

Thesis Project Portfolio

Teaching Endangered World Languages with a Conversational Artificial Intelligence

Application

(Technical Report)

Helpful or Hurtful? Analyzing the Internet's Role in Minority and Indigenous Language

Actor-Networks

(STS Research Paper)

An Undergraduate Thesis

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Charles Edward West Beall

Spring, 2024

Department of Computer Science

Table of Contents

Sociotechnical Synthesis

Teaching Endangered World Languages with a Conversational Artificial Intelligence Application

Helpful or Hurtful? Analyzing the Internet's Role in Minority and Indigenous Language

Actor-Networks

Prospectus

Sociotechnical Synthesis

Endangered languages around the world risk losing all of their speakers by the end of the century, meaning that many important histories and cultural traditions will be forever lost with them. Communities and countries with speakers of these languages face many challenges in preventing the seemingly inevitable extinction of their language to preserve a major part of their identity. These challenges manifest themselves in the form of lack of language resources, unfavorable social pressures, or direct opposition to their use. My technical project proposes the creation of an online language learning platform for endangered languages, and my STS research focuses on analyzing actor-networks of endangered languages with technology such as the internet and software platforms as actors. If fully realized, the proposed language learning platform would be an actor in the explored networks as a software product that language learners and speakers interact with.

My technical project proposes a tool for communities to teach languages to younger generations where sufficient personnel and resources for the task are lacking, with the hope of aiding larger efforts to revitalize endangered languages. More specifically, my technical project is the design of a language learning platform with a chatbot that users can speak to, listen to, and converse with over text in order to learn endangered languages. The platform's chatbot would respond to users' requests to create personalized lessons and exercises. It would be constructed using a large language model that is trained with natural language processing (NLP) techniques for the use of endangered languages. In order to come up with its concept and design, I researched the use of artificial intelligence in language learning, and drew inspiration from existing language learning platforms that host endangered languages such as Duolingo, as well as large language models such as ChatGPT. I also researched methods for training NLP models on

low amounts of data, since endangered languages have fewer resources from which training data can be drawn.

My STS research, on the other hand, takes an in-depth look at various case studies of minority and Indigenous languages, how their speakers interact with technology for communication, and how those actors are affected by government policy. The specific cases that I analyzed were of the Uyghur language in China, the Irish language (Gaelic) in Ireland, the Maori language in New Zealand, and Mayan languages in Central America. Through my research, I found that language networks can be characterized in three main categories. The first type, networks with oppressive governments, is represented by the Uyghur language, because the Chinese government uses internet censorship to limit communication in the Uyghur language. The second type, networks with supportive governments, can be represented by the Irish and Maori languages, because the governments of their respective countries have created policies conducive to the use of those languages on the internet. The third type, networks with limited government impact, where language users must support their own efforts, is embodied by Mayan languages in relation to the Guatemalan and Mexican governments. In all of these cases, when unimpeded, the internet serves as a place in which language users can communicate across vast distances to have cultural conversations and speak in their language.

My technical project and research are far from solving the problems that endangered languages face, but they contribute valuable ideas to this goal. A future team of engineers and researchers should work to build the platform that I have outlined in my technical project, as it is too big an endeavor to take on alone. It will also be important for researchers to gather as much language data as possible from resources in endangered languages for a NLP model to be trained in their use. On the social side, researchers should evaluate the efficacy of policies aimed at

revitalizing endangered languages, and recommend such policies to governments and other international bodies, and figure out how to incentivize malicious governments to instead act in favor of endangered languages, as that is perhaps the largest threat to their future.

I would like to thank all those that inspired and assisted my work. First, I would like to thank my STS professors, Travis Elliott and Dr. Caitlin Wylie, for their guidance and feedback in piecing together my ideas and putting them into writing. I would also like to thank my mom, Alexandra, for inspiring me to take interest in other cultures and learn other languages, and my dad, Jeffrey, for inspiring me to travel and learn about the world around me.