CS460/626 : Natural Language Processing/Speech, NLP and the Web
(Lecture 15, 16– Parser comparison)

Pushpak Bhattacharyya
CSE Dept.,
IIT Bombay
6th and 7th Feb, 2012
Deterministic Parser vs. Probabilistic parser

In probabilistic parser, no hard decision on grammaticality or ungrammaticality, but a score is given.
The problem with “ING” forms

Flying planes can be dangerous.

Adj VBG

Noun – Swimming is good for health

“ING” Verb – He is swimming

Adj – I saw the swimming crocodile

In general, the POS Tagger cannot disambiguate, since arbitrary amount of text can modify the words (e.g., planes which were invented in 20s at the time of ... )
CFG vs PCFG

- PCFG is more powerful than CFG

- “Colorless green ideas sleep furiously”
  - PCFG gives a low score due to counts, even though CFG will pass it

- \( P(S) \) – language modeling
  - It gives how likely is \( S \) to occur in the language
parsers comparison
(Charniack, Collins, Stanford, RASP)

study by masters students:
Avishek, Nikhilesh, Abhishekh and Harshada
Parser comparison: Handling ungrammatical sentences
Charniak (ungrammatical 1)

- Here *has* is tagged as AUX

```
NP  S
  |   |  VP
  |   |   | NP
  |   |   |   | AUX
  |   |   |   |   | VBG
  |   |   |   |   | NP
  |   |   |   |   |   | DT NN
  |   |   |   |   |   | reading the book
  |   |   |   |   |   | Joe has reading the book
  |   |   |   |   |   | has
  |   |   |   |   |   | the book
NNP Joe
```
Win is treated as a verb and it does not make any difference whether it is in the present or the past tense.
Collins (ungrammatical 1)

- *Has*
- should
- have
- been
- AUX.
Collins (ungrammatical 2)

- Same as charniack
Stanford (ungrammatical 1)

- *has* is treated as VBZ and not AUX.
Stanford (ungrammatical 2)

- Same as Charniak
RASP (ungrammatical 1)

- Inaccurate tree
Observation

- For the sentence ‘Joe has reading the book’ Charniak performs the best; it is able to predict that the word ‘has’ in the sentence should actually be an AUX.

- Though both the RASP and Collins can produce a parse tree, they both cannot predict that the sentence is not grammatically correct.

- Stanford performs the worst, it inserts extra ‘S’ nodes into the parse tree.
Observation (contd.)

- For the sentence ‘The book was win by Joe’, all the parsers give the same parse structure which is correct.
Ranking in case of multiple parses
John said Marry sang the song with Max

The parse produced is semantically correct

Charniak (Multiple Parses 1)
Charniak (Multiple Parses 2)

- PP is attached to NP which is one of the correct meanings

I saw a boy with telescope
Collins (Multiple Parses 1)

- Same as Charniak.
Collins (Multiple Parses 2)

- Same as Charniak
Stanford (Multiple Parses 1)

- PP is attached to VP which is one of the correct meanings possible.
Stanford (Multiple Parses 2)

- Same as Charniak.
PP is attached to VP.
RASP (Multiple Parses 2)

- The change in the pos tags as compared to charniak is due to the different corpora but the parse trees are comparable.
Observation

- All of them create one of the correct parses whenever multiple parses are possible.

- All of them produce multiple parse trees and the best is displayed based on the type of the parser:
  - Charniak: Probabilistic Lexicalised Bottom-Up Chart Parser
  - Collins: Head-driven statistical Beam Search Parser
  - Stanford: Probabilistic A* Parser
  - RASP: Probabilistic GLR Parser
Time taken

- 54 instances of the sentence ‘This is just to check the time’ is used to check the time

- Time taken
  - Collins : 40s
  - Stanford : 14s
  - Charniak : 8s
  - RASP : 5s
Embedding Handling
The cat that killed the rat that stole the milk that spilled on the floor that was slippery escaped.
Charniak (Embedding 2)
Collins (Embedding 1)
Collins (Embedding 2)
Stanford (Embedding 1)
Stanford (Embedding 2)
RASP (Embedding 1)
RASP (Embedding 2)
Observation

- For the sentence ‘The cat that killed the rat that stole the milk that spilled on the floor that was slippery escaped.’ all the parsers give the correct results.

- For the sentence ‘John the president of USA which is the most powerful country likes jokes’: RASP, Charniak and Collins give correct parse, i.e., it attaches the verb phrase ‘likes jokes’ to the top NP ‘John’.

- Stanford produces incorrect parse tree; attaches the VP ‘likes’ to wrong NP ‘the president of …’
Handling multiple POS tags
Charniak (multiple pos 1)

Time flies like an arrow

Time flies like an arrow

Fire him immediately

Fire him immediately
Don't toy with the pen
Collins (multiple pos 1)
Collins (multiple pos 2)
Stanford (multiple pos 1)
Stanford (multiple pos 2)
RASP (multiple pos 1)
RASP (multiple pos 2)
Observation

- All but RASP give comparable pos tags. In the sentence ‘Time flies like an arrow’ RASP give flies as noun.
- In sentence ‘Don’t toy with the pen’, all parsers are tagging ‘toy’ as noun.
Repeated Word handling
Buffalo Meanings

- Buffalo
  - Proper Noun
  - animal
  - bully
  (place in USA)
Buffalo buffaloes Buffalo buffaloes buffalo
buffalo Buffalo buffaloes
Collins
Stanford
RASP
Correct parse
Another sentence of same structure

```
S
  NP
    NNS
    Brown
    NNS
    cows

VP
  NP
    NNP
    Black
    NNS
    cows

  VP
    V
    NNS
    cow

  NP
    NP
    white
    NNS
    cows
```
Dependency Tree for Buffalo Example

- Subj: Buffalo buffaloes
- Mod: Buffalo buffaloes
- Main Verb: buffaloes

Example: Buffalo buffaloes
Why do different parsers behave the way they do for the Buffaloes example

- The trees given by the following parsers have syntactic similarity to that of the given examples.
  - Collins
    - *Small village teachers teach small boys good mathematics.*
  - Stanford
    - *John says Americans hunt back, beautiful African buffaloes.*
  - Charniak
    - *Ram declared Shyam said red, blood corpuscles die.*
Observation

- Collins and Charniak come close to producing the correct parse.
- RASP tags all the words as nouns.
Long sentences

- Given a sentence of 394 words, only RASP was able to parse.
One day, Sam left his small, yellow home to head towards the meat-packing plant where he worked, a task which was never completed, as on his way, he tripped, fell, and went careening off of a cliff, landing on and destroying Max, who, incidentally, was also heading to his job at the meat-packing plant, though not the same plant at which Sam worked, which he would be heading to, if he had been aware that that the plant he was currently heading towards had been destroyed just this morning by a mysterious figure clad in black, who hailed from the small, remote country of France, and who took every opportunity he could to destroy small meat-packing plants, due to the fact that as a child, he was tormented, and frightened, and beaten savagely by a family of meat-packing plants who lived next door, and scarred his little mind to the point where he became a twisted and sadistic creature, capable of anything, but specifically capable of destroying meat-packing plants, which he did, and did quite often, much to the chagrin of the people who worked there, such as Max, who was not feeling quite so much chagrin as most others would feel at this point, because he was dead as a result of an individual named Sam, who worked at a competing meat-packing plant, which was no longer a competing plant, because the plant that it would be competing against was, as has already been mentioned, destroyed in, as has not quite yet been mentioned, a massive, mushroom cloud of an explosion, resulting from a heretofore unmentioned horse manure bomb manufactured from manure harvested from the farm of one farmer J. P. Harvenkirk, and more specifically harvested from a large, ungainly, incontinent horse named Seabiscuit, who really wasn't named Seabiscuit, but was actually named Harold, and it completely baffled him why anyone, particularly the author of a very long sentence, would call him Seabiscuit; actually, it didn't baffle him, as he was just a stupid, manure-making horse, who was incapable of cognitive thought for a variety of reasons, one of which was that he was a horse, and the other of which was that he was just knocked unconscious by a flying chunk of a meat-packing plant, which had been blown to pieces just a few moments ago by a shifty character from France.
Partial RASP Parse of the sentence