

ERG 2020 Update

Dan Flickinger

CSLI, Stanford University

DELPH-IN Summit 2021

20 July 2021

Overview of changes from ERG 2018

- Punctuation marks now separate tokens
- Documentation strings now on all instances
- Completion of full Redwoods treebank with FFTB
- Better consistency in MRSs

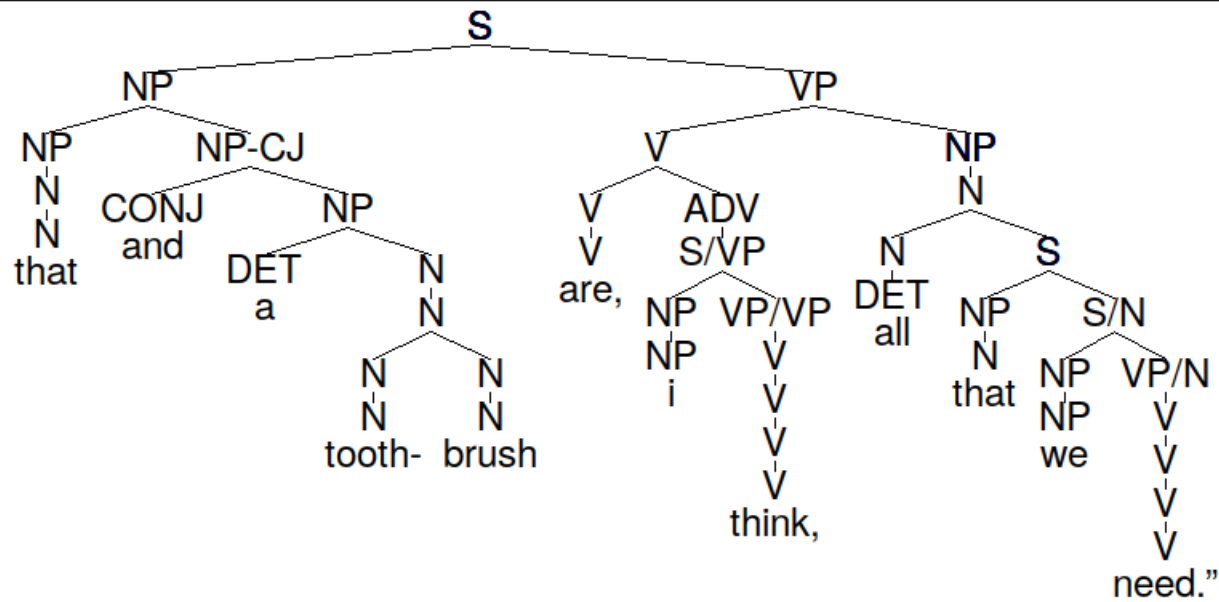


Example of 2018 punctuation analysis

Activities ☐ Unknown ▾

Sun 19:33 ●

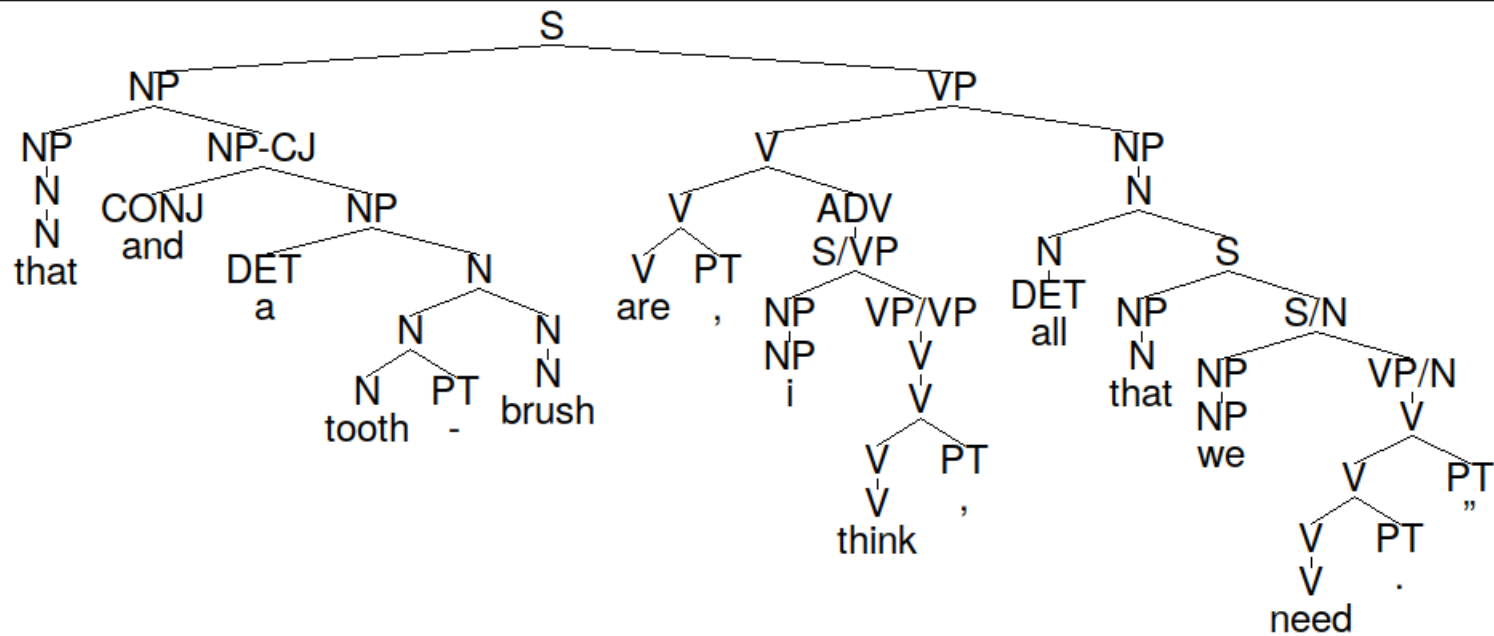
That and a tooth-brush are, I think, all that we need."



Example of 2020 punctuation analysis

Activities ☐ Unknown ▾ Mon 14:37 •

That and a tooth-brush are, I think, all that we need."



Example of in-line documentation strings

```
vp_sbrd-pre-lx_c := subconj_prdp_v_init_lex_rule &  
""
```

Predicative subordinate phrase from lexical VP: pre-head modifier

```
<ex>[Smiling,] Kim arose.  
""
```

```
[ SYNSEM.PUNCT.RPUNCT comma_punct,  
  RNAME scv1 ].
```



Completion of Redwoods update with FFTB

- 2018 included only 6 of 22 sections of WSJ
- 2020 includes all 22 WSJ sections (now FFTB) that were in 1214
- Enabled training of better model for parse selection
Better than versions 1214 and 2018
- Also includes 1000 sentences from section 23 for testing



Scoring model on held-out profiles

- Partial model trained on top 500 parses, missing WSJ sections 6-21
- Full 2020 model trained on top 250 parses, including WSJ sections 0-21

Profile	Partial (800K words)		2020 (1.4M words)	
	Exact tree %	Elem Dep F1	Exact tree %	Elem Dep F1
wsj23	19.2	90.53	32.6	94.56
cb	21.1	89.63	28.7	92.60
ws213	30.4	89.47	36.2	92.12
jhu+k	47.6	93.61	52.2	94.84
vm32	71.6	94.65	67.9	95.42
ecpr	69.9	96.28	68.8	96.34
All	42.6	92.31	48.0	94.28



Better consistency in MRS outputs

- Aiming at improved compliance with (revised) semantic algebra
- Making use of error diagnosis tools already available for treebank
- Fixing bugs as they are reported

