This project focuses on the challenge of developing Automatic Speech Recognition (ASR) systems for code-switching (CS) speech, where two or more languages are used in a single sentence. The main difficulty is the lack of available textual CS data for training satisfactory language models (LMs) for ASR systems. To address this, the authors propose a system for CS text generation using monolingual and parallel data and automatic selection of natural CS examples using linguistic theories. They train an encoder-decoder transformer network to translate between L1 and L2 languages and determine the main language by comparing scores from monolingual LMs. They analyze the naturalness of CS utterances using cross-attentions and Part-of-Speech and dependency analysis. The resulting method can filter out up to 50% of bad CS examples and improve synthetic ASR LM CS corpus. The accuracy of the system is 59% on real CS data, 65% on generated CS data from training on CS data, and 70% on generated CS data from parallel data.