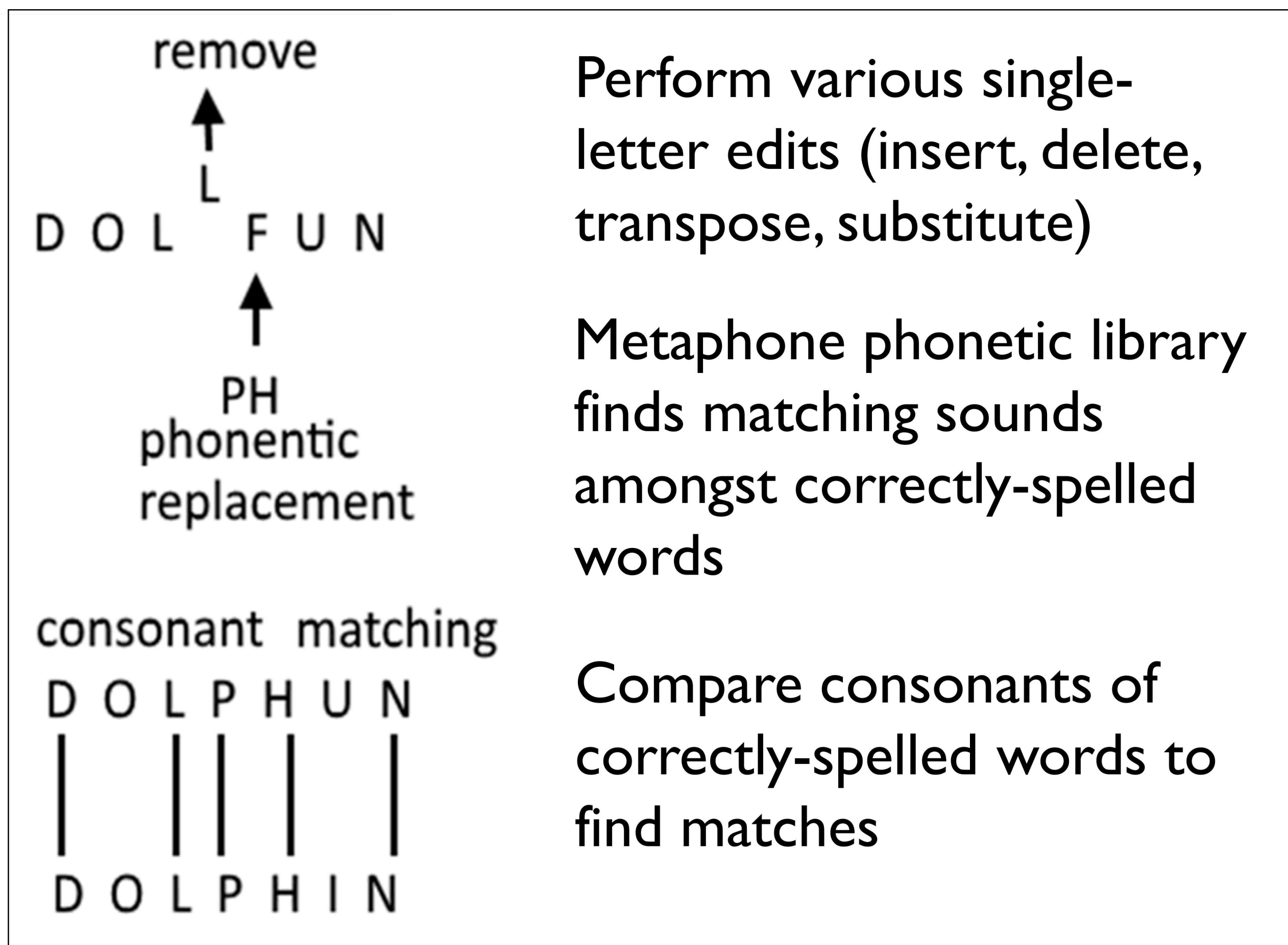
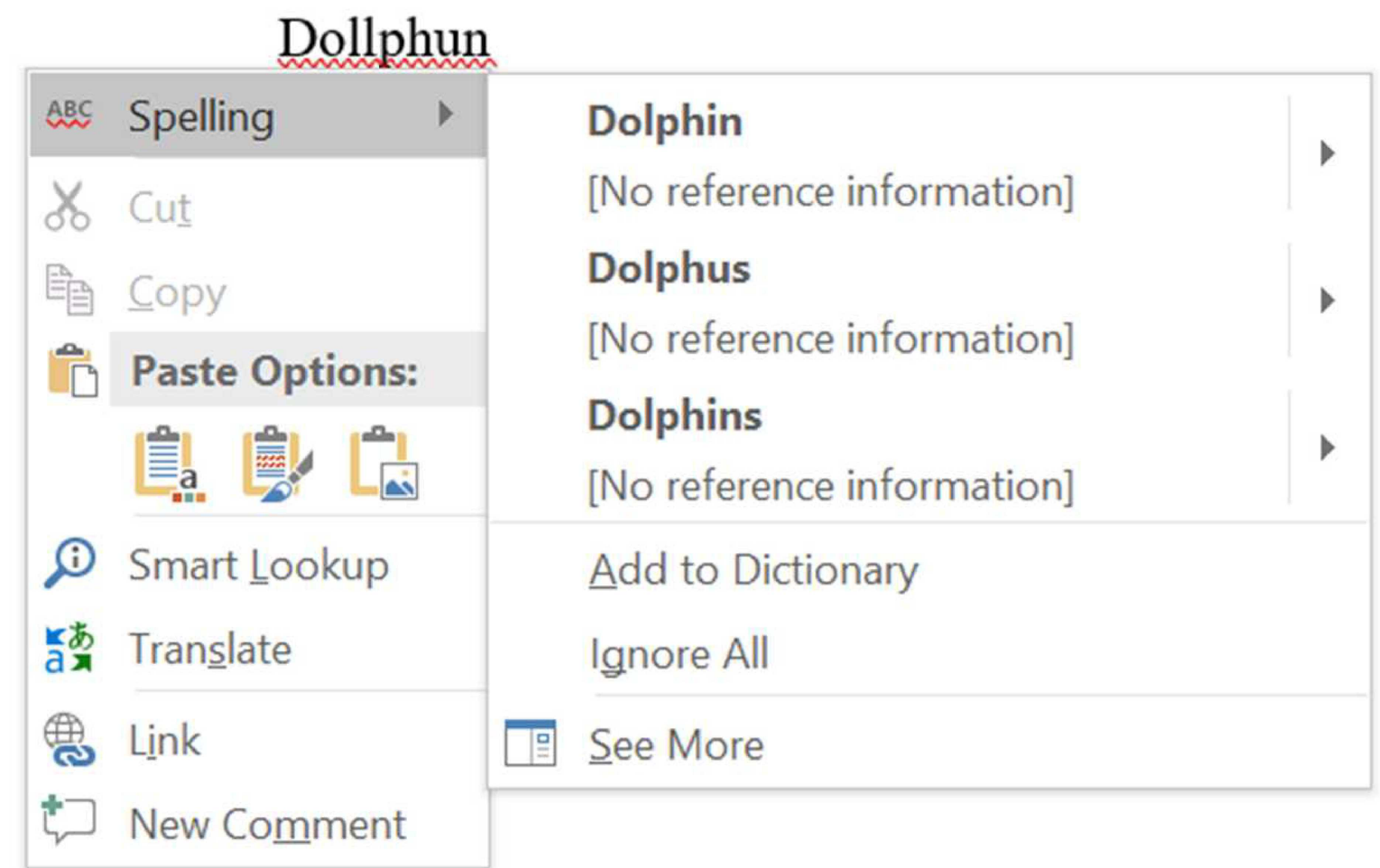




# Phonetic Matching Effect on Natural Language Processing

Lokraj Srinivasan | University of Michigan 2021 | Data Science Edward Walsh | Org 8754

**Abstract:** To create an improved out-of-context spell checking algorithm by adding phonetic matching to the Damerau-Levenshtein minimum edit distance algorithm. Many people sound-out the word to get the correct spelling. Typically sources of error arise from the user's misplacement, omission, or transposition of vowels.



**Algorithm:** The Damerau-Levenshtein minimum distance generates several strings of potential single-character edits (insertions, deletions, transpositions, and substitutions). The Metaphone library then identifies the strings that sound similar to correctly spelled words. Consonants are matched because of the user's tendency to omit, misplace, or add in vowels. The remaining strings are compared to generate suggestions for corrections.

## Kinds of Errors:

- Non-word: "Apple" → Aple"
- Real-word: "Through" → "Thorough"
- Phonetic: "Surprise" → "Surprice"

## Levenshtein Minimum Edit Distance Matrix

Tested with a sample of 187 words

		D	O	L	P	H	I	N
	0	1	2	3	4	5	6	7
D	1	0	1	2	3	4	5	6
O	2	1	0	1	2	3	4	5
L	3	2	1	0	1	2	3	4
F	4	3	2	1	1	2	3	4
U	5	4	3	2	2	2	3	4
N	6	5	4	3	3	3	3	3

