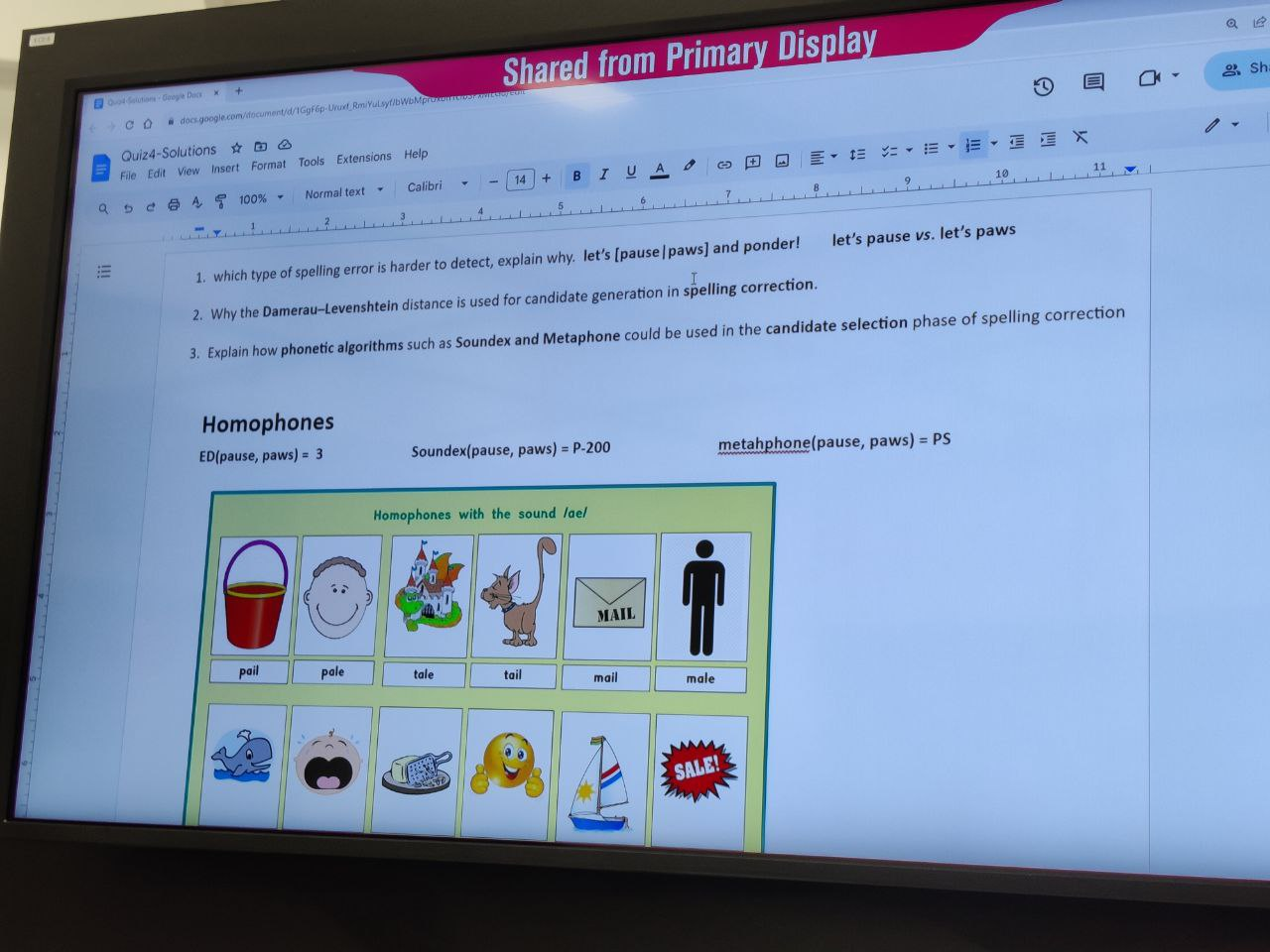
N-grams, Edit distance and Spelling corrections



1. Which type of spelling errors are harder to detect? Explain why.

It is the mistakes that are also words in a dictionary that are harder to detect compared to detecting spelling mistakes that are non-words.

2. Why is the **Damerau-Levenshtein distance** is used for candidate generation in spelling correction

Damerau-Levenshtein distance is used for candidate generation in spelling correction because it considers transposition as well which is a likely error to occur when the text is input from a keyboard.

3. Explain how **phonetic algorithms** such as **Soundex and Metaphone** could be used in the candidate selection phase of spelling correction.

Phonetics can be a second layer for candidate generation when there are instances where the edit distance may be too high and might not produce the correct candidates. For example, the edit distance between the words “pause” and “paws” is 3 and hence might be missed if we were to use only edit distance calculations for our spelling corrections.