Natural Language Processing

Lecture 18a: Meaning Representation Languages

Semantics Road Map

1. Lexical semantics
2. Vector semantics
3. **Meaning representation languages** and **semantic roles**
4. Compositional semantics, semantic parsing
5. Discourse and pragmatics

INTENSION AND EXTENSION

Two Approaches to Semantics

- Intentional
  - Assumes that the word or utterance is intrinsically meaningful
  - Decompositional approaches to lexical semantics are intentional
- Extentional
  - Defines words and utterances by the things in the world of which they are true
  - This lecture will concern extentional models of semantics

Extension

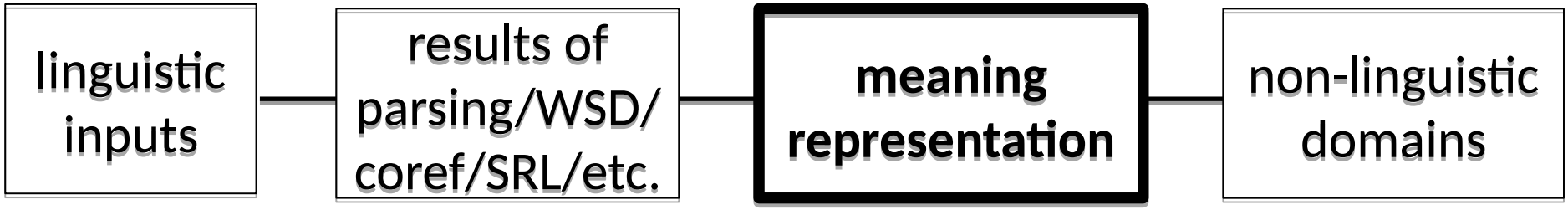
The meaning of *red* is the set of entities in the universe of which the predicate RED is true. Similarly, the meaning of *hit* is the set of <x,y> pairs of which HIT(x, y) is true.

In this lecture…

- We will look at ways of representing the extension of verbs and sentences
- We will also look at semantic roles and how they relate to meaning representation languages (MRLs)

DESIRABLE PROPERTIES OF MEANING REPRESENTATIONS

Meaning Representation?



For what kinds of tasks?

- Answering essay questions on an exam
- Deciding what to order at a restaurant
- Learning an activity via instructions
- Making an investment decision
- Recognizing an insult

Desirable Qualities: Verifiability

We want to be able to determine the truth of our representations.

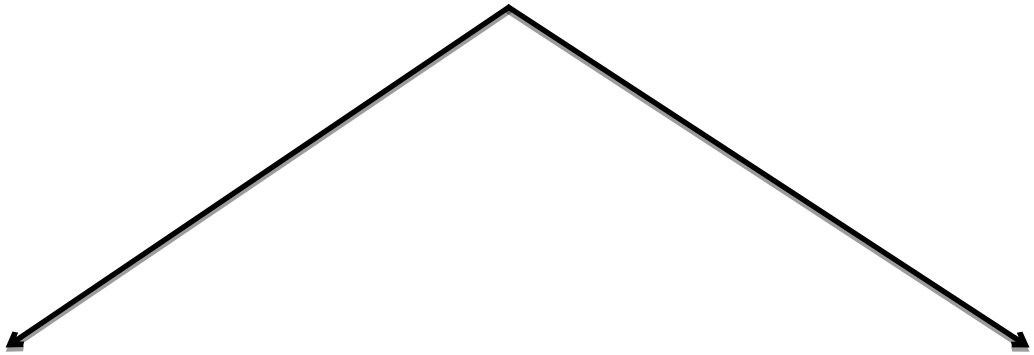
"Does Udipi serve vegetarian food"?

Is SERVE(Udipi, vegetarian food) in our knowledge base?

What is the relationship between the meaning of a sentence and the world as we know it?

Desirable Qualities: Unambiguous Representation

Let's eat somewhere near campus.



(e.g., we want to eat *at* a place near campus)

(e.g., we eat places)

Our MRL must capture precisely one of these meanings—not both.

Desirable Qualities: Canonical Form

- “Mad Mex has vegetarian dishes.”
- “They have vegetarian food at Mad Mex.”
- “Vegetarian dishes are served at Mad Mex.”
- “Mad Mex serves vegetarian fare.”

Inputs that mean the same thing should have the same meaning representation.

Desirable Qualities: Inference, Variables, and Expressiveness

- “Can vegetarians eat at Mad Mex?”

- “I’d like to find a restaurant where I can get vegetarian food.”

  SERVE(x, vegetarian food)

- “Delta flies Boeing 737s from Boston to New York.”

One Limitation: Literality

We will focus on the **basic** requirements for meaning representation.

The basic requirements do not include correctly interpreting statements like:

- "Ford was hemorrhaging money."
- "I could eat a horse."

What entities do we want to represent?

A meaning representation scheme should let us represent:

- **objects** (e.g., people, restaurants, cuisines)
- **properties of objects** (e.g., pickiness, noisiness, spiciness)
- **relations between objects** (e.g., SERVE(Oishii Bento, Japanese))

The Knowledge Base

It contains the things that we "know"

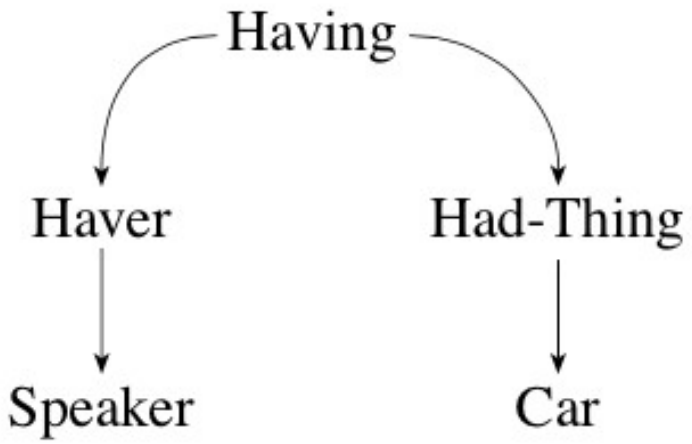
We can query it

Our knowledge base

THE CANDIDATES

“I have a car.”

$$\exists x, y Having(x) \wedge Haver(Speaker, x) \wedge HadThing(y, x) \wedge Car(y)$$



Car
⇑ POSS-BY
Speaker

```
Having
 Haver:   Speaker
 HadThing:  Car
```

FIRST-ORDER LOGIC

MRL #1: First-Order Logic

DressCode(ThePorch)

Cuisine(Udipi)

Functions

SERVES(UnionGrill, AmericanFood)

RESTAURANT(UnionGrill)

Predicates

- HAVE(Speaker, FiveDollars) ∧ ¬ HAVE(Speaker, LotOfTime)
- ∀x PERSON(x) ⇒ HAVE(x, FiveDollars)
- ∃x,y PERSON(x) ∧ RESTAURANT(y) ∧ ¬HASVISITED(x,y)

First Order Logic and Semantics

- Nouns correspond to one-place predicates:

  RESTAURANT(x) is true if x is a member of the set of restaurants

- Adjectives correspond to one-place predicates:

  VEGETARIAN(x) is true if x is a member of the set of things that are vegetarian

- Verbs correspond to one-place, two-place, or three-place predicates

  DINE(x) as in *Noah dined*.

  EAT(x, y) as in *Noah ate American food*.

  GIVE(x, y, z) as in *The bad sushi gave Noah a stomach ache*.

Modus Ponens and Forward Chaining

$$\frac{\textsc{vegetarianrestaurant}(\text{Udipi}) \qquad \forall x\, \textsc{vegetarianrestaurant}(x) \Rightarrow \textsc{serves}(x, \text{VegetarianFood})}{\textsc{serves}(\text{Udipi}, \text{VegetarianFood})}$$

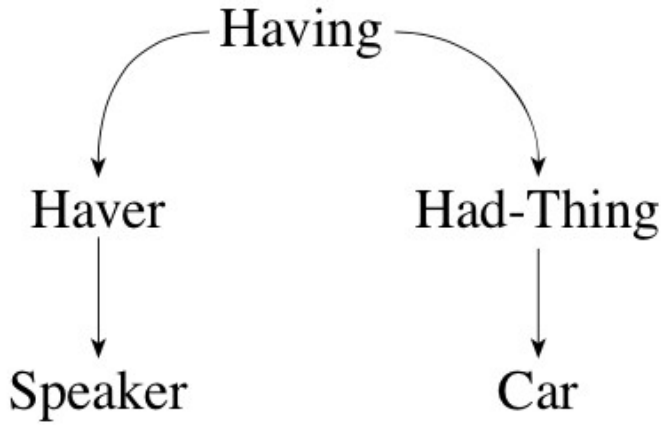
As individual facts are added to a knowledge base, modus ponens can be used to fire applicable implication rules.

First Order Logic: Advantages

- Flexible
- Well-understood
- Widely used

DESCRIPTION LOGICS

MRL #2: Description Logics




```
Having
 Haver:   Speaker
 HadThing:   Car
```

- Goal of description logics: understand and specify semantics for slot-filler representations
- More restrictive than FOL

TBox and ABox

- TBox: contains the knowledge about categories or concepts in the application domain

  All bistros are restaurants

  All restaurants are businesses
- ABox: facts about individuals in the domain

  Udipi is an Indian restaurant

Categories and Subsumption

IndianRestaurant(Udipi)

category domain element

Udipi is an Indian restaurant.

IndianRestaurant $\sqsubseteq$ Restaurant

subsumed subsumer

All Indian restaurants are restaurants.

Negation and Disjunction

IndianRestaurant ⊑ **not** ItalianRestaurant

Indian restaurants can't also be Italian restaurants.

Restaurant ⊑ (**or** ItalianRestaurant IndianRestaurant MexicanRestaurant)

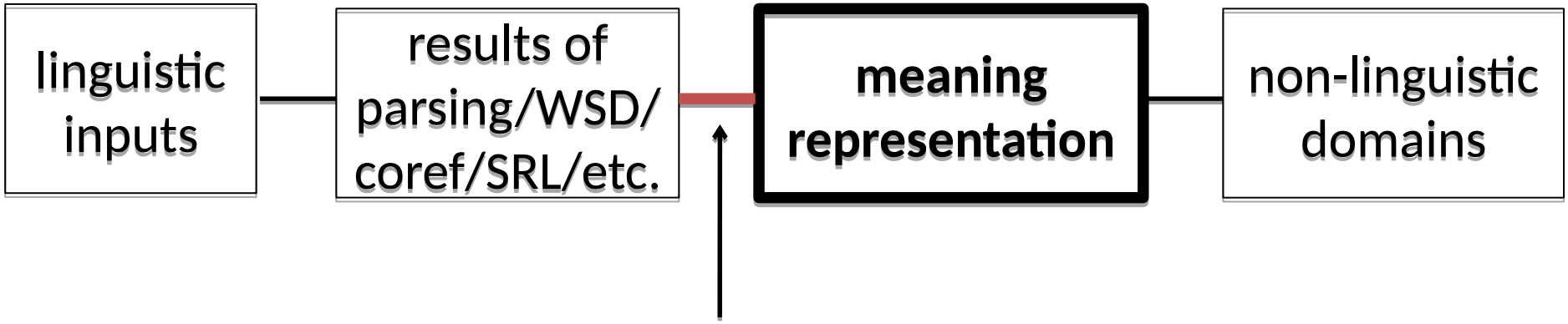
Restaurants are Italian restaurants, Indian restaurants, or Mexican restaurant.

Advantages

- Intuitive hierarchical representation
- Compatible with existing work on ontologies

LOOKING FORWARD

The Missing Link



Compositional semantics / semantic parsing

Natural Language Processing

Lecture 18 part b: Semantic Roles

Semantics Roadmap

- You should already have been convinced that grammatical structure is an important aspect of language
- Now we are discussing *semantics* or *meaning*
- Up until today, we have talked about meaning as something that individual words have (whether in isolation or in context)
- So far today, we have talked about representing the meanings of propositions/sentences in meaning representation languages
- Now, we are going to discuss an enhancement to this view, the notion that individual noun phrases can be characterized as having **roles** relative to a **predicate** or **frame**

- Noah built an ark out of gopher wood.
- He loaded two of every animal onto the ark.
- Noah piloted the ark into stormy weather.
- When the skies cleared, all rejoiced.

- Noah$_1$ built an ark$_2$ out of gopher wood.
- He$_1$ loaded two of every animal onto the ark$_2$.
- Noah$_1$ piloted the ark$_2$ into stormy weather.
- When the skies$_3$ cleared, all$_4$ rejoiced.

Paraphrase

- Noah built an ark out of gopher wood.
- An ark was built by Noah. It was made from gopher wood.
- Noah constructed an ark with wood from a gopher tree.
- Using gopher wood, Noah managed to put together an ark.
- Noah built an ark.
- …

Traditional Semantic Roles

- In the linguistics literature, one sees a number of common terms for semantic roles
  - Agent
  - Patient
  - Theme
  - Force
  - Experiencer
  - Stimulus
  - Recipient
  - Source
  - Goal
  - etc.
- These have their place, and are useful to know if you want to understand what a semantic role is, but are not widely used in NLP
- In NLP, we tend to use finer-grained (and sometimes cryptically named) semantic role labels

Traditional Semantic Roles

- **David** *threw* **the midterms** from **Pausch Bridge** to **the hillside below**.
  - **David**—agent
  - **the midterms**—theme
  - **Pausch Bridge**—source
  - **the hillside below**—goal

Neo-Davidsonian Representation

- David *threw* the midterms from Pausch Bridge to the hillside below
  - THROW(David, midterms, PauschBridge, hillside)
  - $\exists e$ THROW(e) $\wedge$ AGENT(e, David) $\wedge$ THEME(e, midterms) $\wedge$ SOURCE(e, PauschBridge) $\wedge$ GOAL(e, hillside)
- The midterms were *thrown* from Pausch Bridge
  - THROW(midterms, PauschBridge)
  - $\exists e$ THROW(e) $\wedge$ THEME(e, midterms) $\wedge$ SOURCE(e, PauschBridge)

Semantic Role Labeling

Input: a sentence, paragraph, or document

Output: for each predicate*, labeled spans identifying each of its arguments.

*Predicates are sometimes identified in the input, sometimes not.

Predicates

- Noah built an ark out of gopher wood.
- An ark was built by Noah. It was made from gopher wood.
- Noah constructed an ark with wood from a gopher tree.
- Using gopher wood, Noah managed to put together an ark.

Predicates and Arguments

- Noah built an ark out of gopher wood.
- An ark was built by Noah. It was made from gopher wood.
- Noah constructed an ark with wood from a gopher tree.
- Using gopher wood, Noah managed to put together an ark.

Breaking, Eating, Opening

- John broke the window.
- The window broke.
- John is always breaking things.
- The broken window testified to John's malfeasance.

- Eat!
- We ate dinner.
- We already ate.
- The pies were eaten up quickly.
- Our gluttony was complete.

- Open up!
- Someone left the door open.
- John opens the window at night.

Introducing PropBank

- Corpus (PTB) with propositions annotated
  - Predicates (verbs)
  - Arguments (semantic roles)
- Semantic roles are Arg0, Arg1, etc., each with a description
  - Arg0 is typically the most agent-like argument
  - Labels for other arguments are somewhat arbitrary

“Agree” in PropBank

- arg0: agreer
- arg1: proposition
- arg2: other entity agreeing
- The group agreed it wouldn’t make an offer.
- Usually John agrees with Mary on everything

“Fall (move downward)” in PropBank

- arg1: logical subject, patient, thing falling
- arg2: extent, amount fallen
- arg3: starting point
- arg4: ending point
- argM-loc: medium
- Sales fell to $251.2 million from $278.8 million.
- The average junk bond fell by 4.2%.
- The meteor fell through the atmosphere, crashing into Cambridge.

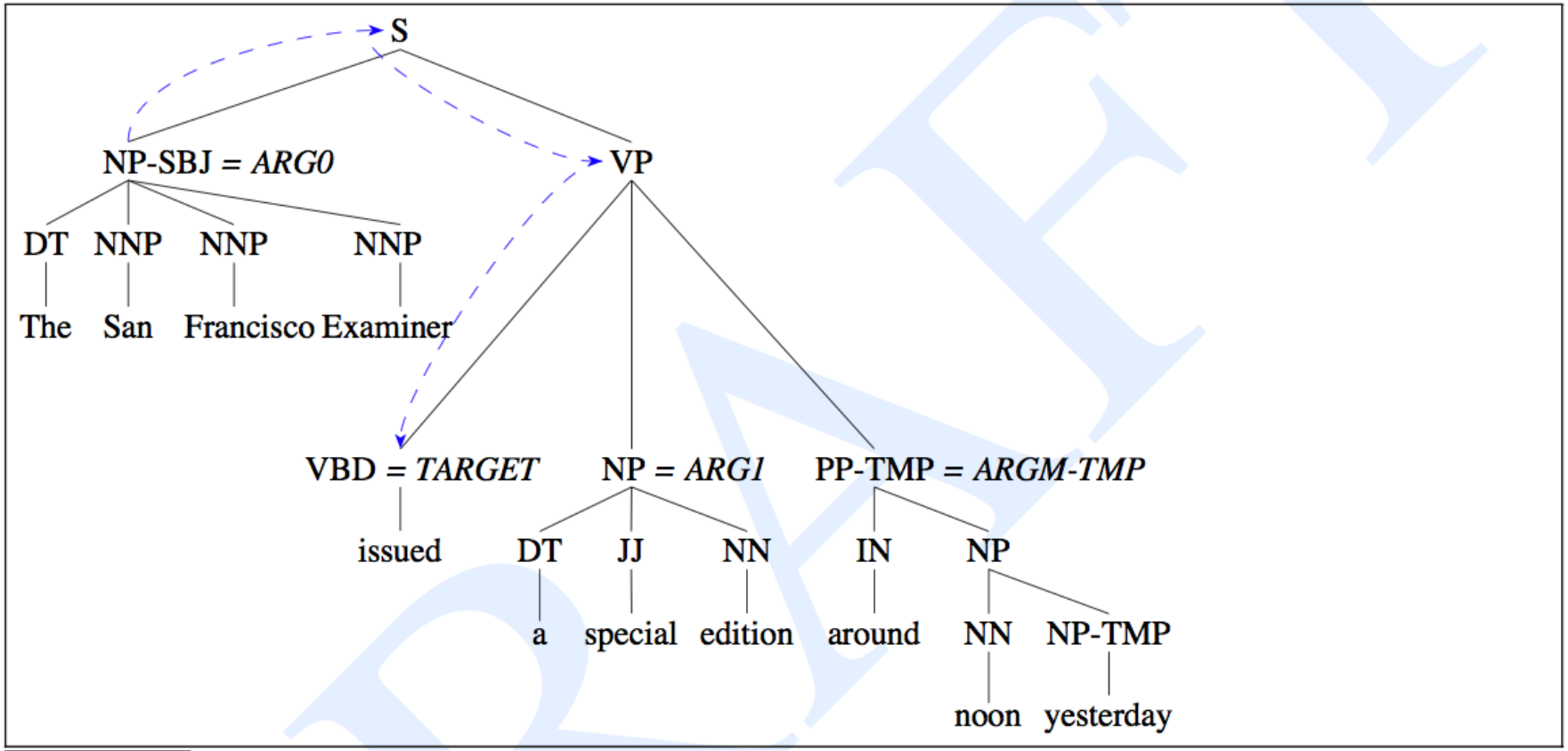


Figure 20.16 Parse tree for a PropBank sentence, showing the PropBank argument labels. The dotted line shows the **path** feature NP↑S↓VP↓VBD for ARG0, the NP-SBJ constituent *the San Francisco Examiner*.

FrameNet

- A **frame** is a schematic representation of a situation involving various participants, and other conceptual roles
- In FrameNet, frames—not verbs—are first-class citizens
  - To a first approximation, verbs that relate to the same situation belong to the same frame
  - Roles are given fine-grained labels that are specific to the frame, but not the verb
  - Frames can center around words other than verbs

change_position_on_a_scale

Core roles	
ATTRIBUTE	scalar property that the ITEM possesses
DIFFERENCE	distance by which an ITEM changes its position
FINAL_STATE	ITEM's state after the change
FINAL_VALUE	position on the scale where ITEM ends up
INITIAL_STATE	ITEM's state before the change
INITIAL_VALUE	position on the scale from which the ITEM moves
ITEM	entity that has a position on the scale
VALUE_RANGE	portion of the scale along which values of ATTRIBUTE fluctuate
Some non-core roles …	
DURATION	length of time over which the change occurs
SPEED	rate of change of the value
GROUP	the group in which an ITEM changes the value of an ATTRIBUTE

- **Verbs**: advance, climb, decline, decrease, diminish, dip, double, drop, dwindle, edge, explode, fall, fluctuate, gain, grow, increase, jump, move, mushroom, plummet, reach, rise, rocket, shift, skyrocket, slide, soar, swell, swing, triple, tumble
- **Nouns**: decline, decrease, escalation, explosion, fall, fluctuation, gain, growth, hike, increase, rise, shift, tumble
- **Adverb**: increasingly

Demo

https://framenet.icsi.berkeley.edu/fndrupal/

How Can We Build an SRL System?

(1) Parse

(2) For each predicate word in the parse:

For each node in the parse:

Classify the node with respect to the predicate