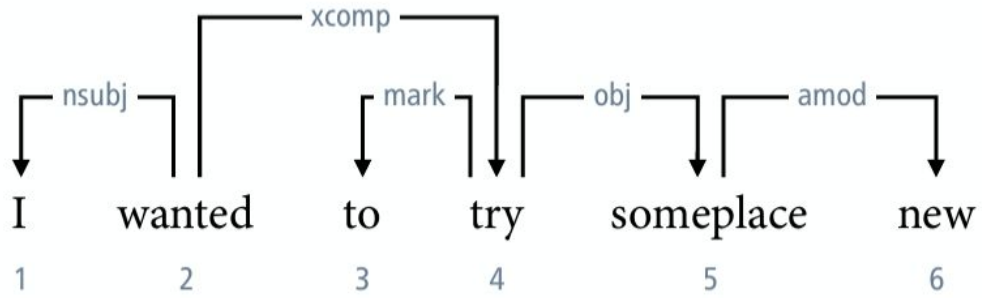
Locally normalized beam search in a dependency parser

Adam N. Alfred J. Angelo V. (Daniel P.)

Dependency parsing

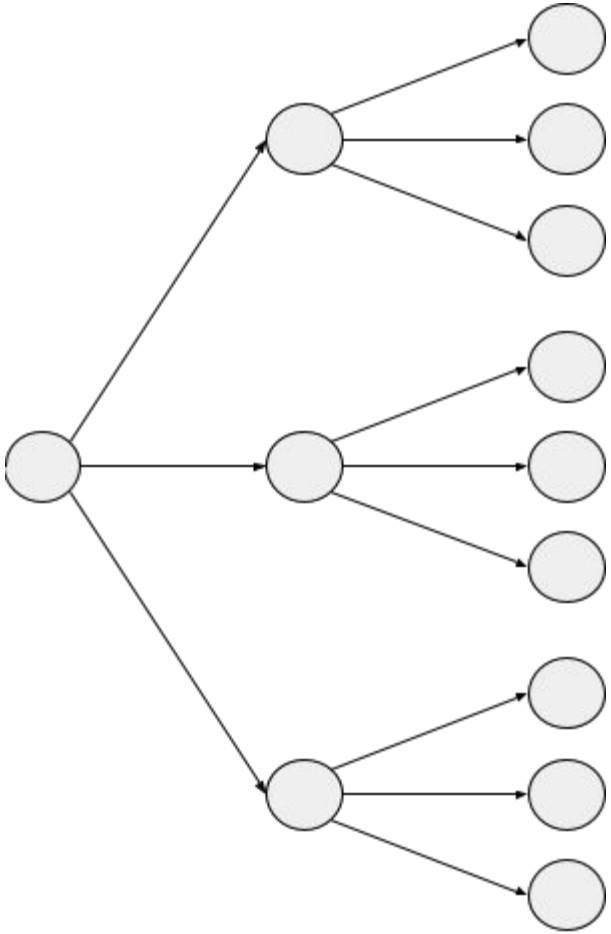


word position	1	2	3	4	5	6
head position	2	0	4	2	4	5
dependency relation	nsubj	root	mark	xcomp	obj	amod

Source: Marco Kuhlmann, lecture 4.04

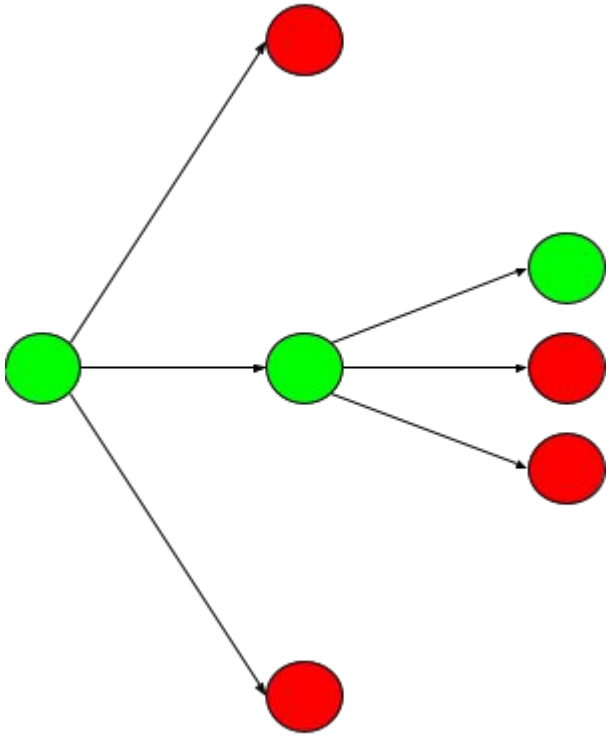
Beam Search

- Score given to each choice
- Baseline (Greedy) only takes optimal choice in each step
- Beam search considers a number of alternatives at once
  - Amount considered = beam size



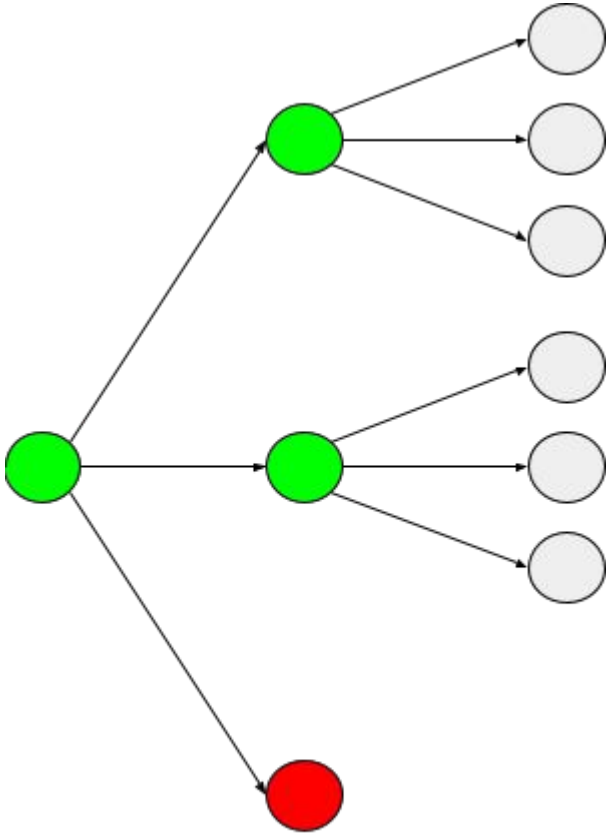
Beam Search

- Score given to each choice
- Baseline (Greedy) only takes optimal choice in each step
- Beam search considers a number of alternatives at once
  - Amount considered = beam size



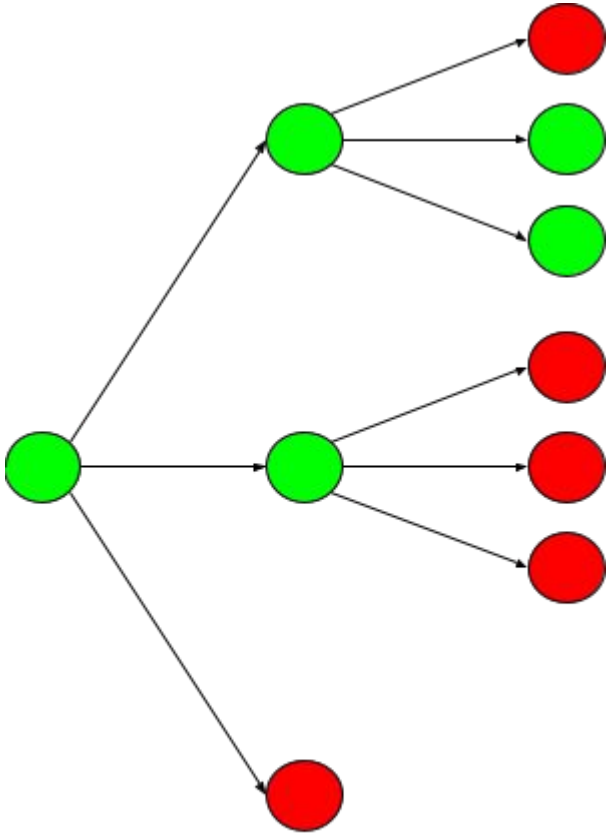
Beam Search

- Score given to each choice
- Baseline (Greedy) only takes optimal choice in each step
- Beam search considers a number of alternatives at once
  - Amount considered = beam size



Beam Search

- Score given to each choice
- Baseline (Greedy) only takes optimal choice in each step
- Beam search considers a number of alternatives at once
  - Amount considered = beam size



Local Normalization

The parser uses log softmax of the score

Local normalization means that the calculated value is used when calculating the value of a sequence

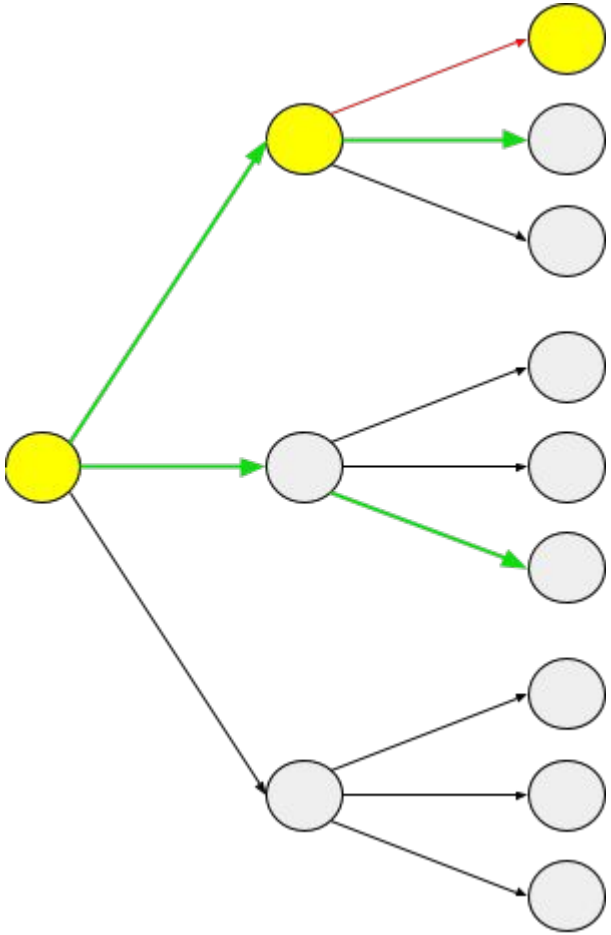
Instead of the raw score

Training

- Local Classifier with error states (Ashish Vaswani, Kenji Sagae, 2016)
- Structured learning
  - Structured perceptron (Collins, 2002)
  - Early update (Collins and Roark, 2004)
  - Global normalization (Andor, 2016)

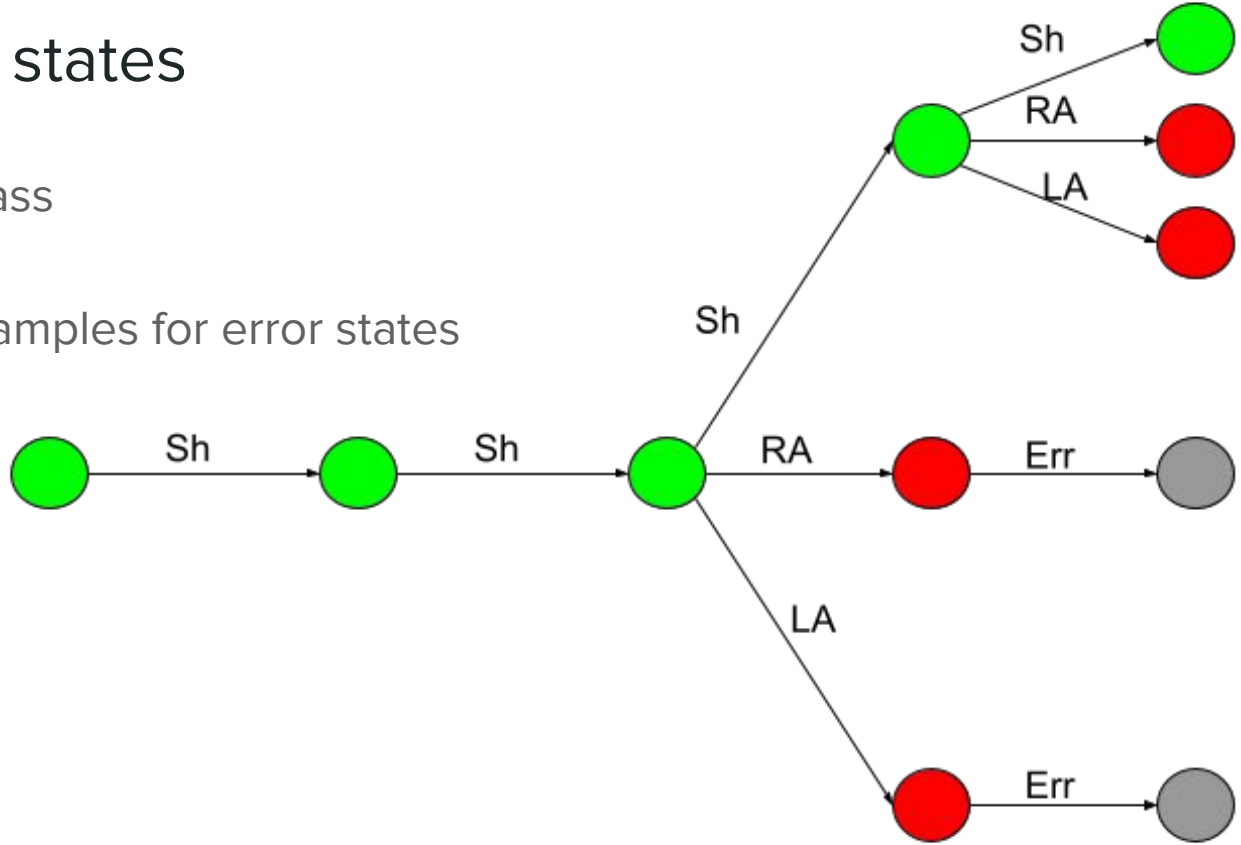
Early update

- Gold sequence of decisions
- Once gold_path falls out off beam
  - Terminate search and update
- If item is final item in gold_path
  - Stop search and update

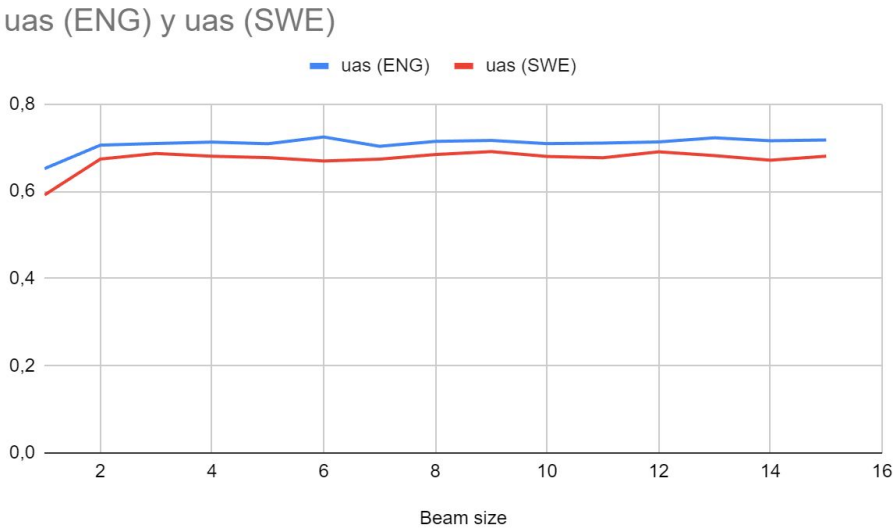


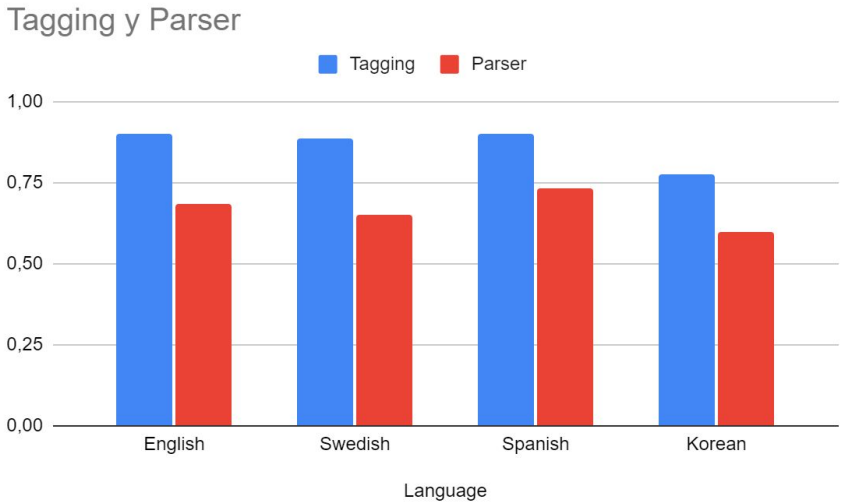
Training with Error states

- Introduce an Error class
- Local classifier
- Generate training examples for error states

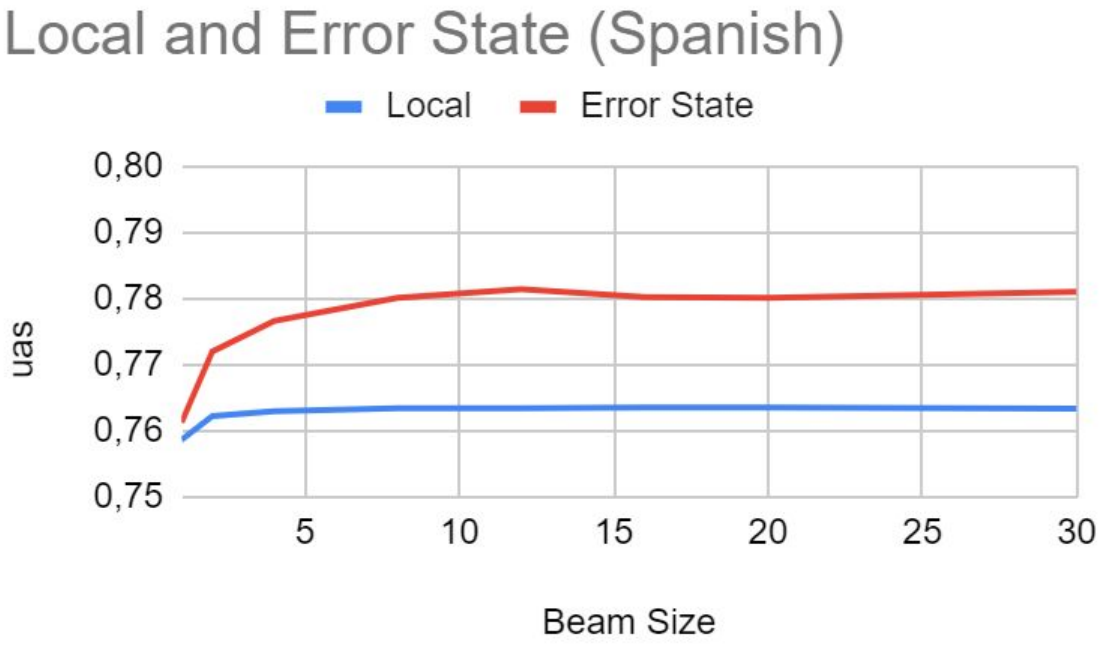


Results


uas (ENG) y uas (SWE)
uas (ENG)
uas (SWE)
0,8
0,6
0,4
0,2
0,0
2
4
6
8
10
12
14
16
Beam size


Tagging y Parser
Tagging
Parser
1,00
0,75
0,50
0,25
0,00
English
Swedish
Spanish
Korean
Language

Results


Local and Error State (Spanish)
Local
Error State
0,80
0,79
0,78
0,77
0,76
0,75
uas
5
10
15
20
25
30
Beam Size

Conclusions

- Beam size quickly plateaus
- Beam search doesn’t appear to particularly favor certain languages.
  - Different baselines
- Training with error states improves UAS

Research papers

- Global normalization
  - Andor et al.
- Error states
  - Vaswani et al.