What Should/Do/Can LSTMs Learn When Parsing Auxiliary Verb Constructions?

Miryam de Lhoneux

Sara Stymne

Joakim Nivre

Computational Linguistics, Volume 46, Number 4. 2020

Dissociated nucleus & transfer relations

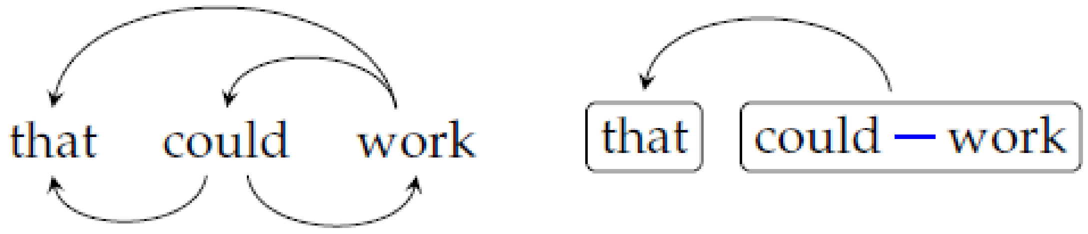

Figure 1
Two different representations of a sentence with auxiliary as used in dependency parsing (left) vs. as can be represented following following Tesnière (1959) (right).

UD vs. MS

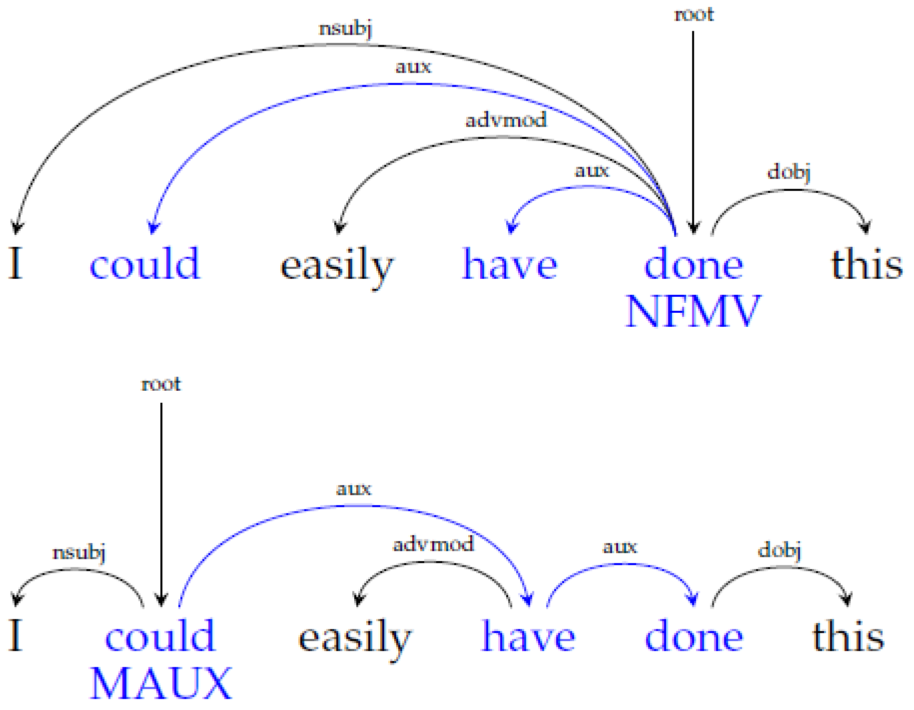

Figure 3
Example sentence with an AVC annotated in UD (top), and in MS (bottom). AVC subtree in thick blue.

Research questions

- RQ1: Is information about agreement and transitivity learned by the parser?
- RQ2: Does a sequential NN-based dependency parser learn the notion of dissociated nucleus?
- RQ3: Does a dependency parser augmented with a recursive layer learn the notion of dissociated nucleus?

Research questions

- RQ1: Is information about agreement and transitivity learned by the parser?

We have clearly seen (1) that transitivity and agreement are learned by the parser and that some information about these tasks is available to the parser when making decisions about FMVs and (2) that this information is not available everywhere in the network and is therefore available specifically when making decisions about FMVs. This answers **RQ1** positively.

Research questions

- RQ2: Does a sequential NN-based dependency parser learn the notion of dissociated nucleus?

We can conclude that a BiLSTM-based parser does not learn the notion of dissociated nucleus for AVCs when working with a representation of AVCs where the main verb is the head such as is the case in UD: The representation of NFMVs contains less information about agreement than the representation of FMVs. However, when using a representation where the auxiliary is the head, a BiLSTM-based parser does seem to learn this notion; it learns a representation of the AVC's head that is similar to the representation of FMVs.

Research questions

- RQ3: Does a dependency parser augmented with a recursive layer learn the notion of dissociated nucleus?

We can therefore conclude that a recursive composition function on top of a BiLSTM allows the model to capture similar information about AVCs and their non-dissociated counterpart, FMVs. This indicates that composing subtree representations with a recursive layer makes it possible for the parser to learn the notion of dissociated nucleus with a representation of AVCs where the head is the main verb.

Overall, it seems that using a UD representation and a recursive composition function is the best option we have to have an accurate parser that captures the notion of dissociated nucleus, given our definition of what it means to capture this notion (that FMV and NFMV representations encode a similar amount of information about agreement and transitivity). This does not improve overall parsing accuracy, which means either that it is not important to capture this notion for parsing or that the benefits of doing so are offset by drawbacks of this method. It would be interesting to find out whether learning this information is important to downstream tasks, which we leave to future work.

Do latent tree learning models identify meaningful structure in sentences?

Adina Williams

Andrew Drozdov

Samuel R. Bowman

Transactions of the Association for Computational Linguistics, Volume 6. 2018

Questão de pesquisa

- Gramáticas induzidas automaticamente refletem princípios linguísticos?
  - Aprendizado automático de árvores com base em vários métodos da literatura, incluindo o melhor deles
  - Comparação com “árvores reais” do Penn Treebank (constituintes)

Conclusões

- Não há correspondência entre o que se aprende e qualquer noção sintática ou semântica conhecida ☹
  - Mas não quer dizer que não seja útil para aplicações específicas (no caso, inferência textual - resultados tão bons quanto os baseados em parsing correto)
  - E há alguns padrões (ruins)
    - Conexão entre duas primeiras e duas últimas palavras
    - Negação ligada à palavra seguinte
    - Inconsistente com palavras funcionais e modificadores

Exemplo

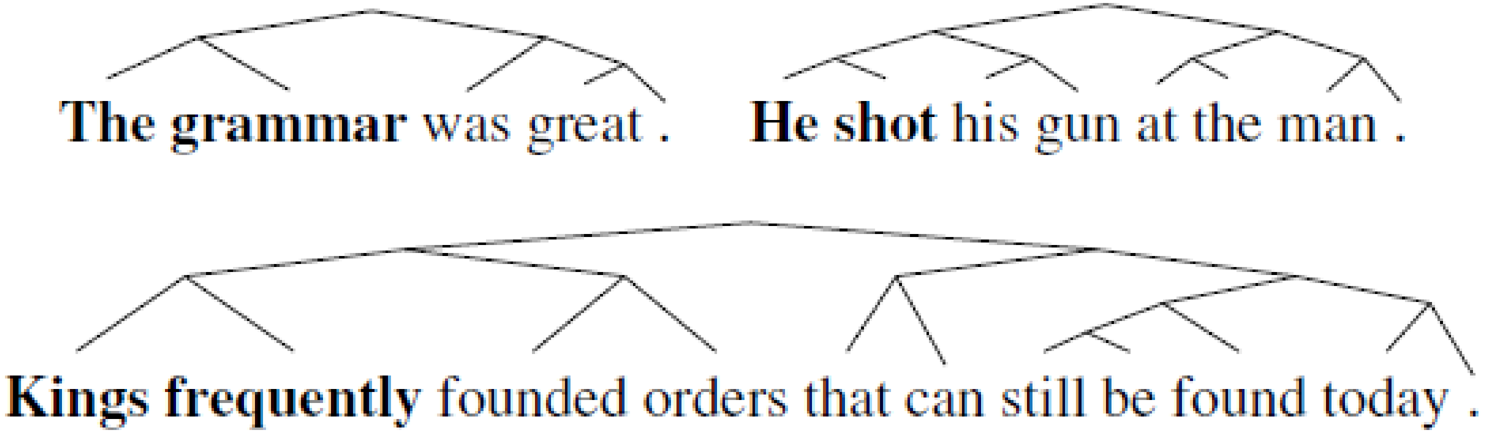

Figure 4: The ST-Gumbel models often form constituents from the first two words of sentences.

Exemplo

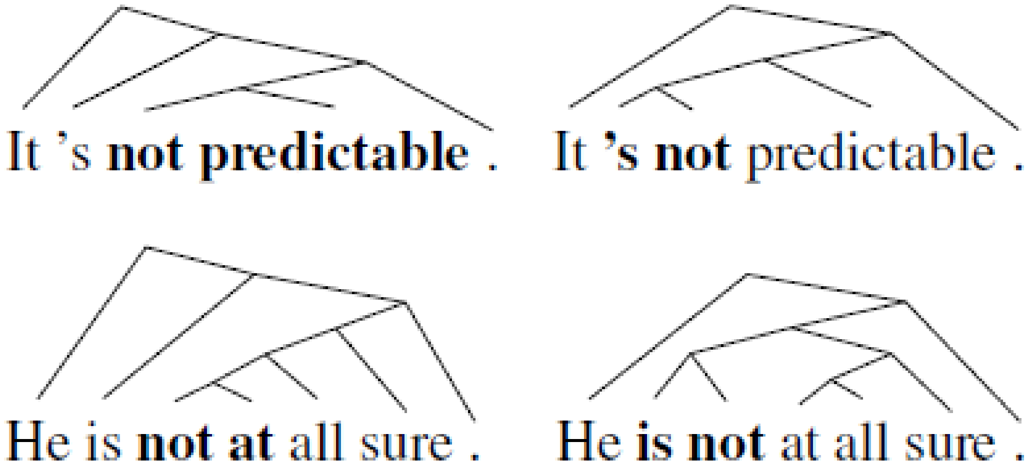

Figure 5: Left: ST-Gumbel models reliably attach negation words to the words directly following them. Right: Stanford Parser trees for comparison.

Exemplo

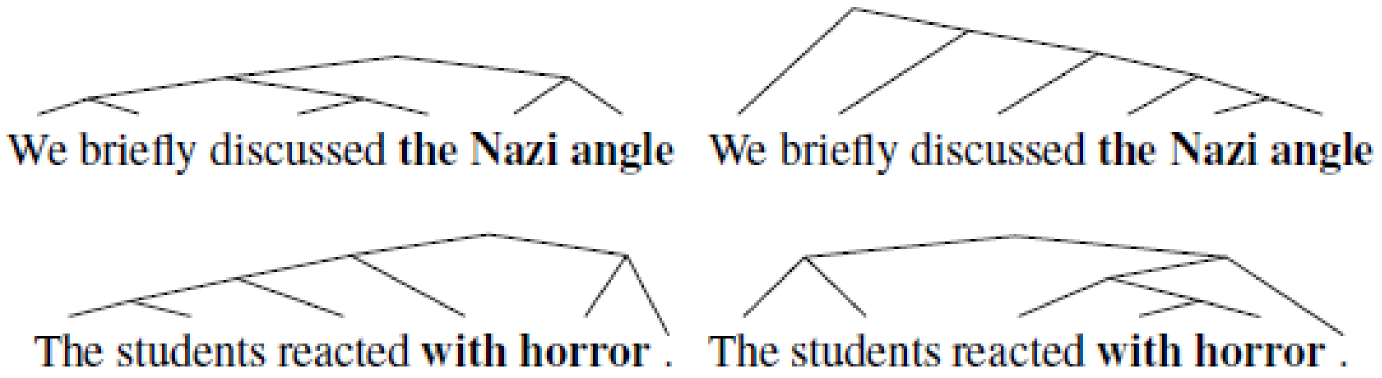

Figure 6: Left: ST-Gumbel models are inconsistent in their treatment of function words and modifiers. Right: Stanford Parser trees for comparison.