nlp_2: Task 22 – Compulsory Task 2

Read up on any innovative technology using NLP (by companies such as Google or IBM, for instance) and write a brief summary about the technology, what it achieves/does, and an overview of how it works (250 -500 words).

I have chosen to write about Grammarly.

The below information has been summarised and interpreted from the following sources:

https://aclanthology.org/W18-6105.pdf
https://www.grammarly.com
https://www.grammarly.com/blog/how-grammarly-uses-ai
https://www.grammarly.com/blog/how-does-grammarly-work/#:~:text=When%20you%20write%20with%20Grammarly,offering%20a%20clearer%20sentence%20structure.
https://www.grammarly.com/blog/nlp-run-on-sentences/
https://www.inbenta.com/en/blog/nlp-use-cases/

Grammarly is a grammar and language correction tool that uses NLP to scan a text, check for language errors and gives suggestions on corrections. According to its website, "Grammarly can help you with everything from fixing basic grammatical mistakes to avoiding overused words, keeping your writing concise, and sounding more confident".

Grammarly's website states that it is "trained on naturally written text" so is "good at spotting issues that occur naturally when people write". It is trained on text copera (large collections of sentences) that its researchers organise and label in a way that its AI algorithms can understand. Grammarly's algorithms learn the rules and patterns of "good writing" by analysing the millions of sentences from the corpora. Grammarly has learned 'correct' and 'incorrect' spelling, grammar and punctuation rules that can it can apply to user inputted text.

Grammarly's website gives commas as an example. The AI model is shown sentences with incorrect commas, so it can learn what a comma mistake looks like. It is then shown sentences with good comma usage, so it can learn how to fix comma mistakes when it finds them.

Grammarly also learns from users' input. When a user accepts or ignores a suggestion, it learns to differentiate between correct and incorrect usage. When lots of users hit "ignore" on a particular suggestion, Grammarly's computational linguists and researchers make adjustments to the algorithms to make it more accurate. This suggests it uses a supervised or semi-supervised machine model.

Another of Grammarly's features is a 'tone detector', where it can detect the tone of an email or other text by analysing a user's word choice, phrasing, punctuation and capitalisation. This could be considered as an example of sentiment analysis.

Grammarly has recently added the ability to detect 'run-on sentences' ("two or more complete sentences that have been improperly squashed together") and offers suitable alternatives where these are detected (e.g. splitting the sentence into two separate sentences, or separating it into two parts by use of a semi-colon or comma).

The run-on sentences feature was developed by being fed millions of example sentences adapted from a corpus of news texts. It uses two machine learning models: a 'Conditional Random Fields' model (roCRF), and a 'Sequence to Sequence model' (roS2S). roCRF "represents a sentence as a sequence of spaces between tokens, labeled to indicate whether a period should be inserted in that space". Each space is represented by contextual features (sequences of tokens, part-of-speech tags, and capitalization flags around each space). roS2S "determines if there is terminal punctuation" (e.g. a period/full stop) following each token *(https://aclanthology.org/W18-6105.pdf)*.