (pp. 206–211). Sofia, Bulgaria: Association for Computational Linguistics.

Ammar, W., Mulcaire, G., Tsvetkov, Y., Lample, G., Dyer, C., & Smith, N. A. (2016). Massively Multilingual Word Embeddings. *ArXiv:1602.01925 [Cs]*. Retrieved from http://arxiv.org/abs/1602.01925

Bhatt, R. M., & Bolonyai, A. (2011). Code-switching and the optimal grammar of bilingual language use. *Bilingualism: Language and Cognition*, *14*(04), 522–546. https://doi.org/10.1017/S1366728910000295

Luong, M.-T., Pham, H., & Manning, C. D. (2015). Effective Approaches to Attention-based Neural Machine Translation. *ArXiv:1508.04025 [Cs]*. Retrieved from http://arxiv.org/abs/1508.04025

Lyu, D.-C., Tan, T.-P., Chng, E.-S., & Li, H. (2015). Mandarin–English code-switching speech corpus in South-East Asia: SEAME. *Language Resources and Evaluation*, *49*(3), 581–600. https://doi.org/10.1007/s10579-015-9303-x

Merity, S., Keskar, N. S., & Socher, R. (2017). Regularizing and Optimizing LSTM Language Models. Retrieved from https://arxiv.org/abs/1708.02182

Sankoff, D. (1998). A formal production-based explanation of the facts of code-switching. *Bilingualism: Language and Cognition*, *1*(1), 39–50. https://doi.org/10.1017/S136672899800011X

Sequence Models and Long-Short Term Memory Networks. (2017). Retrieved February 6, 2018, from http://pytorch.org/tutorials/beginner/nlp/sequence_models_tutorial.html