

# Evaluation of Robust Parsing Methods

Woodley Packard

August 8th, 2017

PCFGs for ERG-1214:

<http://sweaglesw.org/linguistics/csaw/download/>

- ▶ `--pcfg=something.pcfg` command-line option
- ▶ Parse with CSAW and ERG, and some hybrid edges
- ▶ Maxent model picks winner

PCFGs for ERG-1214:

<http://sweaglesw.org/linguistics/csaw/download/>

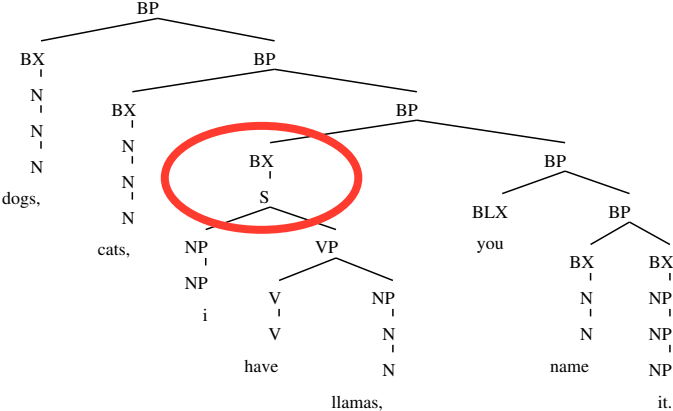
- ▶ `--pcfg=something.pcfg` command-line option
- ▶ Parse with CSAW and ERG, and some hybrid edges
- ▶ Maxent model picks winner
- ▶ ... evaluation soon

# The suspects

- ▶ ERG 1214 by itself Flickinger [2000, 2011]
- ▶ Bridging Flickinger and Packard [2015]
- ▶ Pacman
- ▶ CSAW Zhang and Krieger [2011], Zhang et al. [In prep]
- ▶ Hybrid (`ace --pcfg=...`)

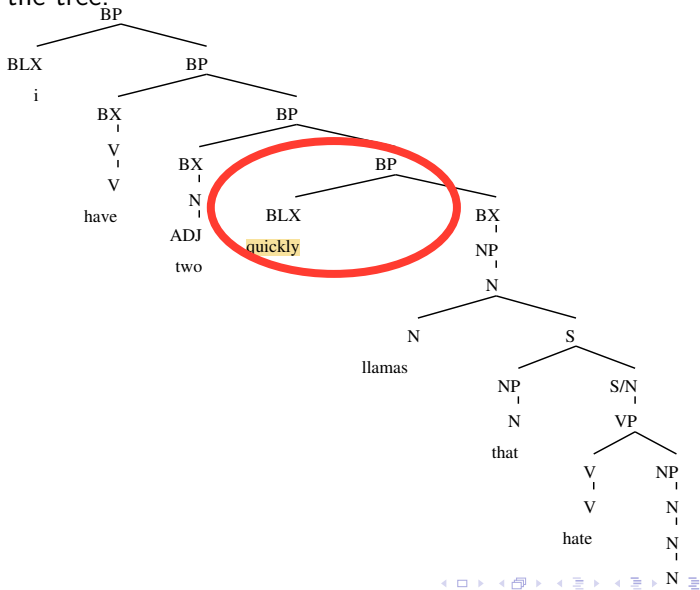
# Bridging

“Sentential” analysis = any concatenation of grammatical subanalyses.



# Bridging

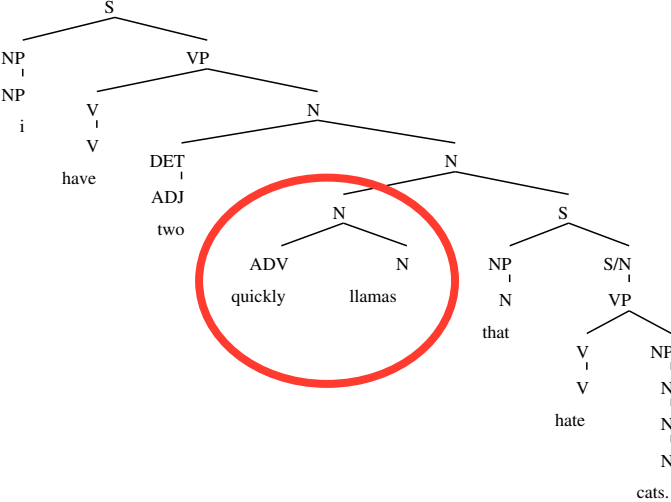
Fails badly if the ungrammaticality would be at a low level in the tree.



# Pacman

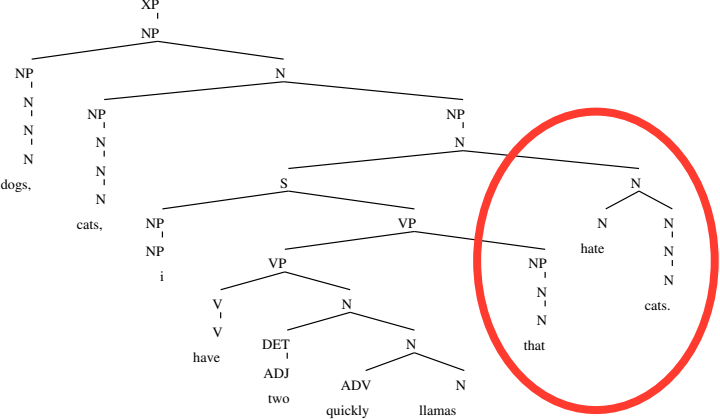
Nbar and Vbar can gobble up arbitrary lexical constituents without changing their own valency (Flickinger).

References



# Pacman

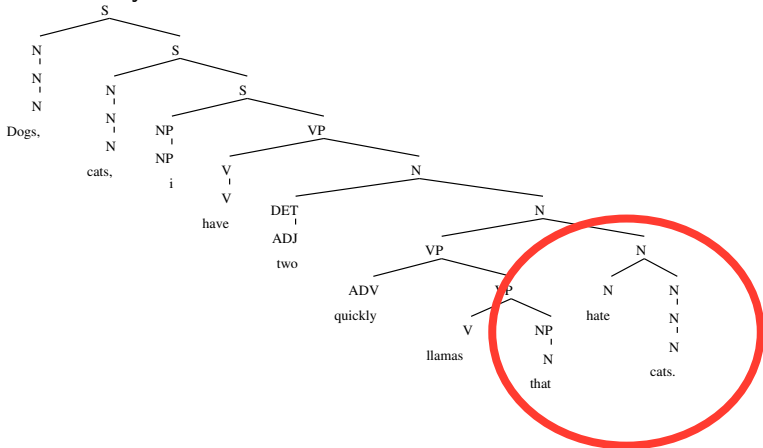
Fails badly if the ungrammaticality would be at a high level in the tree.



References



Uses a PCFG; no moral objections to doing really weird stuff if necessary to maintain order elsewhere.

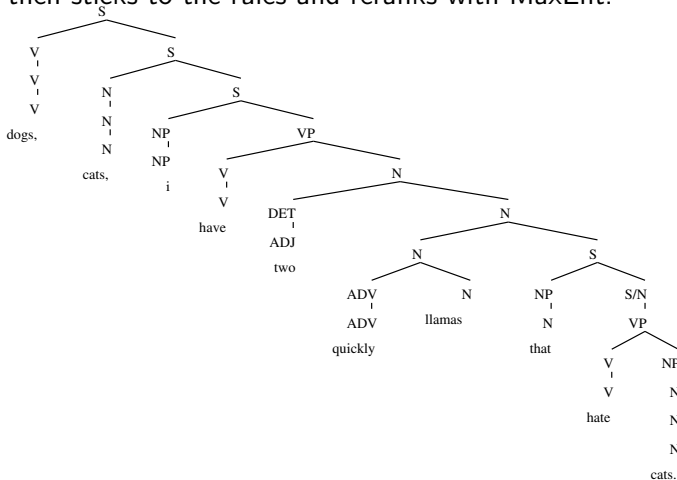


Two versions: different amounts of training data, different degrees of grandparenting.

System	Non-terminals	Rules	Training set
csaw-tb	236	~36K	~100K
csaw-ww	155042	~5M	~50M

# Hybrid

Takes just a single tree recommendation from PCFG, and then sticks to the rules and reranks with MaxEnt.



- ▶ ... is hard, because the interesting text to evaluate on is the text the grammar can't parse by itself, which means there's no gold treebank.
- ▶ ... need to come up with some data, then use EDM  
Zhang et al. [In prep].

# Evaluation datasets

dataset	items	$\frac{\#words}{\#items}$	in 1212?	version
alchemy45	45	28.5	no parse	1212
ws13r	207	27.7	no parse	trunk
semcor	241	24.8	no gold	trunk
wsj00ab	76	25.6	no gold	trunk

# Results (coverage)

Method	alchemy45	ws13r	semcor	wsj00ab
baseline	17.78%	61.35%	88.74%	84.21%
bridging	66.67%	74.40%	90.91%	90.79%
pacnv+ut	53.33%	54.11%	87.45%	85.53%
csaw-tb	97.78%	77.29%	98.27%	98.68%
csaw-ww	<b>100.00%</b>	83.09%	<b>100.00%</b>	<b>100.00%</b>
hybrid-tb	97.78%	88.41%	<b>100.00%</b>	98.68%
hybrid-ww	<b>100.00%</b>	<b>90.82%</b>	99.13%	<b>100.00%</b>

# Results (time in seconds per item)

Method	alchemy45	ws13r	semcor	wsj00ab
baseline	44.1	31.4	7.7	10.6
bridging	41.2	48.3	22.4	21.9
pacnv+ut	29.0	89.7	16.6	27.6
csaw-tb	<b>2.1</b>	<b>1.0</b>	<b>0.6</b>	<b>0.7</b>
csaw-ww	247.0	224.3	135.8	132.2
hybrid-tb	39.2	20.4	13.4	14.1
hybrid-ww	219.2	228.2	122.0	119.8

# Results (EDM F1)

Method	alchemy45	ws13r	semcor	wsj00ab
baseline	28.83	52.28	79.60	72.78
bridging	42.07	44.31	69.61	66.11
pacnv+ut	43.85	42.27	72.70	66.16
csaw-tb	68.51	48.87	67.81	67.26
csaw-ww	<b>77.11</b>	60.98	78.51	74.48
hybrid-tb	69.76	63.73	78.76	76.17
hybrid-ww	75.56	<b>68.47</b>	<b>81.52</b>	<b>78.61</b>



- ▶ Coverage improvement can be dramatic
- ▶ `csaw-tb` is fast
- ▶ Accuracy costs speed
- ▶ None of the systems is incredibly accurate
- ▶ Future work: extrinsic evaluation

# Thank You!

## References:

- Dan Flickinger. On building a more efficient grammar by exploiting types. *Natural Language Engineering*, 6(01): 15–28, 2000.
- Dan Flickinger. Accuracy v. Robustness in grammar engineering. In Emily M. Bender and Jennifer E. Arnold, editors, *Language from a Cognitive Perspective: Grammar, Usage and Processing*, pages 31–50. CSLI Publications, Stanford, CA, USA, 2011.
- Dan Flickinger and Woodley Packard. Robust parsing in HPSG: Bridging the coverage chasm. In *poster presented at the 22nd International Conference on HPSG*, 2015.
- Yi Zhang and Hans-Ulrich Krieger. Large-scale corpus-driven PCFG approximation of an HPSG. In *12th International Conference on Parsing Technologies*, pages 198–208. Association for Computational Linguistics, 2011.
- Yi Zhang, Stephan Oepen, Rebecca Drīdan, Dan Flickinger, and Hans-Ulrich Krieger. Robust parsing, meaning.