Computational Semantics

Human Language from a Computational Perspective
May 16, 2018

Motivation

Computers can understand language and interact with us

- Personal assistants
- Intelligence analysis
- Robotics

How do they do it?

Semantics

Semantics = meaning.

Computational semantics: algorithms to build meaning representation and use it to solve semantic tasks

Word sense disambiguation

A word may have more than one sense (meaning).

Which sense of the word is being used?

- the new power **plant** has a production capacity of up to 96 MW
- a **plant**-based diet is still seen as an incredibly hard task to manage

Here both are nouns (NN),

but different parts of speech are possible (wish...)

wordnetweb.princeton.edu/perl/webwn

Resource: WordNet

Synset: collection of words with a specific meaning

1. (n) **plant**, works, industrial plant
2. (n) **plant**, flora, plant life
3. (v) **plant**, set
4. (v) implant, engraft, embed, imbed, **plant**
5. (v) establish, found, **plant**, constitute, institute

wordnetweb.princeton.edu/perl/webwn

Resource: WordNet

Relations between words/synsets

1. (n) **plant**, works, industrial plant
   - (n) building complex, complex
     - (n) structure, construction
       - (n) artifact, artefact
         - (n) whole, unit
           - (n) object, physical object
             - (n) physical entity
               - (n) entity

2. (n) **plant**, flora, plant life
   - (n) organism, being
     - (n) living thing, animate thing
       - (n) whole, unit
         - (n) object, physical object
           - (n) physical entity
             - (n) entity

hypernymy (“is a” relation)

Other possible relations:
meronym (“part of” relation),
pertainym (“of” relation)
entailment (“if-then” relation)

wordnetweb.princeton.edu/perl/webwn

Resource: WordNet

Gloss (definition) for each synset

1. (n) plant, works, industrial plant (**buildings for carrying on industrial labor**)
2. (n) plant, flora, plant life (**(botany) a living organism lacking the power of locomotion**)
3. (v) plant, set (**put or set (seeds, seedlings, or plants) into the ground**)
4. (v) implant, engraft, embed, imbed, plant (**fix or set securely or deeply**)
5. (v) establish, found, plant, constitute, institute (**set up or lay the groundwork for**)

wordnetweb.princeton.edu/perl/webwn

Resource: WordNet

Frequency counts for each synset

1. (n) plant, works, industrial plant (**63**)
2. (n) plant, flora, plant life (**37**)
3. (v) plant, set (**8**)
4. (v) implant, engraft, embed, imbed, plant (**2**)
5. (v) establish, found, plant, constitute, institute (**1**)

WSD algorithm

An algorithm can use WordNet (WN) to determine which sense of a word is used in a given sentence.

WSD algorithm

Input: a list of tokens (sentence), target word, WordNet

Output: sense number (according to WN) for target word

wsd(this plant is in the industrial zone, plant, WN) = 1

wsd(a plant is a living thing, plant, WN) = 2

Lesk algorithm for WSD

1. For each sense, get the gloss (definition).
2. Count how many words in the sentence are also in the gloss (overlap).
3. If all senses have a zero overlap, return the sense with the highest frequency.
4. Otherwise, return the sense with the highest overlap.

Lesk algorithm for WSD

```
lesk(L, w, WN):                ▷ L: list of tokens, w: target word
    S ← synsets(WN, w)
    s ← argmax(freq(S))        ▷ get sense with highest frequency
    m ← 0
    i ← 1
    while i ≤ len(S):          ▷ find sense with highest overlap with L
        if overlap(L, gloss(S[i])) > m:
            m ← overlap(L, gloss(S[i]))
            s ←S[i]
        i ← i + 1
    return s
```

Lesk algorithm for WSD

number	gloss	frequency
1	buildings for carrying on **industrial** labor	63
2	(botany) a living organism lacking **the** power of locomotion	37
3	put or set (seeds, seedlings, or plants) into **the** ground	8
4	**plant** (fix or set securely or deeply)	2
5	set up or lay **the** groundwork for	1

lesk(THIS PLANT IS IN THE INDUSTRIAL ZONE, PLANT, WN) = 1 ✓

Lesk algorithm for WSD

number	gloss	frequency
1	buildings for carrying on industrial labor	63
2	(botany) **a living** organism lacking the power of locomotion	37
3	put or set (seeds, seedlings, or plants) into the ground	8
4	**plant** (fix or set securely or deeply)	2
5	set up or lay the groundwork for	1

lesk(A PLANT IS A LIVING THING, PLANT, WN) = 2 ✓

Lesk algorithm for WSD

number	gloss	frequency
1	buildings for carrying on industrial labor	63
2	(botany) **a** living organism lacking the **power** of locomotion	37
3	put or set (seeds, seedlings, or plants) into the ground	8
4	**plant** (fix or set securely or deeply)	2
5	set up or lay the groundwork for	1

lesk(SHE WORKS IN A POWER PLANT, PLANT, WN) = 2

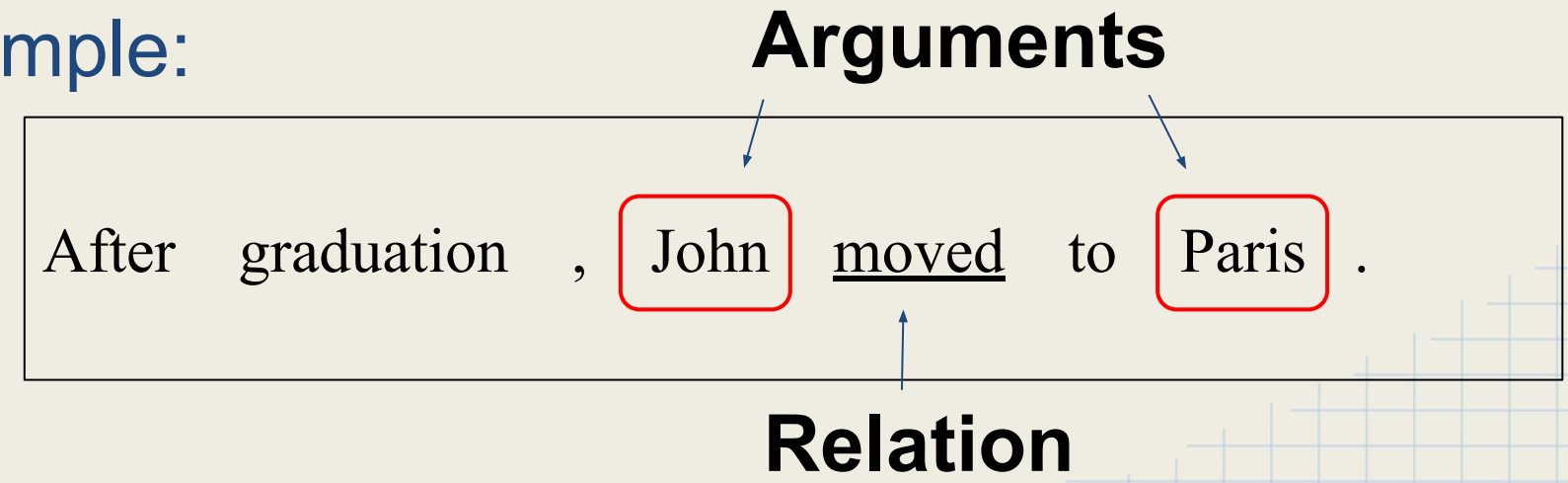
There are better algorithms today.

Semantic role labeling

What are the semantic relations within the sentence?

(Specifically, predicates and their arguments.)

Example:



propbank.github.io

Resource: PropBank

Roleset: collection of relations with a specific meaning, and a definition of **numbered** roles

give.01 , ***transfer***

Roles:

Arg0: *giver*

Arg1: *thing given*

Arg2: *entity given to*

give.17 , ***yield when pressure is applied***

Roles:

Arg1: *thing yielding*

Arg2: *to what*

propbank.github.io

Resource: PropBank

Examples:

The PLO failed to win membership in the World Health Organization.
Arg0 **Arg1**

Wall Street's old guard seems to be winning the program-trading battle.
Arg0 **Arg2**

She won grant money for the school.
Arg0 **Arg1** **Arg3**

A wine that wins high ratings from the critics will eventually move.
Arg0 **Arg1** **Arg4**

He won D.T.'s queen for two minor pieces and two pawns.
Arg0 **Arg1** **Arg5**

win.01 , ***win, win a prize/contest***

Roles:

Arg0: *winner*

Arg1: *prize*

Arg2: *contest*

Arg3: *beneficiary*

Arg4: *loser, giver of prize*

Arg5: *in exchange for*

SRL algorithm

An algorithm can use PropBank (PB) to determine which phrases in the sentence correspond to which arguments.

Let’s assume we already know which word in the sentence is the relation, and which sense (that is, roleset) is used…

SRL algorithm

Input: a list of tokens, target word (relation), sense, phrase structure of the sentence, PB

Output: list of phrases and their argument number

srl(SHE GAVE HIM THE KEY, GAVE, 1, [SHE: NP, HIM: NP, THE KEY: NP], PB) =

[SHE: ARG0, HIM: ARG1, THE KEY: ARG2]

Simple SRL algorithm

Informal description:

Mark each noun phrase (NP) or prepositional phrase (PP) as an argument,

according to the order in the sentence and the order of argument definitions.

Simple SRL algorithm

srl(SHE GAVE THE KEY TO HIM, GAVE, 1, [SHE: NP, THE KEY: NP, TO HIM: PP], PB)

SHE GAVE THE KEY TO HIM

NP relation NP PP

SHE GAVE THE KEY TO HIM

ARG0 ARG1 ARG2

give.01 , ***transfer***

Roles:

Arg0: *giver*

Arg1: *thing given*

Arg2: *entity given to*

[SHE: ARG0, THE KEY: ARG1, TO HIM: ARG2] ✓

Simple SRL algorithm

srl(SHE WAS GIVEN THE KEY BY HIM, GIVEN, 1, [SHE: NP, THE KEY: NP, BY HIM: PP], PB)

SHE WAS GIVEN THE KEY BY HIM
NP relation NP PP

SHE WAS GIVEN THE KEY BY HIM
ARG0 ARG1 ARG2

give.01 , ***transfer***

Roles:

Arg0: *giver*

Arg1: *thing given*

Arg2: *entity given to*

[SHE: ARG0, THE KEY: ARG1, BY HIM: ARG2]

It should be: [SHE: ARG2, THE KEY: ARG1, BY HIM: ARG0]

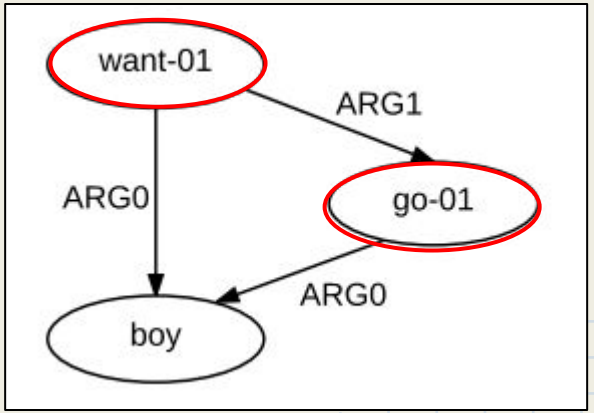
SRL algorithm

We can improve the algorithm by adding rules for passive verbs and so on, but there are better methods today.

Abstract Meaning Representation

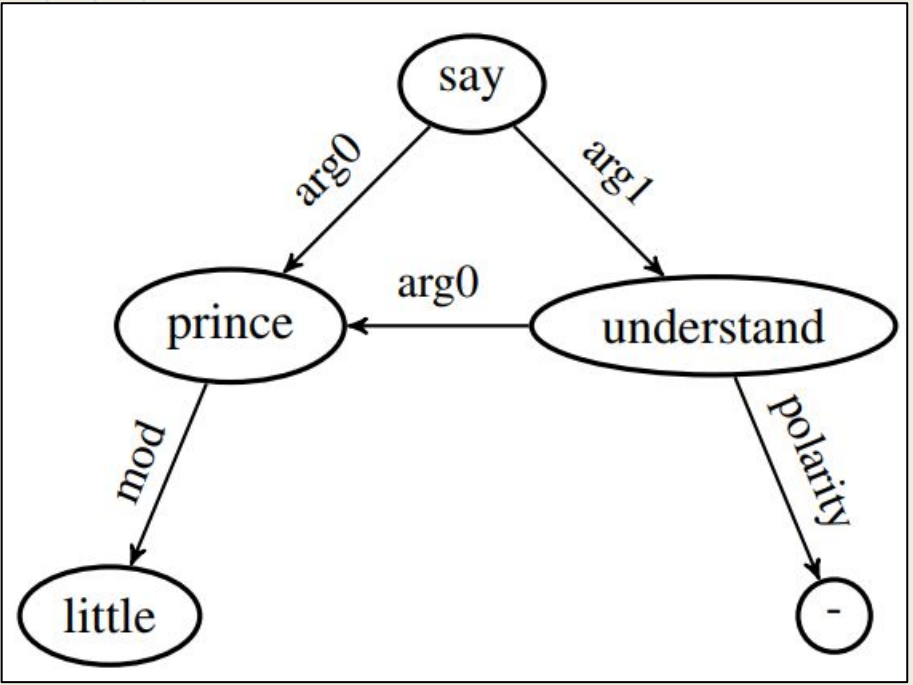
Convert the sentence into an abstract graph that describes all entities and relations

The boy wants to go

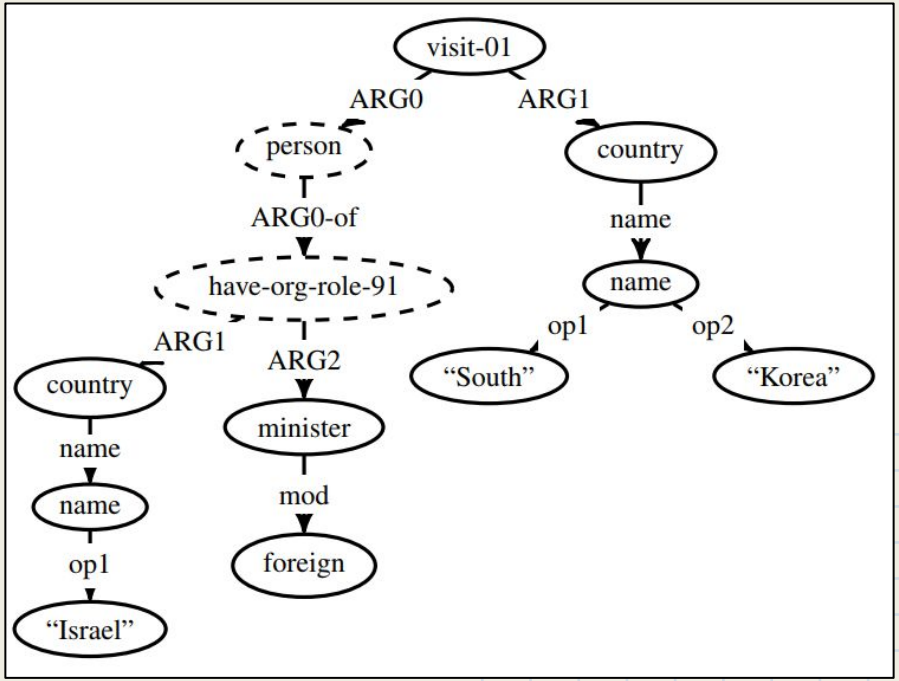


Using PropBank for relation and role definitions

Abstract Meaning Representation



“I do not understand”, said the little prince.

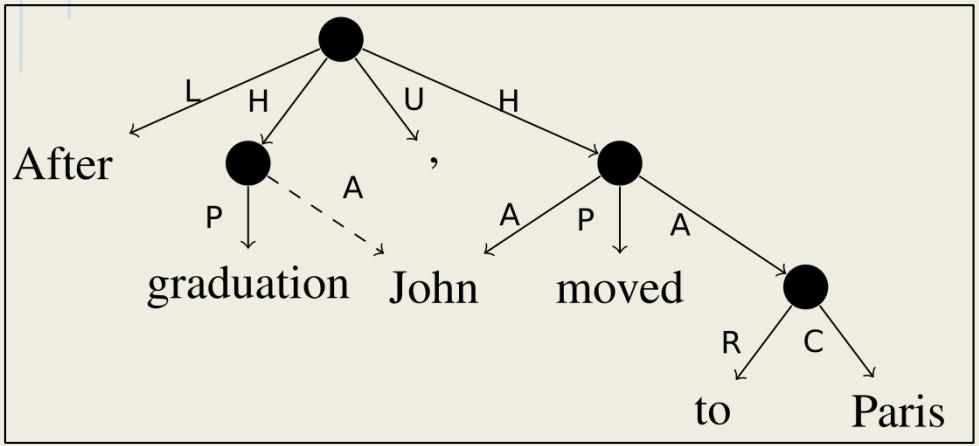


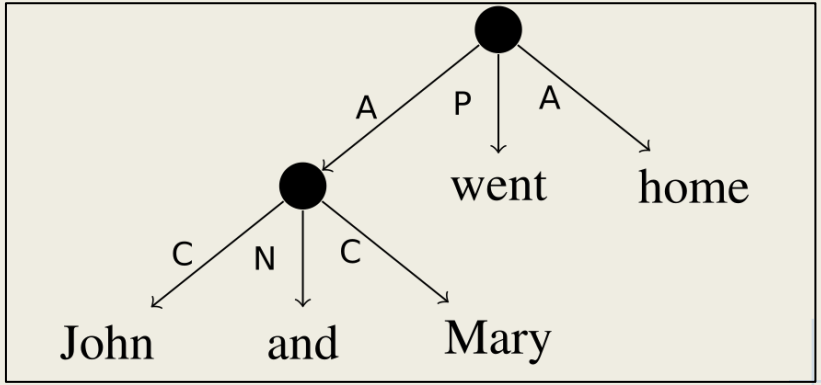
Israel foreign minister visits South Korea.

www.cs.huji.ac.il/~oabend/ucca.html

UCCA

Universal Conceptual Cognitive Annotation





H	parallel scene	D	adverbial	N	connector	U	punctuation
A	participant	E	elaborator	P	process	LR	link relation
C	center	G	ground	R	relator	LA	link argument
L	linker	S	state	F	function		

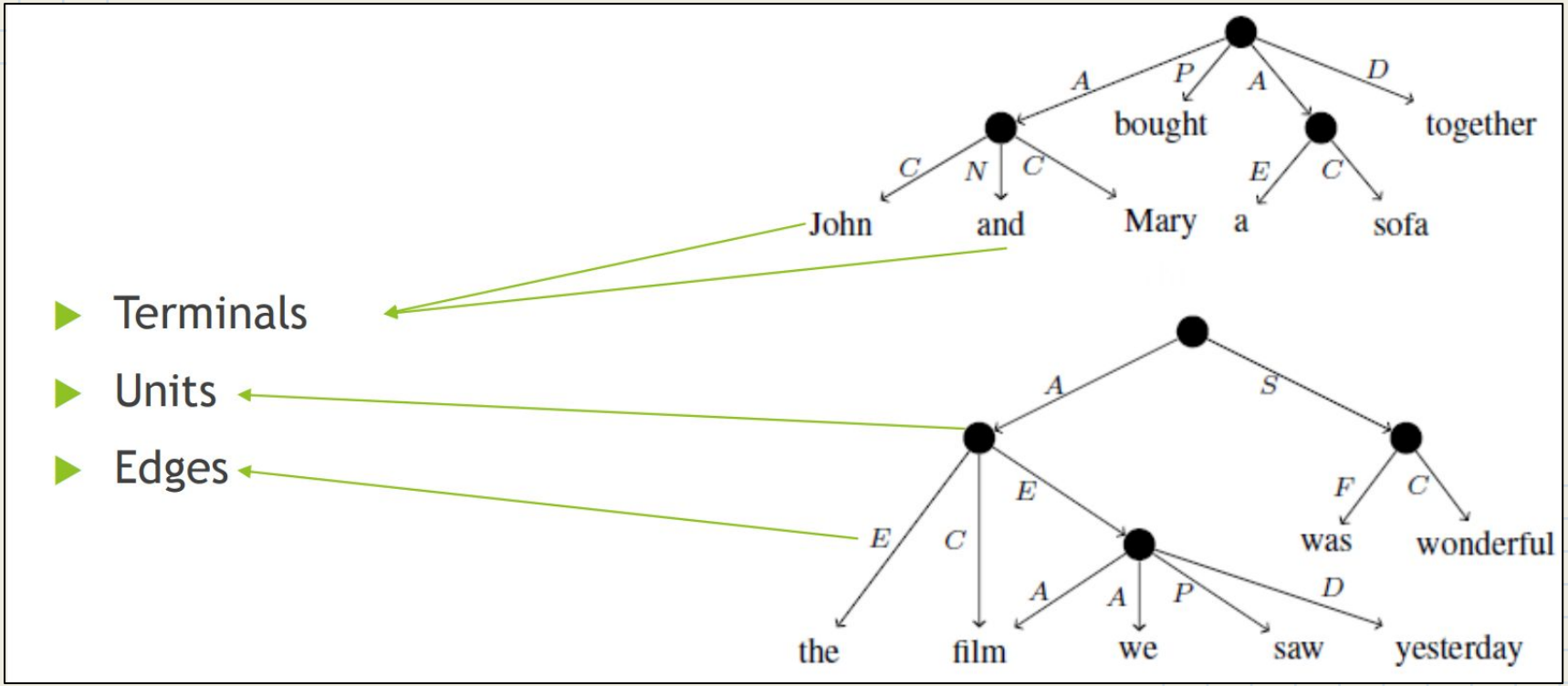
www.cs.huji.ac.il/~oabend/ucca.html

Universal Conceptual Cognitive Annotation

Abb.	Category	Short Definition
		Scene Elements
P	**Process**	The main relation of a Scene that evolves in time (usually an action or movement).
S	**State**	The main relation of a Scene that does not evolve in time.
A	**Participant**	A participant in a Scene in a broad sense (including locations, abstract entities and Scenes serving as arguments).
D	**Adverbial**	A secondary relation in a Scene (including temporal relations).
		Elements of Non-Scene Units
C	**Center**	Necessary for the conceptualization of the parent unit.
E	**Elaborator**	A non-Scene relation which applies to a single Center.
N	**Connector**	A non-Scene relation which applies to two or more Centers, highlighting a common feature.
R	**Relator**	All other types of non-Scene relations. Two varieties: (1) Rs that relate a C to some super-ordinate relation, and (2) Rs that relate two Cs pertaining to different aspects of the parent unit.
		Inter-Scene Relations
H	**Parallel Scene**	A Scene linked to other Scenes by regular linkage (e.g., temporal, logical, purposive).
L	**Linker**	A relation between two or more Hs (e.g., "when", "if", "in order to").
G	**Ground**	A relation between the speech event and the uttered Scene (e.g., "surprisingly", "in my opinion").
		Other
F	**Function**	Does not introduce a relation or participant. Required by the structural pattern it appears in.

www.cs.huji.ac.il/~oabend/ucca.html

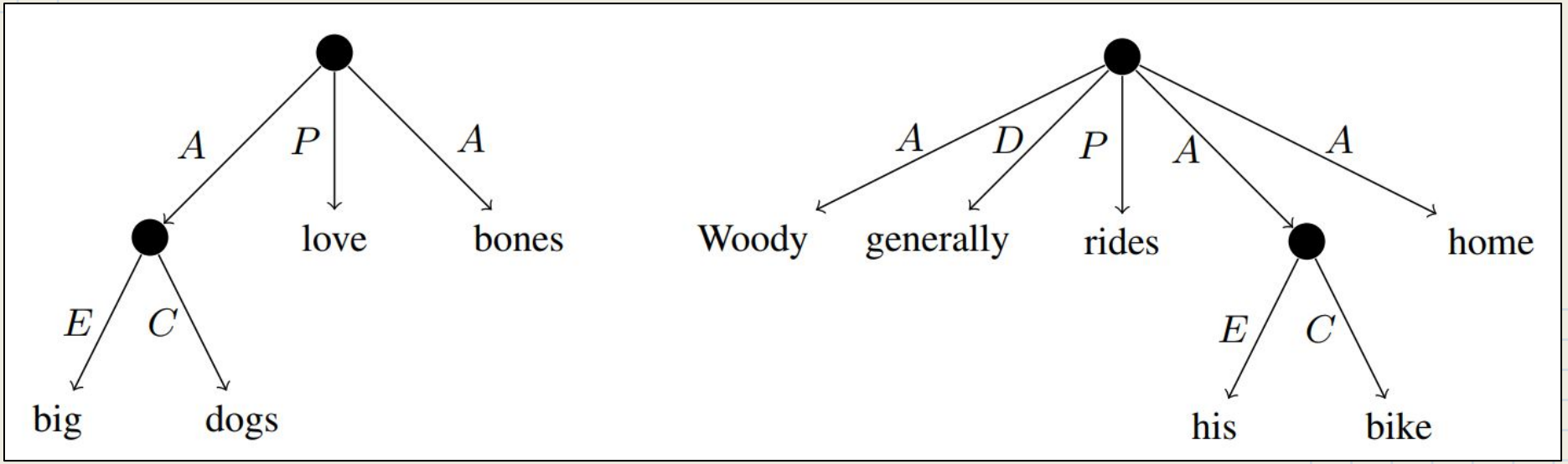
Universal Conceptual Cognitive Annotation



- Terminals
- Units
- Edges

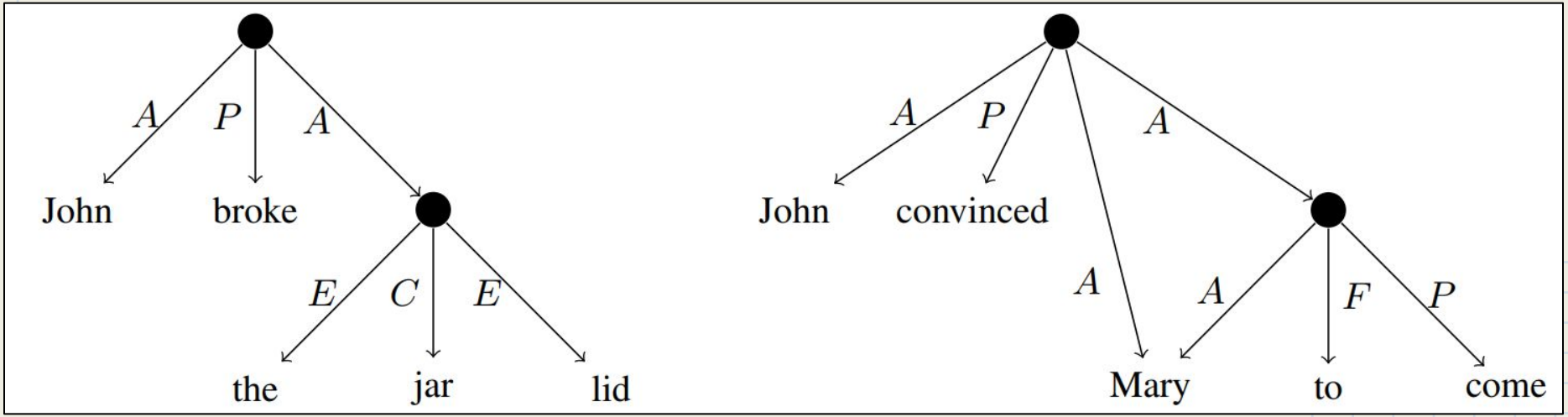
www.cs.huji.ac.il/~oabend/ucca.html

Universal Conceptual Cognitive Annotation



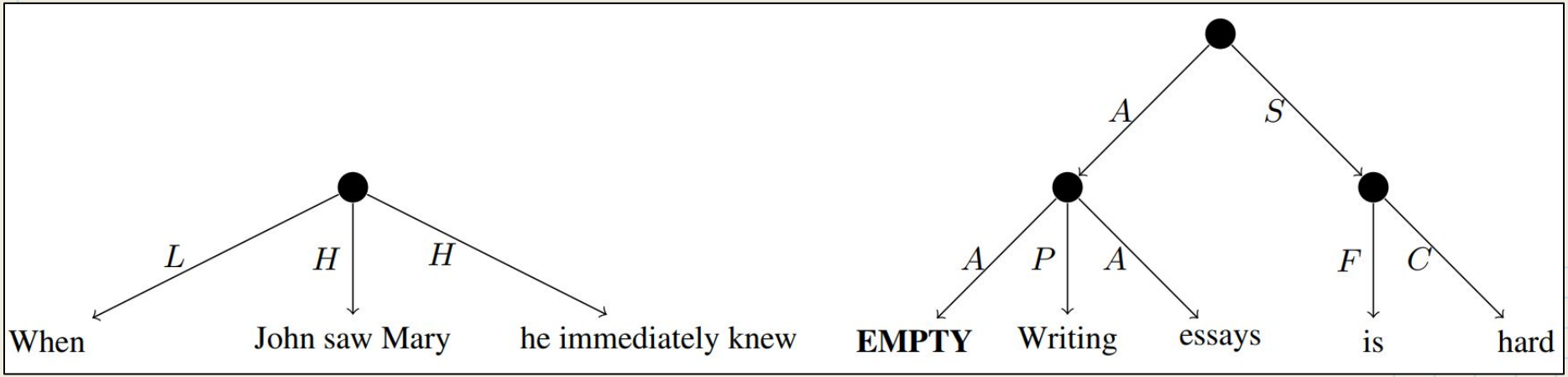
www.cs.huji.ac.il/~oabend/ucca.html

Universal Conceptual Cognitive Annotation



www.cs.huji.ac.il/~oabend/ucca.html

Universal Conceptual Cognitive Annotation



www.cs.huji.ac.il/~oabend/ucca.html

Universal Conceptual Cognitive Annotation