One of my goals to learn computer science is to implement a translator that can translate PutongHua to ChaoshanHua. In the project of my C++ course, when I was a freshman, I implemented a prototype that translates PutongHua to characters by [Xunfei]() API and then converted the characters to ChaoshanHua according to the recorded audio of each word (Lol, how naïve it was). It's far from satisfactory because our program doesn't consider the semantics of the language! Due to its complexity, I postponed this project and continued my studies in computer science. Thanks to the deep learning I learned under the supervision of my advisor Professor Si Wu, I am reconsidering this project with serious commitment!

I am from the Chaoshan region, Guangdong. There is a dialect called ChaoshanHua widely spoken in the Chaoshan region. ChaoshanHua is significantly different from Putonghua (i.e., a widely-accepted, standard language spoken in China) due to its isolation geologically and cultural traditions. For example, PutongHua is believed as one of the most challenging languages because it's based on hieroglyphs, and it contains 21 consonants, 13 vowels and four tones. While ChaoshanHua has no characters, it contains even 18 consonants, 64 vowels, and eight tones! It's insanely difficult for non-local residents to learn how to speak ChaoshanHua. Indeed, there are dialects everywhere around the world. However, ChaoshanHua, with more than 30 million users, brought negligible negative impacts to the Chaoshan Region, economically and culturally. Honestly, the fundamental reasons for its complexity are as follows:

- Geologically, the Chaoshan region is located in the mountainous area where communities and villages are separated by mountains. In ancient times, people hardly communicated with outsiders due to geological inconvenience. It's discovered that many ancient pronunciations that date back to Qin Dynasty (200 B.C.) are found in ChaoshanHua!

- Culturally: The geological isolation exacerbates the cultural isolation of the Chaoshan region. People married, commercialized with local residences, which endows the anti-diversity social attitude. The undeveloped situation of the Chaoshan region is related to the anti-diversity attitude. When I was young, adults believed it was a shame to marry a non-Chaoshanese.

- Ideologically: It was pretty feudal in the Chaoshan region 30-40 years ago; For example, women are incredibly less literate than men, and family culture is widely accepted. It was believed that a qualified woman should take good care of the family.

By reading so far, you may understand my feelings about my hometown. Even though we have been developing in recent fifteen years due to the development of China, many people, especially middle-aged women, are unable to recognize the PutongHua. They are unwelcomed to the new era. As a typical woman in the Chaoshan region, my mom took good care of her four children, and we are all college students now. But she remains illiterate as many women out there do.

One of my goals to learn computer science is to implement a translator that can translate PutongHua to ChaoshanHua. In the project of my C++ course, when I was a freshman, I implemented a prototype that translates PutongHua to characters by [Xunfei]() API and then converted the characters to ChaoshanHua according to the recorded audio of each word (Lol, how naïve it was). It's far from satisfactory because our program doesn't consider the semantics of the language! Due to its complexity, I postponed this project and continued my studies in computer science. Thanks to the deep learning I learned under the supervision of my advisor Professor Si Wu, I am reconsidering this project with serious commitment!

Therefore, this blog will show you how this project is working on!