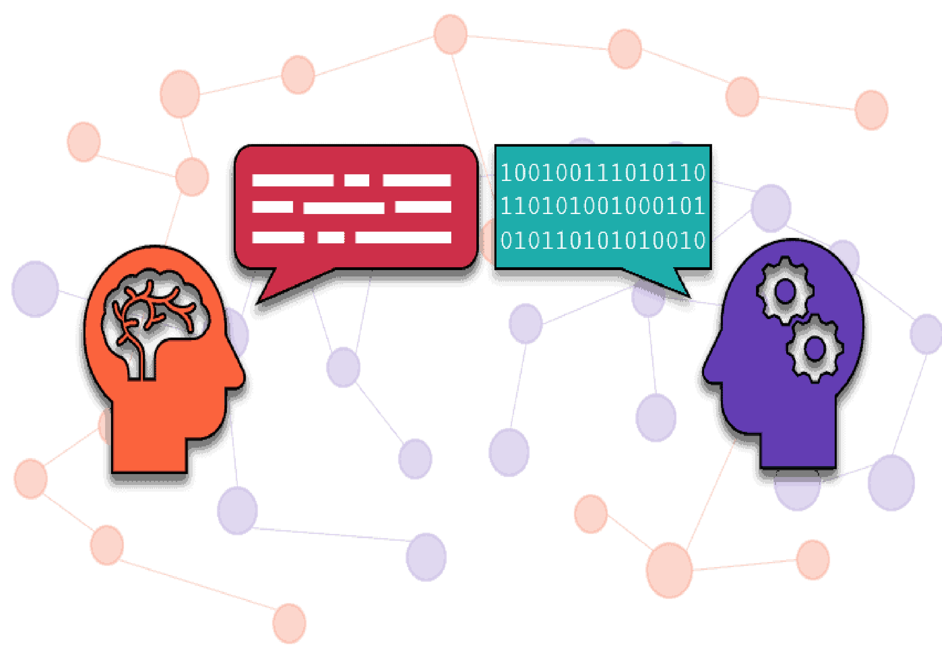
Meaning Representation Parsing



Meaning Representation Parsing is a crucial task within Natural Language Processing. Our main task was to investigate Semantic Graph Parsing, which aims to create graphical meaning representations for natural language that can be interpreted by computers. We present research on two semantic parsers: Sequence-To-Sequence-Based and Graph-Based, using the LinGo Redwoods corpus as our data source and Elementary Dependency Structures as our parsing framework. Our objectives were to compare pre-trained transformer neural networks to LSTMs, using metrics such as precision, recall, F1 score and SMATCH (for whole sentence semantic structures). Models were trained using the Centre for High Performance Computing cluster.

Sequence-To-Sequence-Based

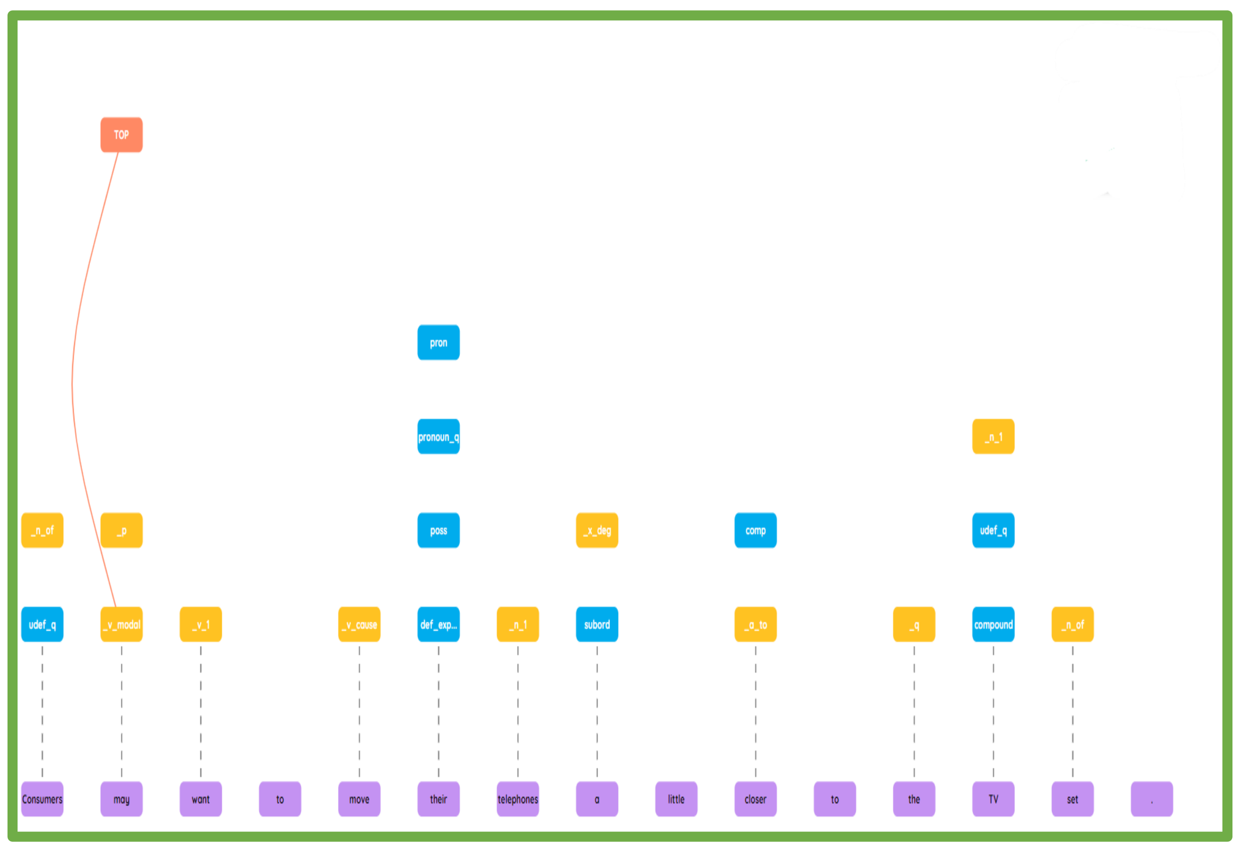
By Chase Ting Chong

Seq2seq approaches serialize meaning representation graphs using graph linearization techniques. Seq2seq neural networks can then be trained to directly predict the serialization given an input sentence. We investigated the use of the BART transformer with a modified PENMAN serialization and compare this to previous RNN based approaches; as well as the impact of pretraining the BART model. Our results show an improvement of 5.64 F1 score over a previous RNN approach, and the use of pre-training significantly improves F1 score by 21.11.

"The results were in line with analysts' expectations."

```
(< 3 3 > _p :ARG1 (< 1 1 > _n_of :BV-of (< 0 0 > _q))
            :ARG2 (< 4 4 > _n_of :BV-of (< 4 4 > idiom_q_i))
            :ARG1-of (< 5 5 > _p  :ARG2 (< 8 8 > _n_of :BV-of (< 7 7 > def_explicit_q)

        :ARG1-of (< 7 7 > poss :ARG2 (< 6 6 > _n_1 :BV-of (< 6 6 > udef_q))))))
```



Graph shows nodes generated by BERT model. Yellow is the surface nodes and blue is the abstract.

Graph-Based: Node Prediction

By Jane Imrie

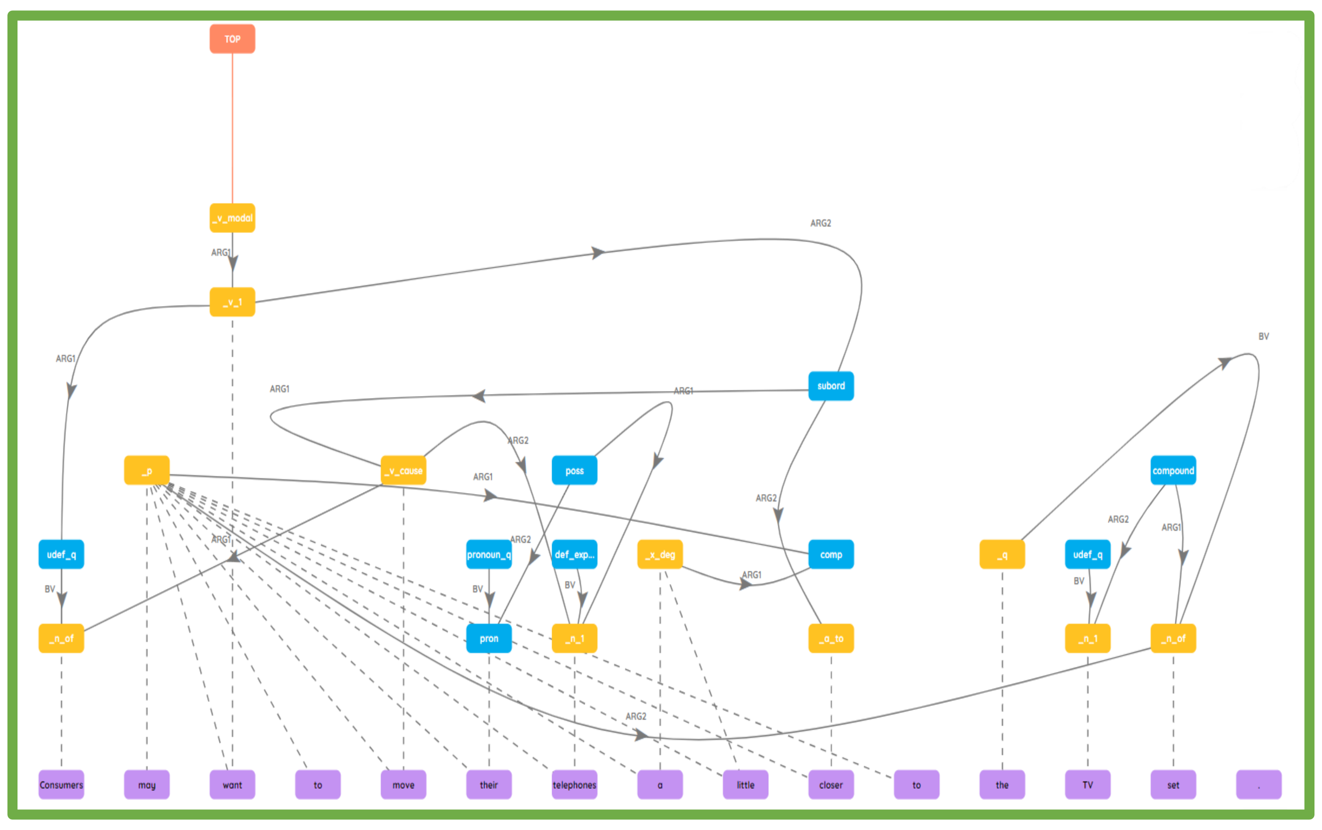
Node prediction/concept identification is the first stage in the graph-based parsing process. It consists of finding concepts which are triggered by tokens within a sentence. A sentence can have surface concepts, which can be derived from a token's orthography, or abstract, which show the influence of grammatical constructions. Nodes can correspond to multiple tokens, sub-tokens or singular tokens. The accuracy of a pre-trained and fine-tunable SpanBERT transformer and Glove embeddings with a BiLSTM were compared, using the precision, recall and F1 scores, with the transformer substantially outperforming the LSTM across all categories.

Graph-Based: Edge Prediction

By Claudia Greenberg

Edge Prediction involves predicting the relationships between the nodes of a graph to amplify the meaning representation of a sentence. The two parts of the prediction are: which nodes relate; and which types of relationships exist. A comparative study was conducted comparing: Pre-trained BERT versus Non-trained LSTM; and Maximum Entropy versus Maximum Margin Loss Functions. We present high accuracy results for the BERT module. However, limited accuracy results were presented for the LSTM module, resulting in inconclusive answers to our research questions.

This graph presents the same sentence as shown in the above graph, with the inclusion of the edges connecting the nodes.



Team:
Claudia Greenberg | GRNCLA009
Jane Imrie | IMRJAN001
Chase Ting Chong | TNGCHA001

University of Cape Town
Department of Computer Science
dept@cs.uct.ac.za
Supervisor:
Dr Jan Buys | jan.buys@uct.ac.za