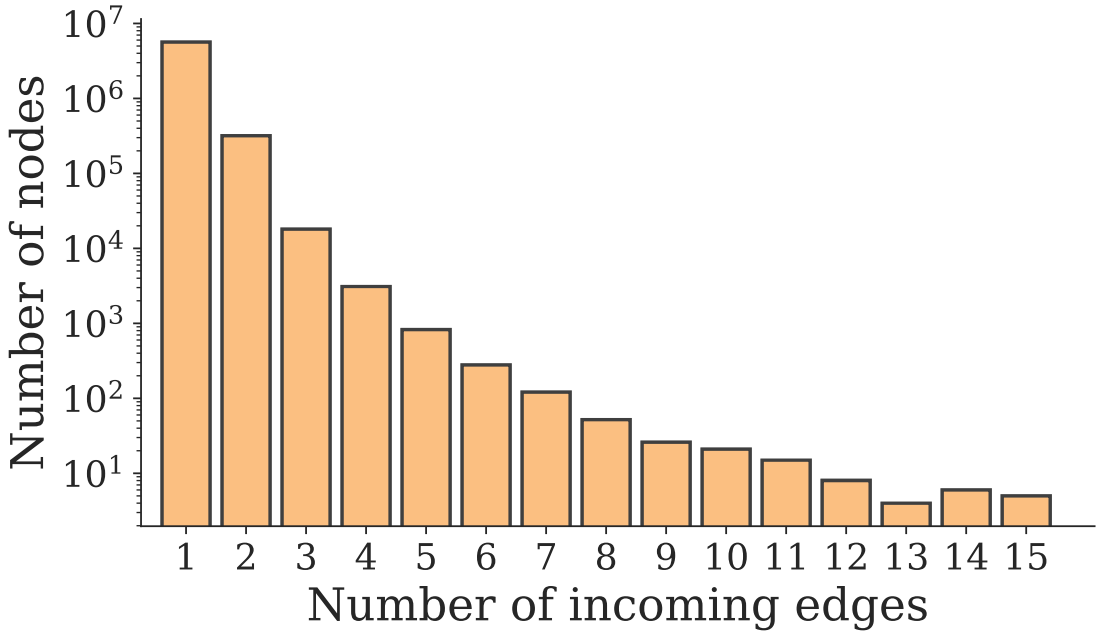
Splitting EUD graphs into trees: A quick and clatty approach

Mark Anderson, Carlos Gómez Rodríguez

Finished last of full submissions! (º ^ º)

Incoming edges

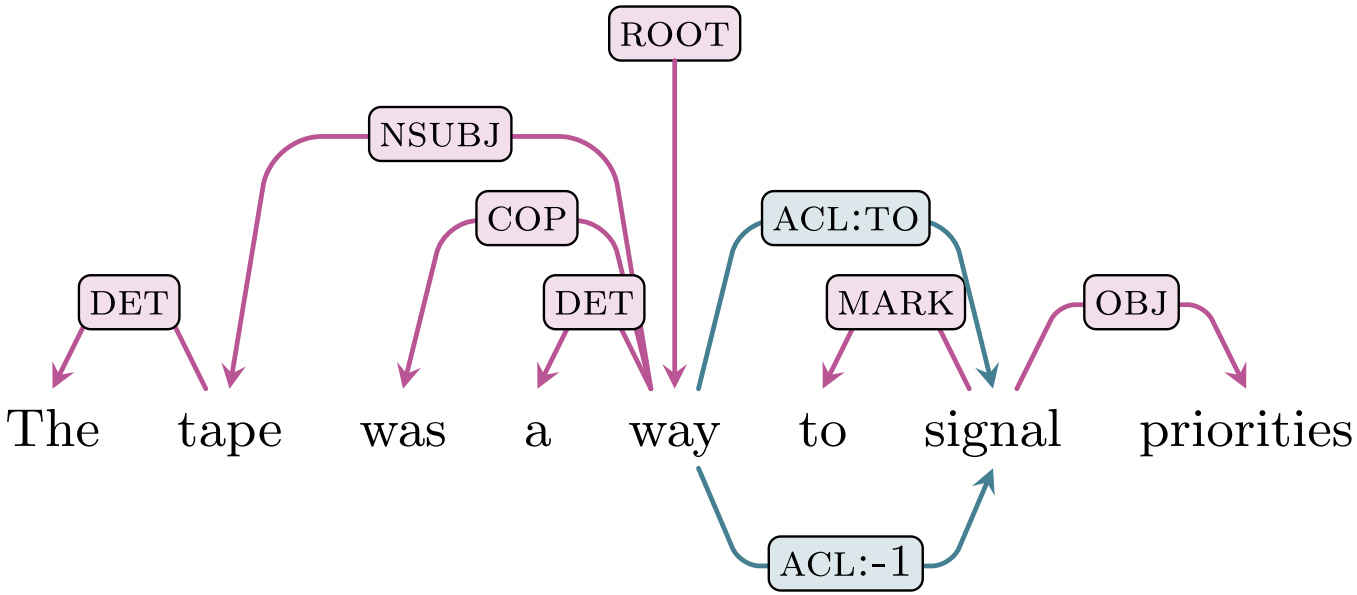


Quick and clatty

- Use sequence-labelling parser
- Bracketing encoding[1]
- Split each treebank into 4 treebanks
- Each new treebank is a tree that captures certain EUD edges
- Train parsers, predict trees for each split
- Collate predicted edges to form EUD graph

[1]M. Strzyz, D. Vilares, and C. Gómez-Rodríguez., *Viable dependency parsing as sequence labeling*, 2019

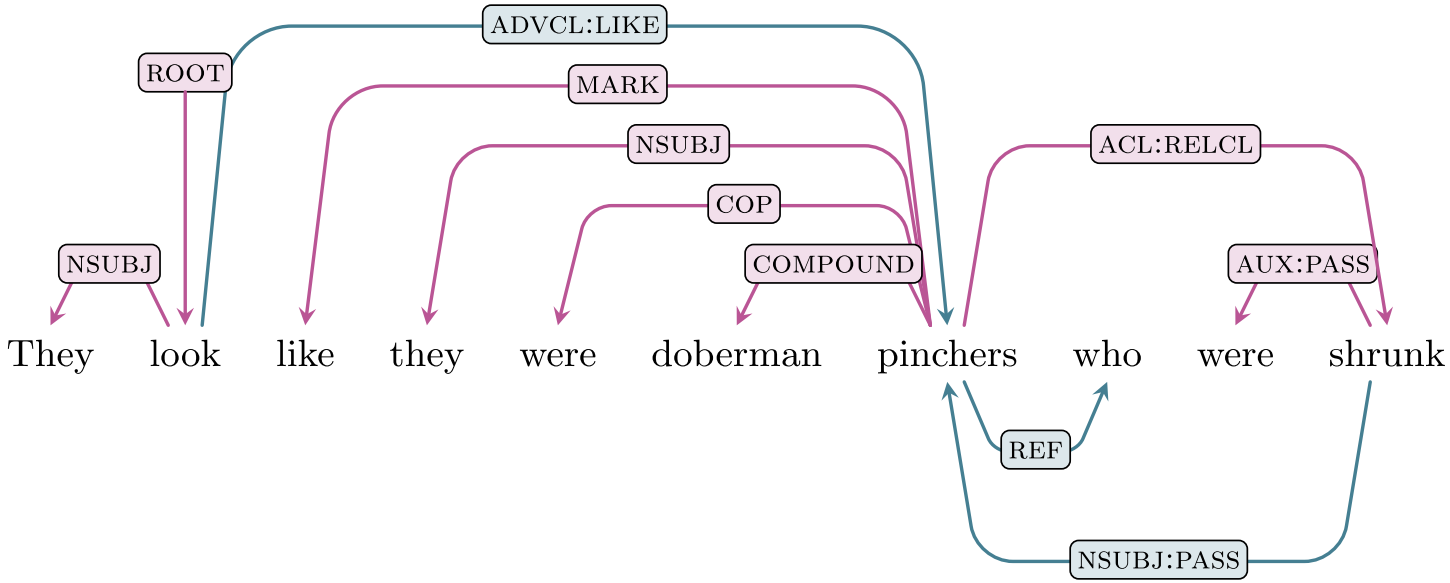
Basic tree (forest)



Edges that correspond to UD tree and relative case marking. Edge-less nodes get attached to dummy root.

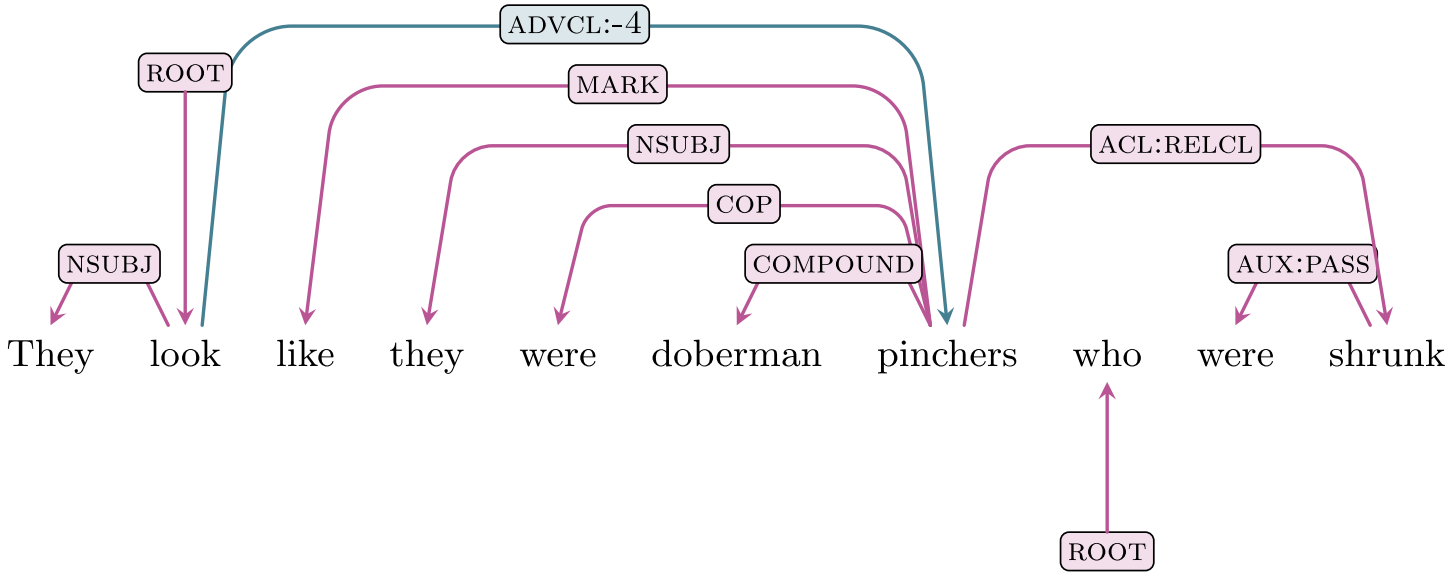
No multi-word case marking coverage. :(

Relative tree (initial graph)



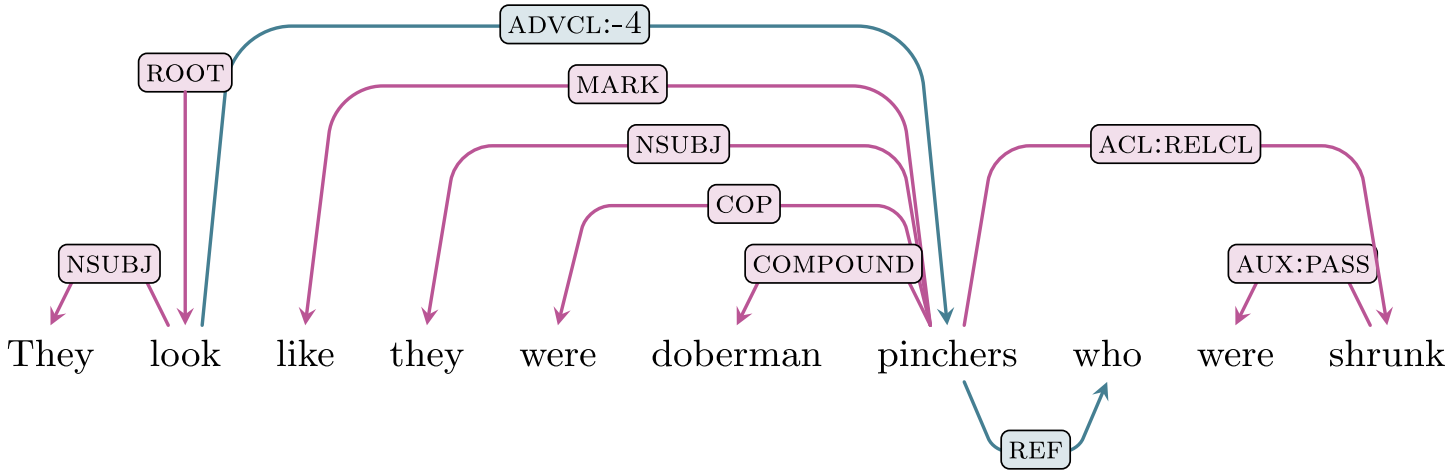
Bit silly.

Relative tree (form basic tree)



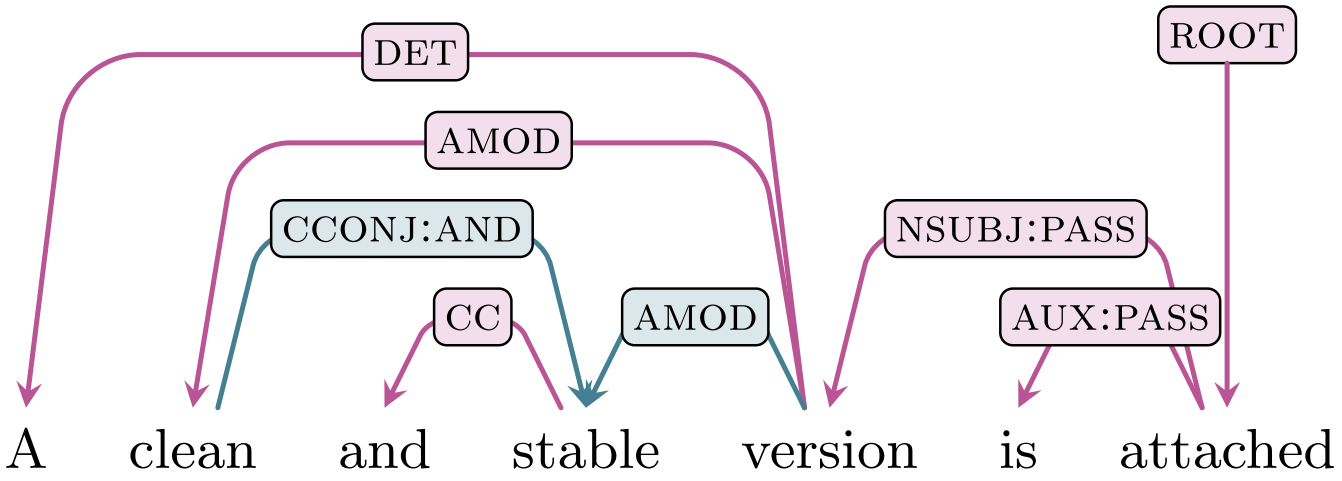
Bit silly.

Relative tree (exchange ref edge)

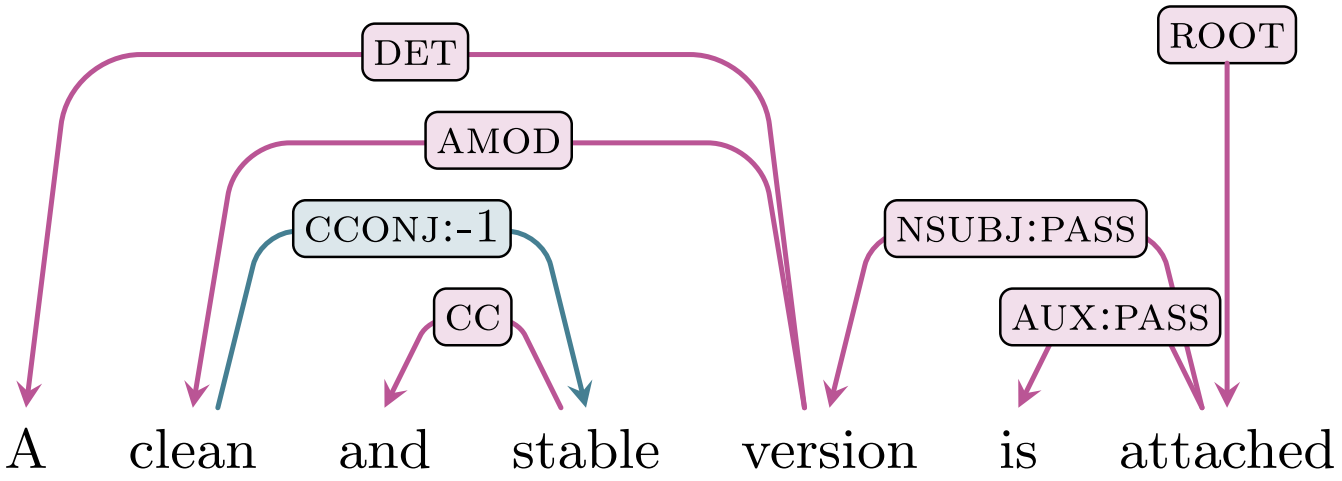


Bit silly.

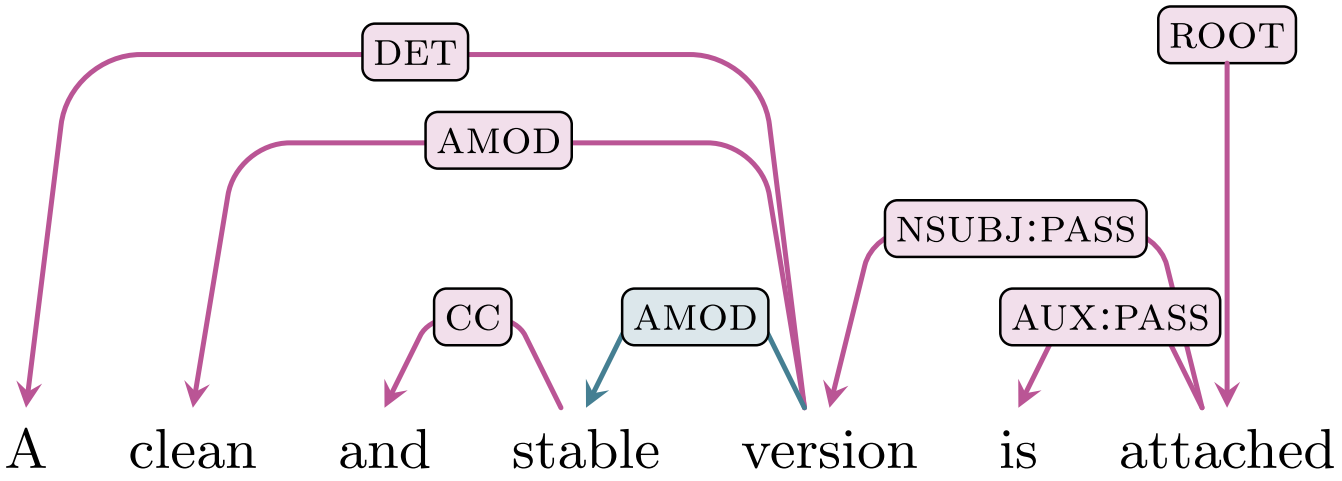
Conjunct tree (initial graph)



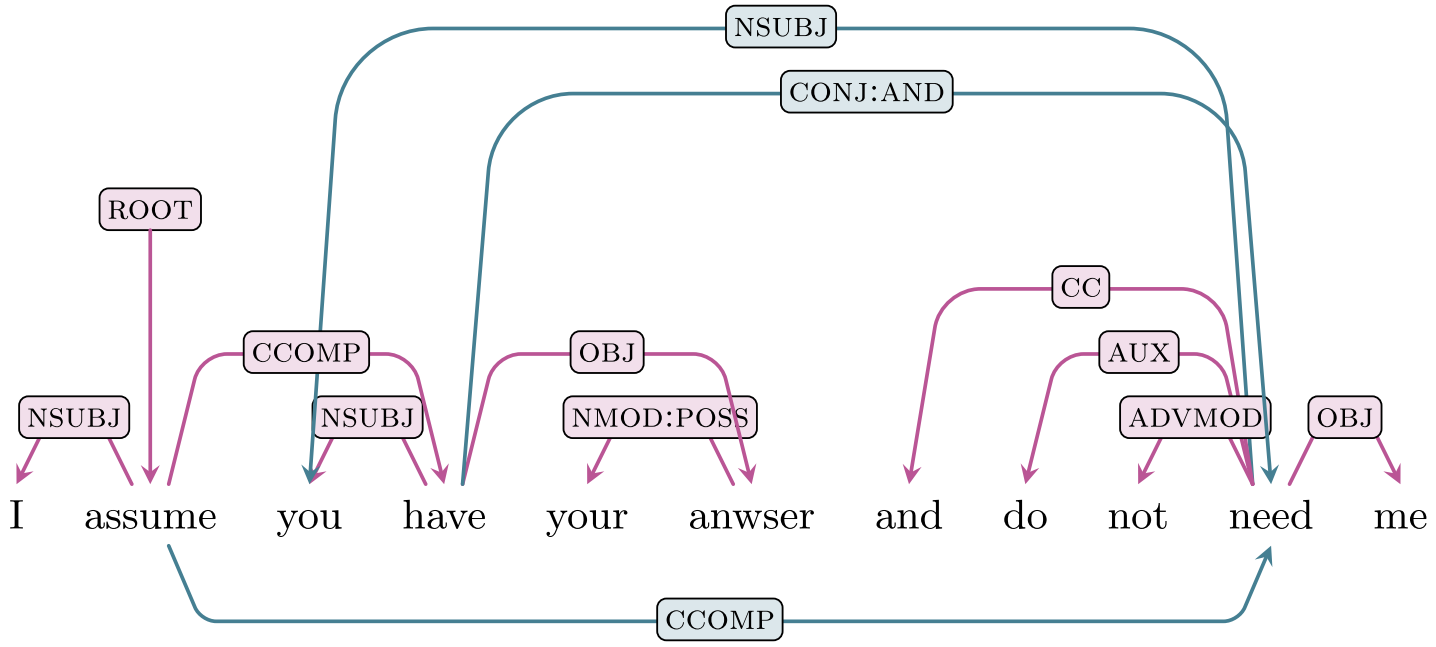
Conjunct tree (form basic tree)



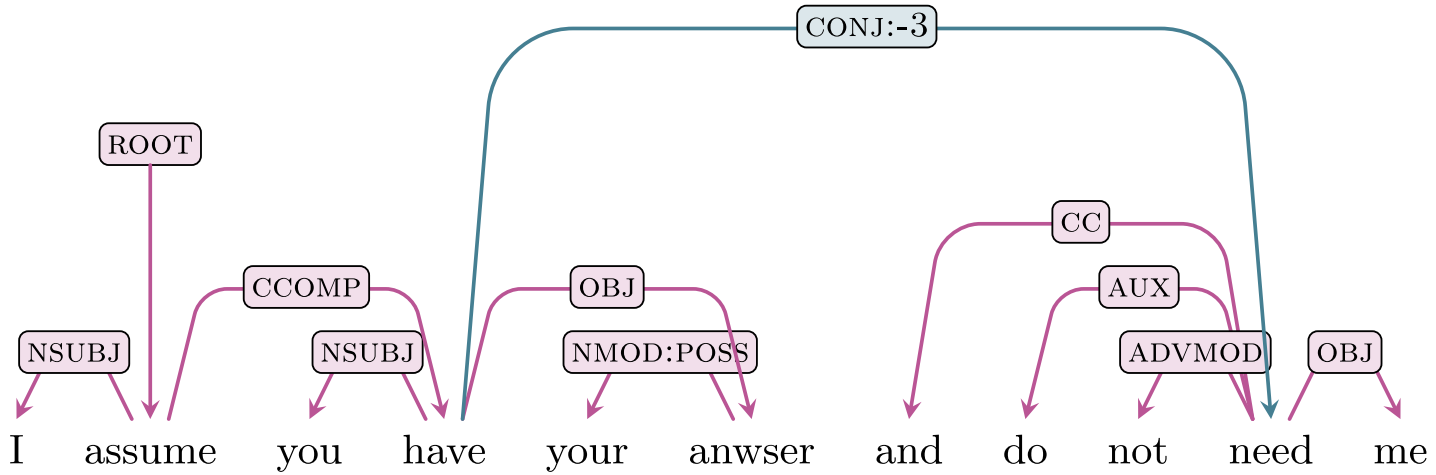
Conjunct tree (replace cconj)



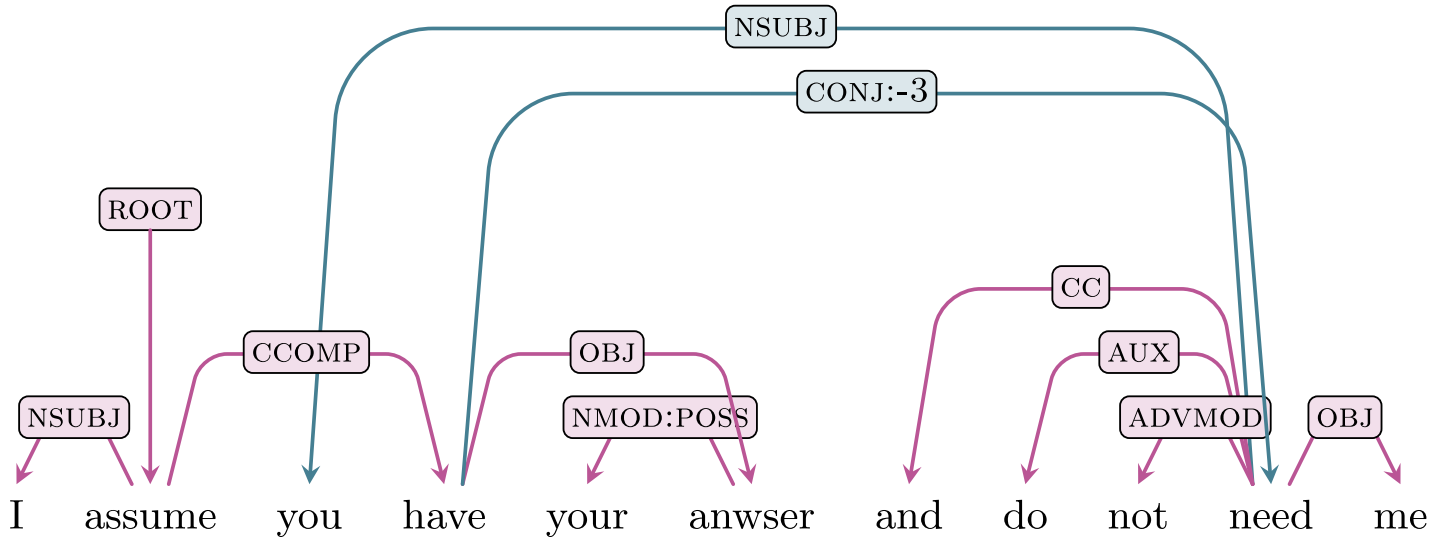
Control tree (initial graph)



Control tree (form basic tree)



Control tree (propagate nsubj)

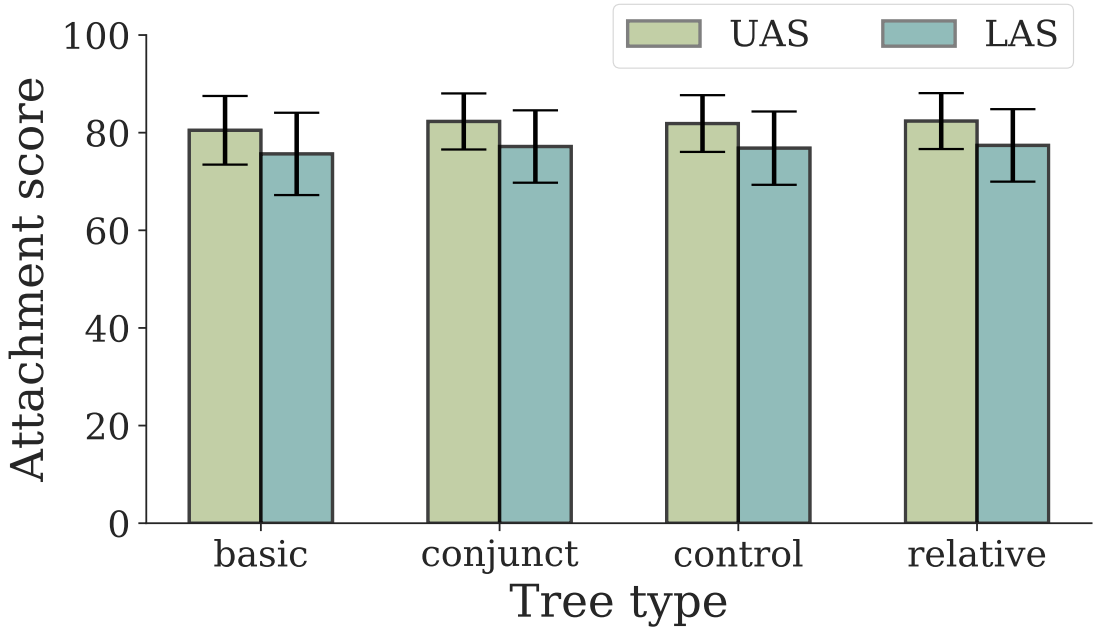


Should have also added the ccomp edge ... :(

	EULAS	ELAS
ar-padt	94.04	81.91
bg-btb	97.09	97.06
cs-cac	94.72	93.18
cs-fictree	94.21	91.75
cs-pdt	94.41	92.36
en-ewt	97.44	97.44
en-gum	97.09	97.09
et-edt	95.61	92.35
et-ewt	95.75	91.27
fi-tdt	92.73	87.13
fr-sequoia	96.22	96.22
it-isdt	96.32	95.98

	EULAS	ELAS
lt-alksnis	94.08	87.35
lv-lvtb	93.77	93.77
nl-alpino	98.07	98.01
nl-lassysmall	97.34	97.30
pl-lfg	99.02	99.02
pl-pdb	96.37	96.19
ru-syntagrus	97.97	97.68
sk-snk	96.23	94.18
sv-talbanken	96.31	96.31
ta-ttb	97.62	93.39
uk-iu	96.35	95.97

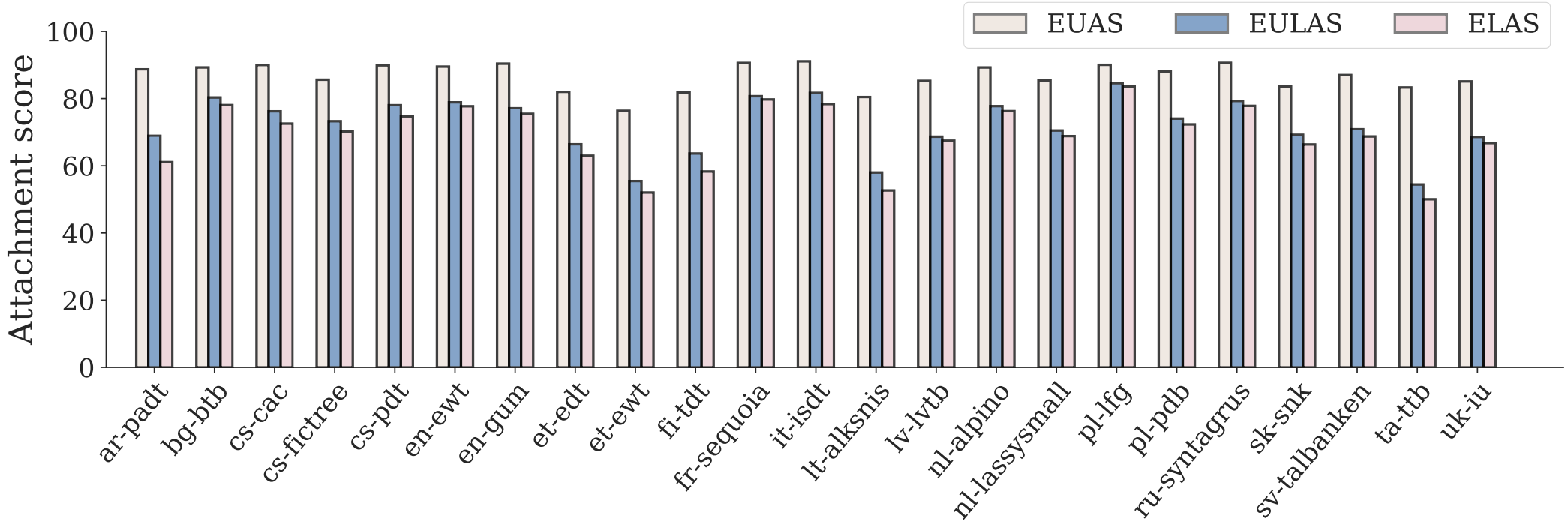
Tree-type performance



Average performance

Metric	Score
EUAS	86.66
EULAS	72.02
ELAS	69.21

Treebank performance



"A BREADTH-FIRST EXPLORATION OF THE SEARCH SPACE OF PARSING TECHNIQUES"