# Language nerds – Comma placement in sentences


## Group members:
S082971 – Jens Peter Träff
S151714 – Marcus Pagh
S153528 – Poul Gunnar Pii Svane
S154438 – Theis Kierkegaard Hauge


## Motivation:
Comma placement is notoriously difficult, especially in the danish language. For publishers (and potentially others) it would be a significant help to have an automatic way to check that the commas are placed correctly, as well as placing them automically.

## Background:
In danish there are multiple rules for the correct comma placement, and quite a few exceptions to the general rules. That makes the problem very hard for any automatic hardcoded solver. The problem however seems to fall naturally into a category, that can be solved by machine learning methods, particularly Recurrent-neural networks.
The problem can be stated as a classification problem (is the comma placed at the right position in the sentence?). We expect to utilize this idea and model it using a recurrent neural network, as this will allow the model to learn from the temporal dimensions of the sentence.

## Milestones:

| | |
|---|---|
| 30. oktober: | Data is ready and pruned so it can be used. Encoding is chosen |
| 6. november: | First model is ready to be trained and data is encoded. |
| 13. november | Evaluate on first model – figure out what changes needs be made |
| 20. november | Second and improved model is ready to be trained |
| 4. december: | Primary development ends. (training could still continue) |
| 18. december: | project presentation. |
| 22. december: | Final report is handed in. |

## References:
At the time of writing we plan on using the RNN specified in the lab3 exercises. We then plan to adjust the model to our needs during the project.


https://www.cs.toronto.edu/~graves/preprint.pdf

http://www.deeplearningbook.org/contents/rnn.html