

CS224N Assignment 3: NMT Systems Analysis

January 16, 2026

2. Analyzing NMT Systems (29 points)

(a) (3 points)

When encoding a Mandarin Chinese sequence, the tokenizer maps the sequence into vocabulary items, each consisting of one or more characters. In Mandarin, each character is often a word or a morpheme. For example, (electricity), (brain), and (computer).

Adding a 1D Convolutional layer after the embedding layer allows the model to capture local patterns and n-gram features in the sequence of embeddings. This is especially useful for Chinese, where the meaning of a word can depend on the combination of adjacent characters. For instance, while and have their own meanings, together as they form a new word (computer). The convolutional layer can learn to recognize such meaningful combinations, improving the model's ability to represent and translate phrases that are not just the sum of their parts. This helps the NMT system better capture the compositional structure of Mandarin words and phrases, leading to improved translation quality.