Deep Semantic Analysis of Big Data

Dr. Jesse English

Unbound Concepts

jesse@unboundconcepts.com

Unbound Concepts

- Individualized education
  - One size does not fit all
- Improving literacy
  - Reading more is good
  - Reading more at your level is better
- Common Core

unbound concepts

Unbound Concepts – Reading Levels

- “...at a third grade reading level”
  - This kind of metric is inaccurate, and far too coarse grained to be of use to educators and students
- A – Z reading levels
  - Focus on K – 6
  - Level bands
- Reading Levels ↔ Textual Complexity

unbound concepts

Unbound Concepts – Textual Complexity

- Word counts, high frequency words, etc. are poor indicators
  - John Steinbeck: low word counts, very complex texts
  - Harry Potter novels: very odd words, low textual complexity
- Need deep semantic understanding of texts to judge complexity
- Must be able to produce meaning representations over very large corpora rapidly

unbound concepts

Meaning Representations

- Machine tractable form of text
  - Lossless meaning (explicit)
  - Ideally captures all intention
    - “I really needed a snack.” - Why?
  - Ideally eliminates all ambiguity
    - “Mary had a little lamb.” - How?
    - “It was delicious.”
  - Ideally resolves coreferences in text
    - “The president spoke today. He said we should all just chill out.” - Who?

WIMs

- Weakly Inferred Meanings
- Thematic, frame-based meaning representation
- Produced from syntactic parses coupled with a phrasal lexicon derived from commonly available knowledge resources
- Linear processing time growth

WIMs – Answering Questions Like a Boss

Who?	The actor of the event
What?	The theme of the event
When?	Relative or absolute timing of the event
Where?	Relative or absolute location of the event
Why?	The purpose of the actor's involvement in the event
How?	The instrument used by the actor to fulfill the event

WIMs – Example

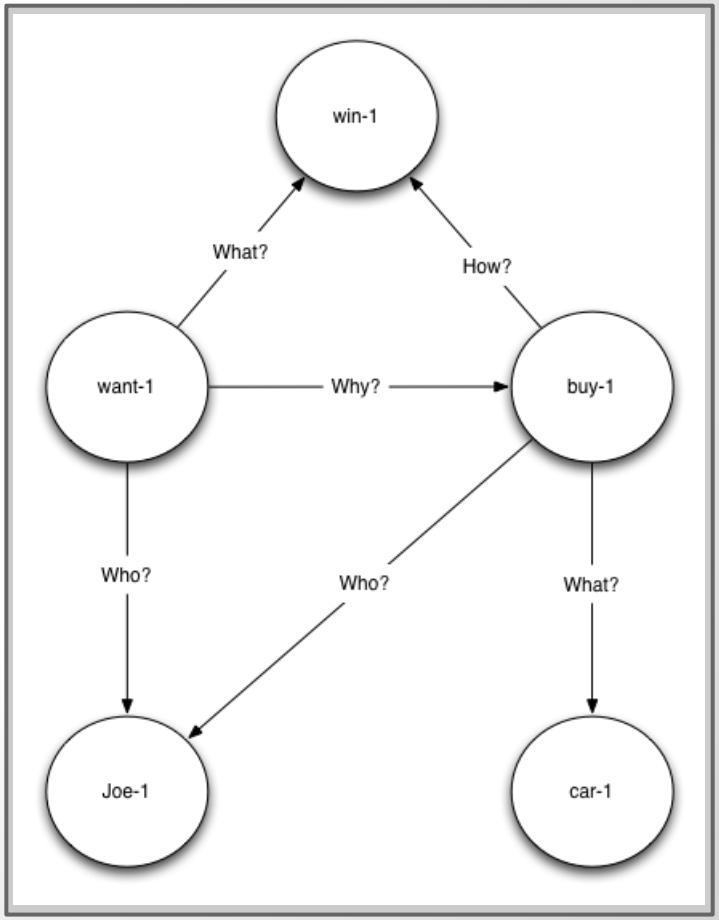
"Joe fixed the faucet with a rubberband."

- fixed-1
  - Who: Joe-1
  - What: faucet-1
  - How: rubberband-1
  - Why: faucet-1.broken=true
  - Effect: faucet-1.broken=false

WIMs – Example The Second

"Joe wanted to win, so he could buy a car."

- want-1
  - Who: Joe-1
  - What: win-1
  - Why: buy-1
- buy-1
  - Who: Joe-1
  - What: car-1
  - How: win-1

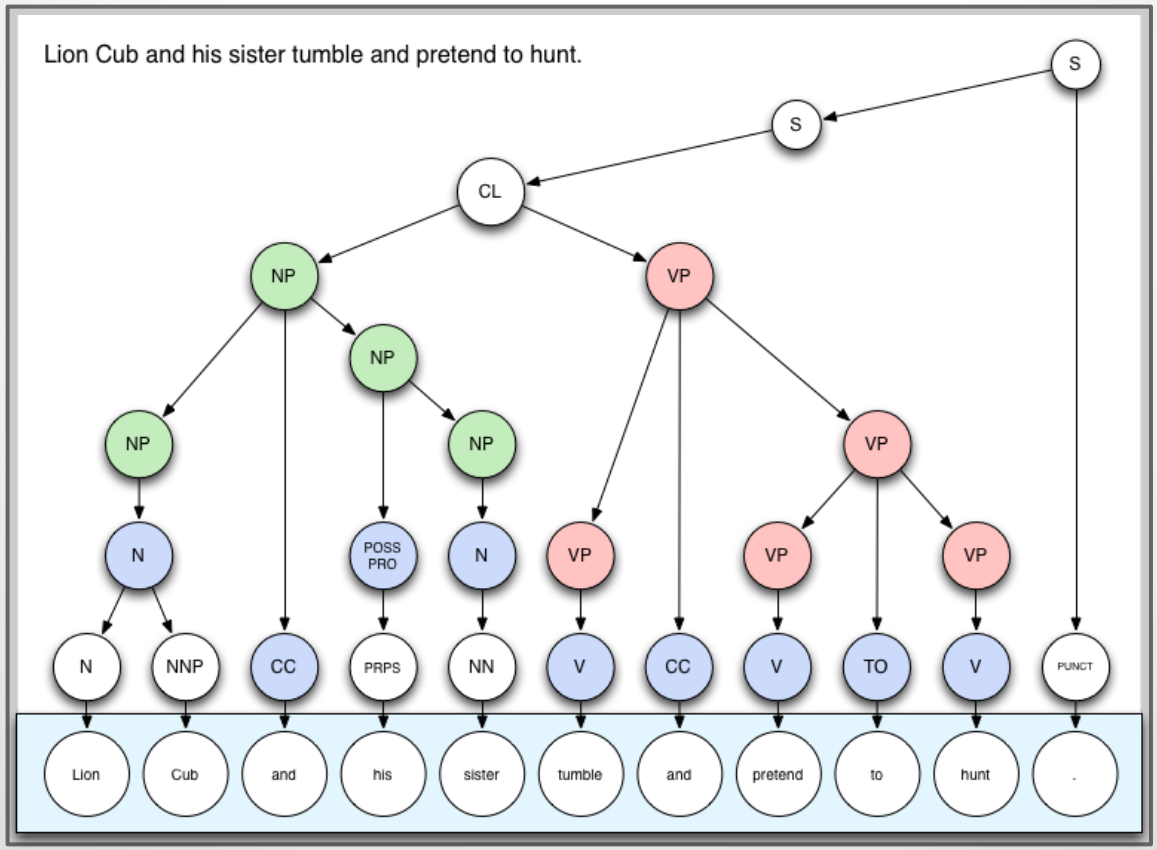


WIMs – From A to Z

1. Start with a syntactic parse
2. Extract the heads from each verb phrase
3. Disambiguate each verb by mapping it to grammar-friendly verb frame template
4. Use the disambiguated verb sense's WIM template, aligned with the verb frame template, to produce a WIM for the verb

Syntactic Parse

- Can use any grammar / parser
- Specifics of the later stages of WIM analysis depend on the grammar but can be shaped to match



Verb Heads

"Joe wanted to win, so he could buy a car."

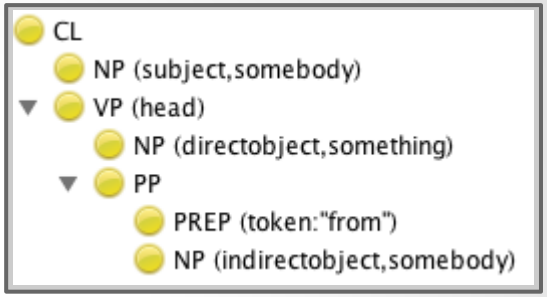
- Identify each verb phrase in the text
- Extract the head of the verb phrase (the active verb)
- Each of these verbs will become a WIM frame after disambiguation

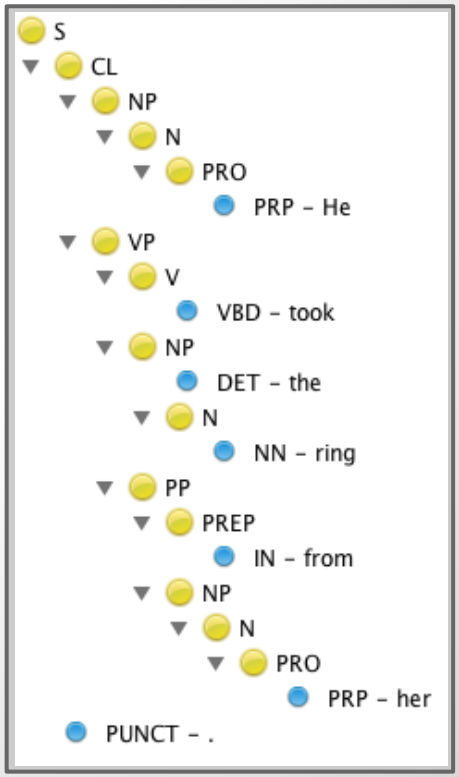
Disambiguation

- A lexicon of verbs, each containing two elements:
  - A grammatical verb phrase mapping
  - A verb phrase → WIM frame template mapping
- Grammatical verb phrase mapping is a partial syntactic tree, with each element being a functional mapping
- WIM frame template is a mapping of verb phrase elements (subject, head, direct object, etc.) to WIM relations (who, what, how, etc.)

Example Verb Frame Mapping

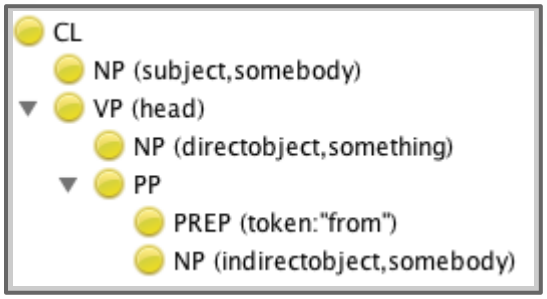
- "He took the ring from her."
  - One verbal sense of take includes this verb frame mapping
  - Mapping is grammar dependent
  - Each parameter is a function that must evaluate to true for the grammatical node to validate
  - Full validation means a match; any failure means the sense is no longer a candidate



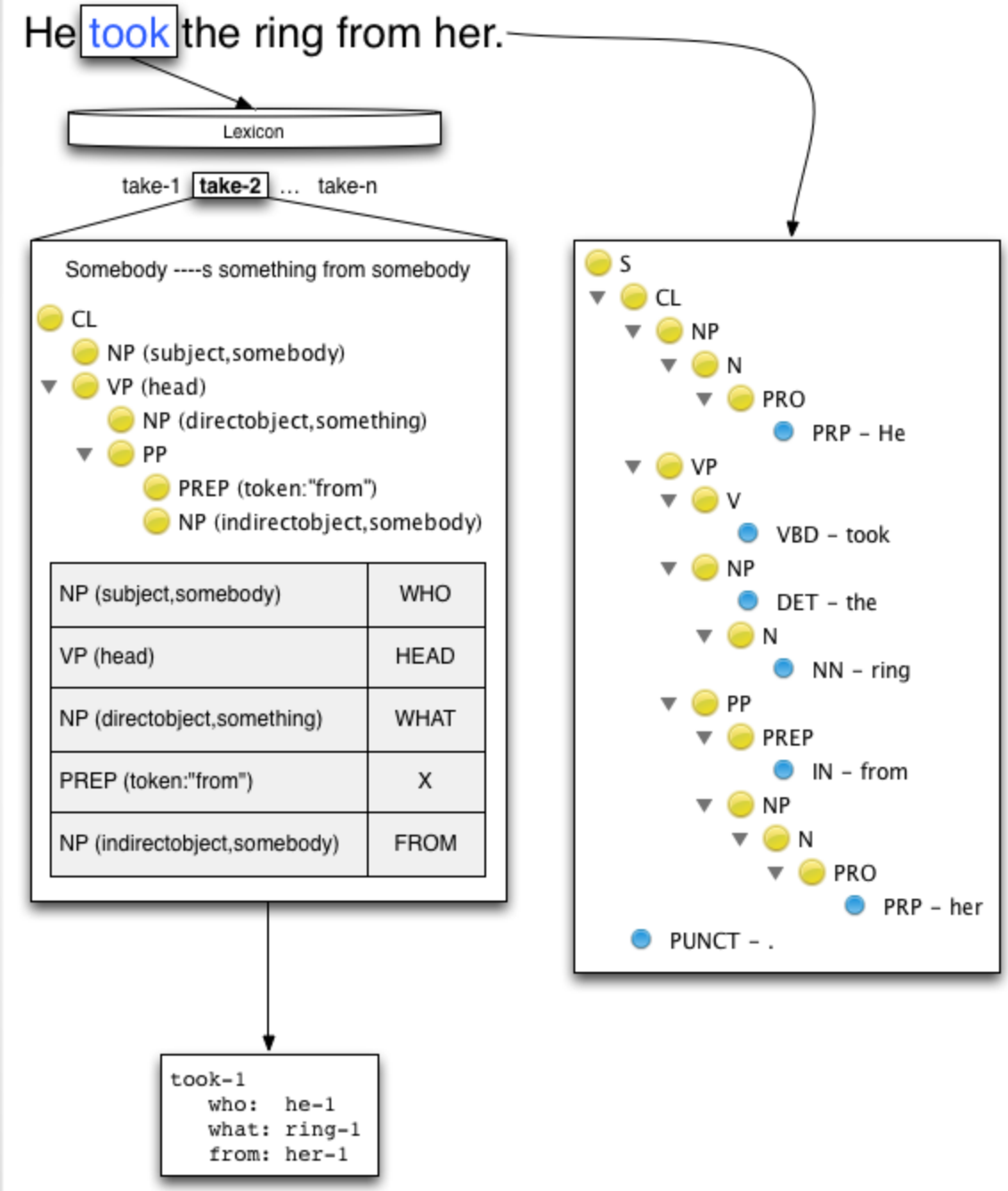


Example WIM Template

"He took the ring from her."

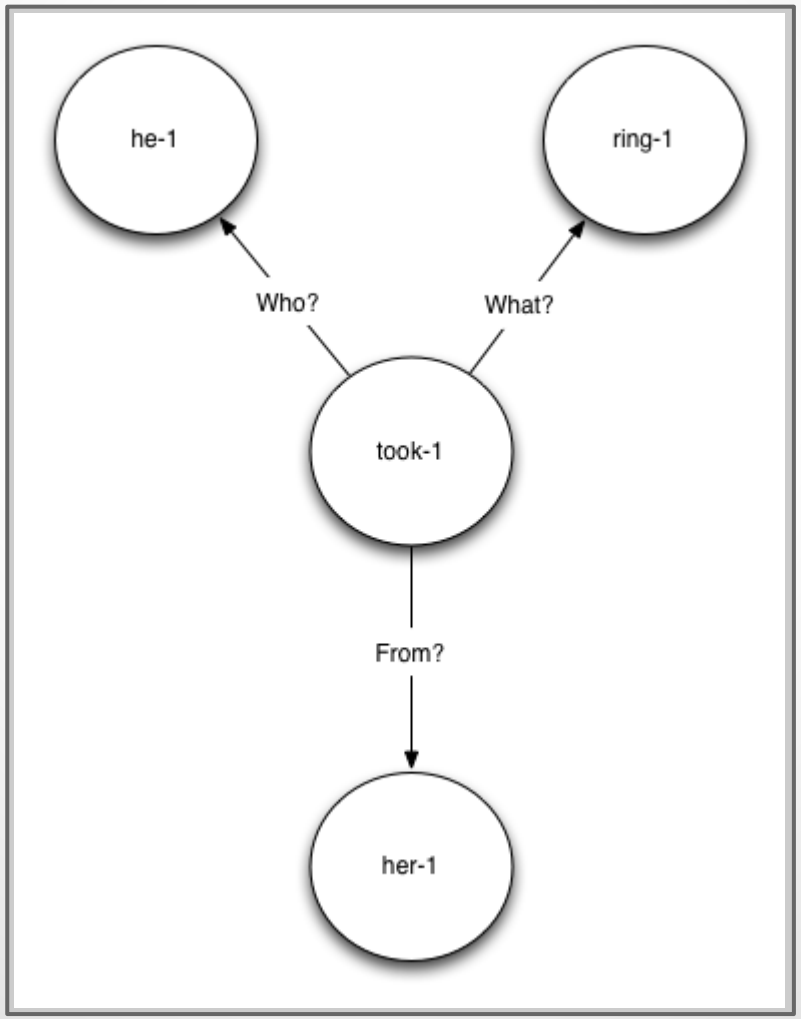


NP (subject,somebody)	WHO
VP (head)	HEAD
NP (directobject,something)	WHAT
PREP (token:"from")	X
NP (indirectobject,somebody)	FROM


He took the ring from her.
Lexicon
take-1 take-2 … take-n
Somebody ----s something from somebody
CL
NP (subject,somebody)
VP (head)
NP (directobject,something)
PP
PREP (token:"from")
NP (indirectobject,somebody)
NP (subject,somebody) WHO
VP (head) HEAD
NP (directobject,something) WHAT
PREP (token:"from") X
NP (indirectobject,somebody) FROM
took-1
who: he-1
what: ring-1
from: her-1
S
CL
NP
N
PRO
PRP - He
VP
V
VBD - took
NP
DET - the
N
NN - ring
PP
PREP
IN - from
NP
N
PRO
PRP - her
PUNCT - .

Example WIM

“He took the ring from her.”



Verb Frame Mapping Acquisition

- Must have a lexicon of verbs that provide both frame maps and WIM templates
- Phase 1:
  - Circumvent the knowledge acquisition bottleneck
  - Use WordNet verb frames
  - Convert set of verb frames to set of frame maps and WIM templates
  - Manual acquisition as needed

WordNet Verb Frame Example

- take (v) “get hold of”
  - Somebody ----s something
  - Somebody ----s somebody
  - Something ----s somebody
  - Something ----s something
- Somebody ----s something
  - WHO? head WHAT?

Going Forward (Phase 2)

- Expand coverage of lexical senses by:
  - Multiple frame mappings per sense
  - More phrasal senses
  - More WIM template relationship coverage
- Open source initiative with distributable python package

Questions?