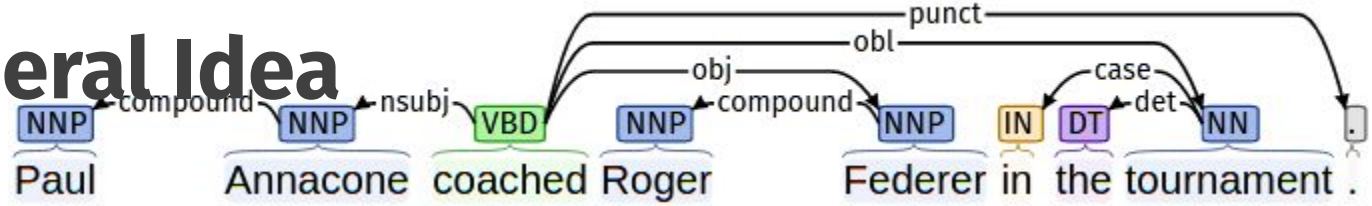
MANCHESTER 1824
The University of Manchester

Extraction Template	Open Pattern
1. (arg1; be {rel} {prep}; arg2)	{arg1} ↑nsubjpass↑ {rel:postag=VBN} ↓{prep_*}↓ {arg2}
2. (arg1; {rel}; arg2)	{arg1} ↑nsubj↑ {rel:postag=VBD} ↓dobj↓ {arg2}
3. (arg1; be {rel} by; arg2)	{arg1} ↑nsubjpass↑ {rel:postag=VBN} ↓agent↓ {arg2}
4. (arg1; be {rel} of; arg2)	{rel:postag=NN;**type=Person**} ↑nn↑ {arg1} ↓nn↓ {arg2}
5. (arg1; be {rel} {prep}; arg2)	{arg1} ↑nsubjpass↑ {slot:postag=VBN;**lex** ∈**announce**|**name**|**choose...**} ↓dobj↓ {rel:postag=NN} ↓{prep_*}↓ {arg2}

OLLIE: General Idea



- Take high-confidence extractions from ReVerb
- Map to a large corpus of dependency parsed sentences (that contain extraction's words)
- generate *open patterns* (paths in dependency parse) by generalising from observed patterns (subject to some constraints)

(*Annacone*, *coach*, *Federer*)

"*Now coached by Annacone, Federer is winning more titles than ever*"

"*Paul Annacone coached Roger Federer in the tournament.*"

Mausam et al 2012. Open Language Learning for Information Extraction. EMNLP

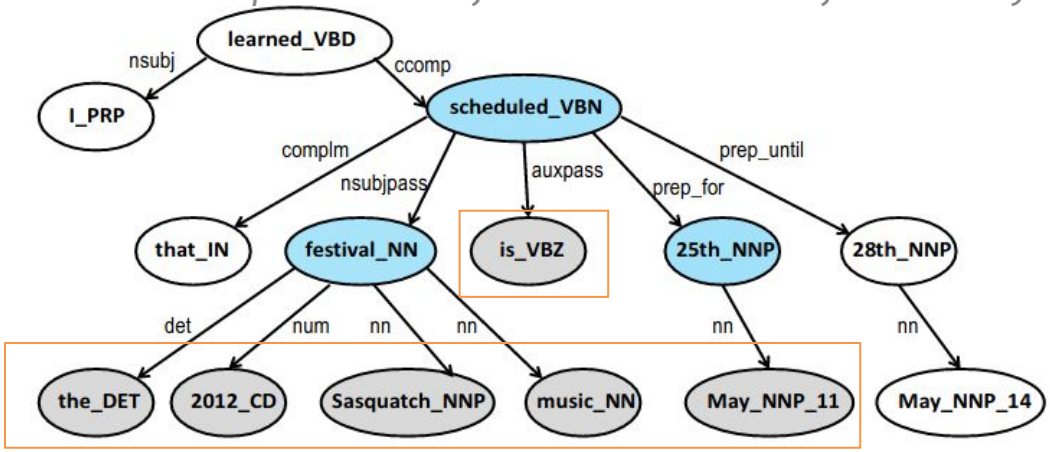
MANCHESTER 1824
The University of Manchester

Extraction Template	Open Pattern
1. (arg1; be {rel} {prep}; arg2)	{arg1} ↑nsubjpass↑ {rel:postag=VBN} ↓{prep_*}↓ {arg2}

OLLIE: Application

- For an input sentence (with dep. parse)
  - Find a matching open pattern
  - Expand arguments along dependency edges

"I learned that the 2012 Sasquatch music festival is scheduled for 25th May until 28th May."



(*the 2012 Sasquatch music festival*, *be scheduled for*, *25 May*)

MANCHESTER 1824
The University of Manchester

OLLIE: Advantages

- Noun-mediated relations
  - *"Bill gates is founder of Microsoft."*
  - Many useful facts are mediated as nouns!
- Argument collection
  - *"John and Mary married."*
  - Relations don't have to be strictly between arguments anymore!

MANCHESTER 1824
The University of Manchester

Stanford OpenIE: Idea

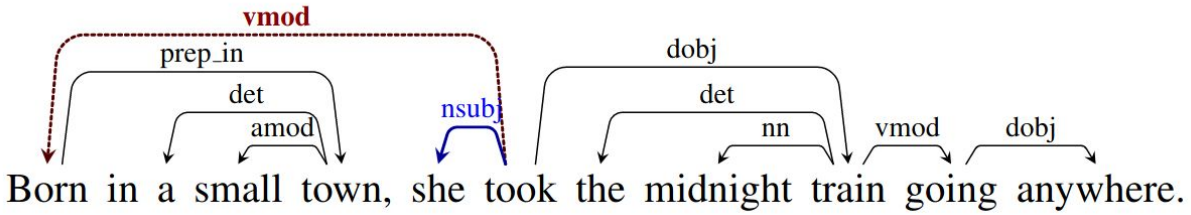
- Stanford OpenIE
  - (learn to) Split to atomic clauses
  - Further remove words with natural logic reasoning
  - Use patterns to extract information from them

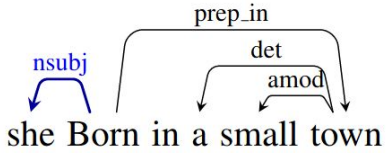
Angeli, G et al 2015. Leveraging Linguistic Structure For Open Domain Information Extraction. ACL

MANCHESTER 1824
The University of Manchester

OpenIE: break into clauses

- Follow dependency structure
- Split along dependency edges
- train classifier to decide, which edges to split





MANCHESTER 1824
The University of Manchester

OpenIE: Refine clauses

- Delete words to yield more clauses
  - “some cute rabbits eat carrots” ⇒ “some rabbits eat carrots”
  - “He has a big gun” ⇒ “he has a gun”
  - “Alice went to the playground with Bob” ⇒ “ Alice went to the playground”
- Yields overlapping arguments
  - Broader coverage
  - More general facts

(he, has, big gun) ✔
(he, has, gun) ✔

MANCHESTER 1824
The University of Manchester

OpenIE: Refine clauses

- Delete words to yield more clauses
  - "all cute rabbits eat carrots" ⇉ "all rabbits eat carrots"
  - "He has a fake gun" ⇉ "he has a gun"
  - "Alice is friends with Bob" ⇉ " Alice is friends"

Natural logic: hard-coded rules (monotonicity here)

non-subjective adjectives: lexicon

prepositional attachment:
co-occurrence heuristic

(he, has, fake gun) ✔
(he, has, gun) ✗

MANCHESTER 1824
The University of Manchester

OpenIE: extract from clauses

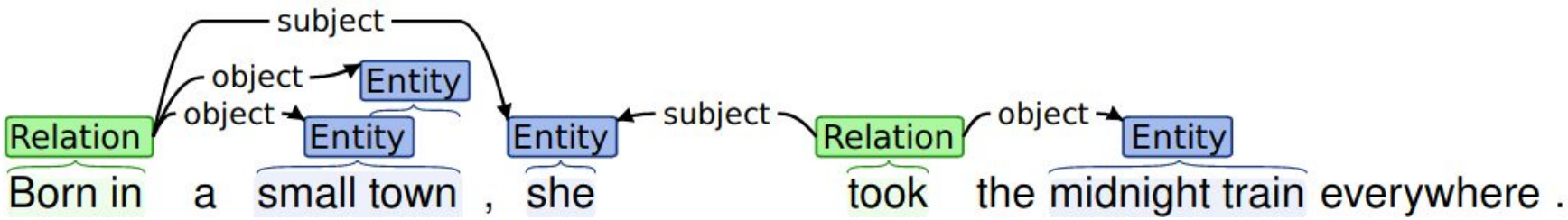
- Clauses “atomic” (can’t be further broken down)
- Simple enough to apply pattern-based extraction

Input	Extraction
cats play with yarn	(cats; play with; yarn)
fish like to swim	(fish; like to; swim)
cats have tails	(cats; have; tails)
cats are cute	(cats; are; cute)
Tom and Jerry are fighting	(Tom; fighting; Jerry)
There are cats with tails	(cats; have; tails)

MANCHESTER 1824
The University of Manchester

OpenIE: put together

"Born in a small town, she took the midnight train everywhere."



(she, born in, small town)
(she, born in, town)
(she, took, the midnight train)