

Grammarly uses innovative NLP technology to not only correct grammar and spelling but also enhance the overall quality of written content by providing suggestions for clarity, tone and style. For example, while composing an email or a document it could be hard to articulate thoughts clearly and important ideas might be buried in text or the paragraph structure. Grammarly helps people communicate more effectively by providing efficient suggestions for improvements of email or any texts in documents.

To achieve this Grammarly breaks the the technical problem into two key components:

1. Extracting the main points from the text
2. Detecting where the reader is likely to focus their attention

The first point is a familiar problem in ML and NLP but grammarly uses a different approach for this problem which involves separating action items from main points. For example, an action item could be "Please complete this task by end of day" and a main point could be "Project has been delayed by 3 weeks". This separation of creating separate buckets for action and main items improves accuracy and enables future ML-powered features around action items.

The second component involves assigning an importance score to main points in a message. This score is determined using annotated data, where scores are averaged or normalized by different people to achieve a general consensus. Grammarly adopted this framework to collect truth data with annotators identifying main points and action item sentences, along with judging the importance of each main point in a set of representative email texts.

The next step was developing the model.

1. **Extractive summarization:** Grammarly initially uses LexRank and TextRank, based on the PageRank algorithm, to extract main points from the text. While PageRank measures the number of links between pages, LexRank and TextRank consider the number of words that a sentence has in common with other sentences in the text. Though a deeper learning algorithm like BERTSUM the more modern approach to text summarisation is highly effective, it requires a large dataset, which Grammarly initially lacked.

2. **Developing Features:** In-house analytical linguists create a robust set of language-based features. For smaller texts, main points are typically found at the beginning or end, and in larger documents, the title or subject indicates this. These approaches are used to identify features.

3. **Optimization:** The model initially faced challenges with longer texts, leading to various optimization strategies. Instead of working with all sentences, a sampling approach is used, selecting random sentences scored against features. Subsequent sampling biases towards top-scoring sentences, resulting in a best subset. Additional strategies involve dropping non-essential sentences like salutations, improving response times, and reducing latency.

Overall, Grammarly achieved a model that performed with 80% precision, and with the development of robust annotated datasets, were recently able to move to a deep learning approach that preserved the same level of precision while improving recall from 40% up to 62%.

**Reference:** <https://www.grammarly.com/blog/engineering/nlp-ml-identify-main-points/>