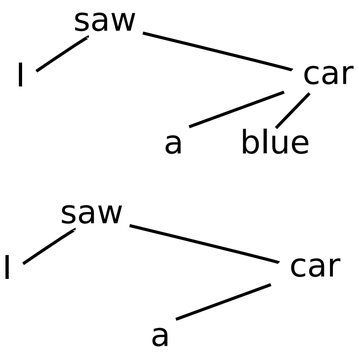
Inducing syntactic trees from BERT representations

Rudolf Rosa and David Mareček
Charles University, Institute of Formal and Applied Linguistics

Reducibility of a word in a sentence

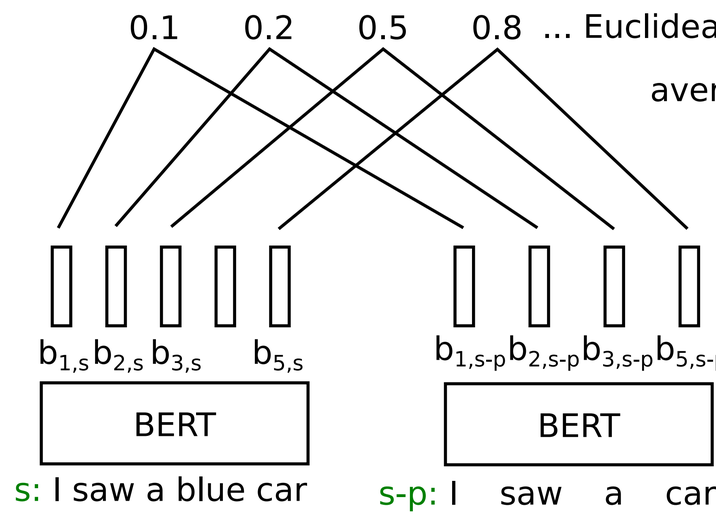
Reducibility in a syntactic dependency tree
- theoretical binary concept
- if a word is reducible, it can be removed (reduced) from the sentence without the loss of grammatical correctness
- when generalized, this constitutes one of the principles for representing sentences by dependency trees



Reducibility in a corpus
- attempt to estimate word reducibility from a corpus
- a word is declared reducible if the corpus contains a pair of identical sentences except for the presence or absence of the word
- extremely sparse
- can be used to estimate part-of-speech reducibility; e.g. adjectives are more reducible than nouns

...
I saw a blue car.
...
I saw a car.
...
...
I saw a [ADJ] car.
...
I saw a car.
...

Reducibility in BERT representations
- estimation of a reducibility score based on removability of the word
- a word is highly reducible if the average of BERT representations of the other words in the sentence does not change much when the word is removed
- can be also generalized to reducibility of phrases (remove a whole phrase)



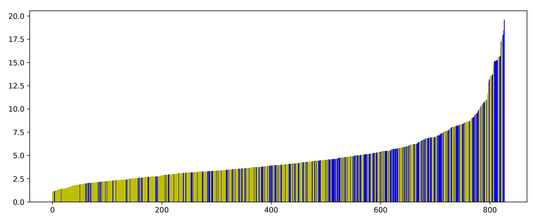
Reducibilities of all words in a sentence (smaller words are more reducible)

I SAW A BLUE CAR

It is possible to estimate **roots** and **leaves** of dependency trees from **unsupervised BERT** representations

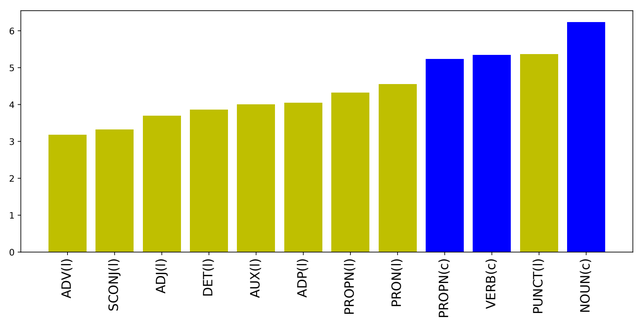
Linguistic properties of BERT-based reducibilities

Reducibility of leaf/non-leaf
- leaf nodes (yellow) tend to be more reducible than non-leaf nodes (blue)



- scores vary among sentences
- empirical threshold: 1.2x avg
- leaf identification acc. 74.5%
- baseline all-is-leaf: 65.8%

Reducibility per part-of-speech
- average reducibility per POS, leaf (yellow) versus non-leaf (blue)



Dependency edge direction
- head tends to be less reducible than dependent
- true for 70.6% of edges
- baseline right-chain: 65.8%

Low reduciblity of root
- root tends to be the least reducible word: it is so in 34% of sentences
- if we ignore punctuation: 46%
- random baseline: 13%

I RAN ACROSS THIS ITEM ON THE INTERNET .

THE US TROOPS FIRED INTO THE HOSTILE CROWD , KILLING 4 .

THE HYMN WAS SUNG AT MY FIRST INAUGURAL CHURCH SERVICE AS GOVERNOR .

THE HYMN TALKS ABOUT SERVING SOMETHING GREATER THAN YOURSELF IN LIFE .

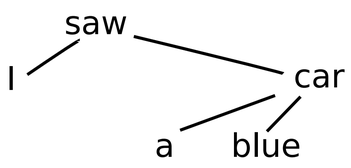
THE HOTTEST ITEM ON CHRISTMAS WISH LISTS THIS YEAR IS NUCLEAR WEAPONS .

WHEN THEIR PRECIOUS CARTOONS ARE RELEASED I HIGHLY DOUBT IT WILL LOOK LIKE THE END OF THE WORLD

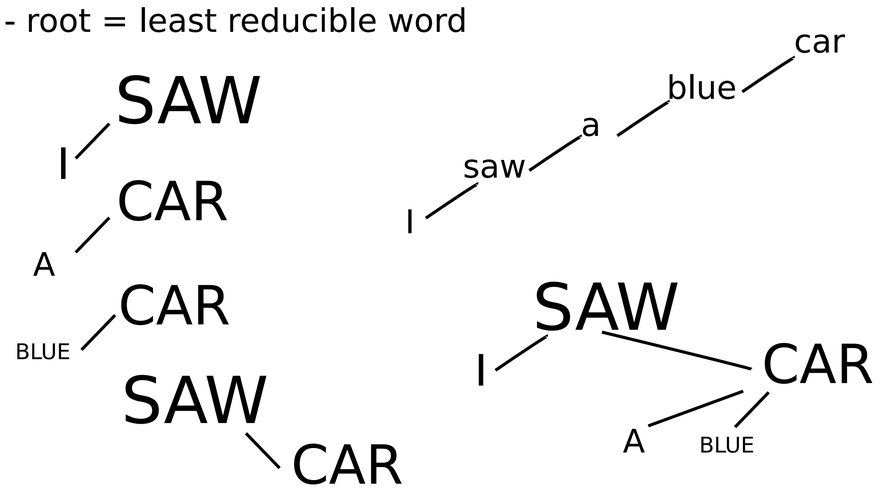
Building dependency trees using BERT-based reducibilities

Algorithm D: subtree bracketing
- compute reducibilities of words and phrases
- iterate: denote the most reducible phrase as a subtree by brackets

I saw a blue car
I saw a (blue) car
I saw (a) (blue) car
(I) saw (a) (blue) car
(I) saw ((a) (blue) car)



Algorithm R: modified right-chain
- head of a word = nearest subsequent less reducible word
- if no such word: attach to the root
- root = least reducible word



Parsing accuracy (UAS)

baseline (right chain)	29.5%
algorithm D	31.1%
algorithm R	33.1%
enforce low reducibility for punctuation	
algorithm D	37.0%
algorithm R	40.6%

This work has been supported by the grant 18-02196S of the Czech Science Foundation and uses languages resources and tools developed and stored by the LINDAT/CLARIN project (LM2015071)!