

Intelligent Systems (AI-2)

Computer Science cpssc422, Lecture 23

Oct 30, 2019

Slide credit: Probase Microsoft Research Asia, YAGO Max Planck
Institute, National Lib. Of Medicine, NIH

To summarize: Truth in first-order logic

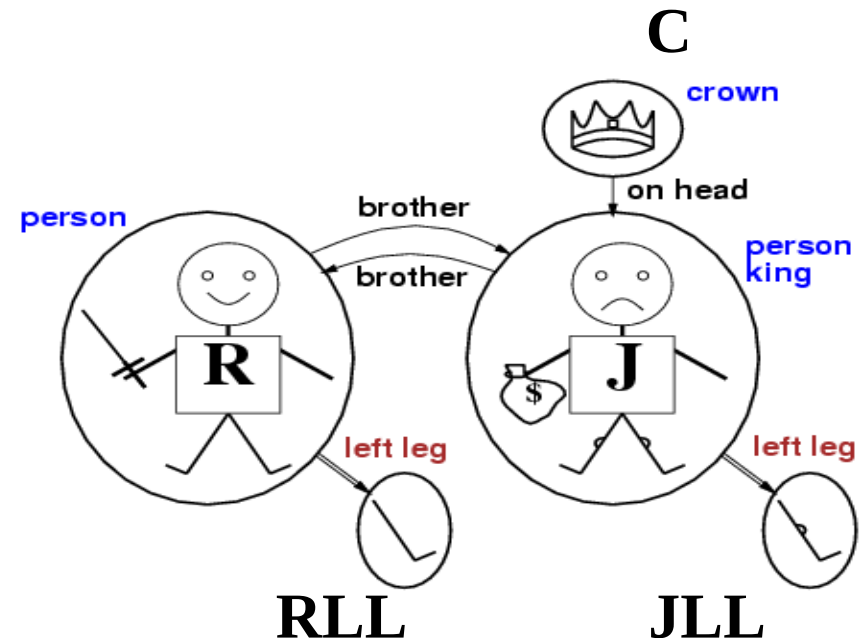
Sentences are true with respect to an **interpretation**

World contains objects (**domain elements**)

Interpretation specifies referents for

constant symbols	→	objects
predicate symbols	→	relations
function symbols	→	functional relations

- An **atomic sentence** $\text{predicate}(\text{term}_1, \dots, \text{term}_n)$ **is true** iff the objects referred to by $\text{term}_1, \dots, \text{term}_n$ are in the **relation** referred to by **predicate**



Quantifiers

Allows us to express

- **Properties of collections of objects** instead of enumerating objects by name

Universal: “for all” $\forall$

- **Properties of an unspecified object**

Existential: “there exists” $\exists$

Universal quantification

$\forall \langle \text{variables} \rangle \langle \text{sentence} \rangle$

Everyone at UBC is smart:

$$\forall x \text{ At}(x, \text{UBC}) \Rightarrow \text{Smart}(x)$$

$\forall x P(x)$ is true in an interpretation I iff P is true with x being each possible object in I

Equivalent to the conjunction of instantiations of P

$$\begin{aligned} & \text{At}(\text{KingJohn}, \text{UBC}) \Rightarrow \text{Smart}(\text{KingJohn}) \\ & \wedge \text{At}(\text{Richard}, \text{UBC}) \Rightarrow \text{Smart}(\text{Richard}) \\ & \wedge \text{At}(\text{Ralphie}, \text{UBC}) \Rightarrow \text{Smart}(\text{Ralphie}) \\ & \wedge \dots \end{aligned}$$

Existential quantification

$\exists$ <variables> <sentence>

Someone at UBC is smart:

$$\exists x \text{ At}(x, \text{UBC}) \wedge \text{Smart}(x)$$

$\exists x P(x)$ is true in an interpretation I iff P is true with x being some possible object in I

Equivalent to the **disjunction** of **instantiations** of P

$$\text{At}(\text{KingJohn}, \text{UBC}) \wedge \text{Smart}(\text{KingJohn})$$

$$\vee \text{ At}(\text{Richard}, \text{UBC}) \wedge \text{Smart}(\text{Richard})$$

$$\vee \text{ At}(\text{Ralphie}, \text{UBC}) \wedge \text{Smart}(\text{Ralphie})$$

$$\vee \dots$$

Properties of quantifiers

$\exists x \forall y$ is **not** the same as $\forall y \exists x$

$\exists x \forall y \text{ Loves}(x,y)$

- “There is a person who loves everyone in the world”

$\forall y \exists x \text{ Loves}(x,y)$

- “Everyone in the world is loved by at least one person”

Quantifier duality: each can be expressed using the other

$\forall x \text{ Likes}(x, \text{IceCream}) \quad \neg \exists x \neg \text{Likes}(x, \text{IceCream})$

$\exists x \text{ Likes}(x, \text{Broccoli}) \quad \neg \forall x \neg \text{Likes}(x, \text{Broccoli})$

FOL: Inference

Resolution Procedure can be generalized to FOL

- **Every formula** can be rewritten in **logically equivalent CNF**
 - Additional rewriting rules for quantifiers
- **Similar Resolution step**, but variables need to be unified (like in DATALOG)

$$\begin{array}{l} \left\{ \begin{array}{l} \text{In}(x, y) \vee \neg \text{Charged}(x) \\ \neg \text{In}(z, v) \vee \text{Connected}(z) \end{array} \right. \quad \theta = \{z/x, v/y\} \\ \rightarrow \boxed{\neg \text{Charged}(x) \vee \text{Connected}(x)} \end{array}$$

NLP Practical Goal for FOL: the ultimate Web question-answering system?

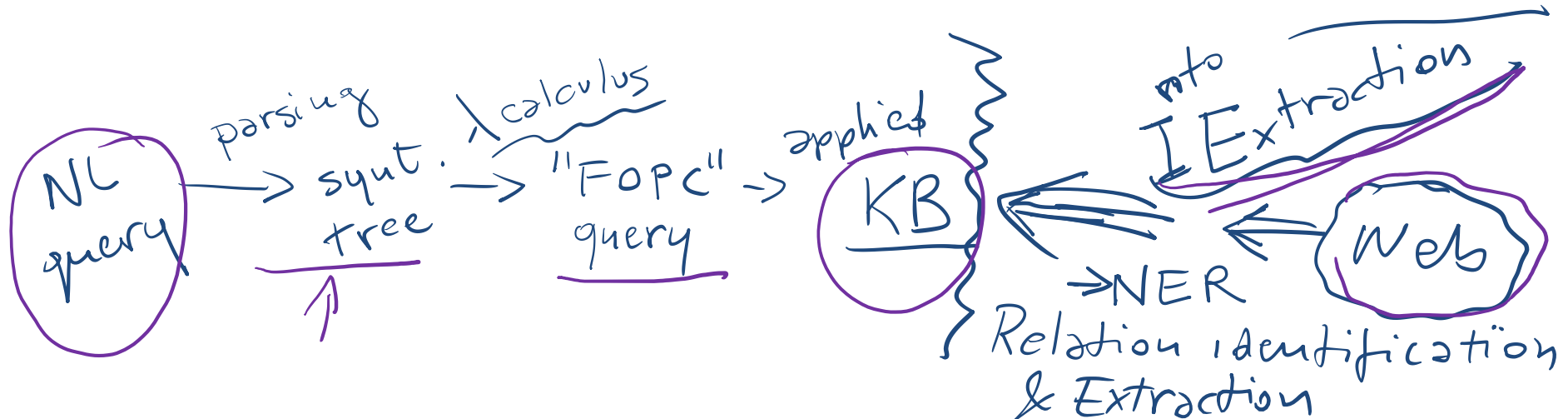
Map NL queries into FOL so that answers can be effectively computed

What African countries are not on the Mediterranean Sea?

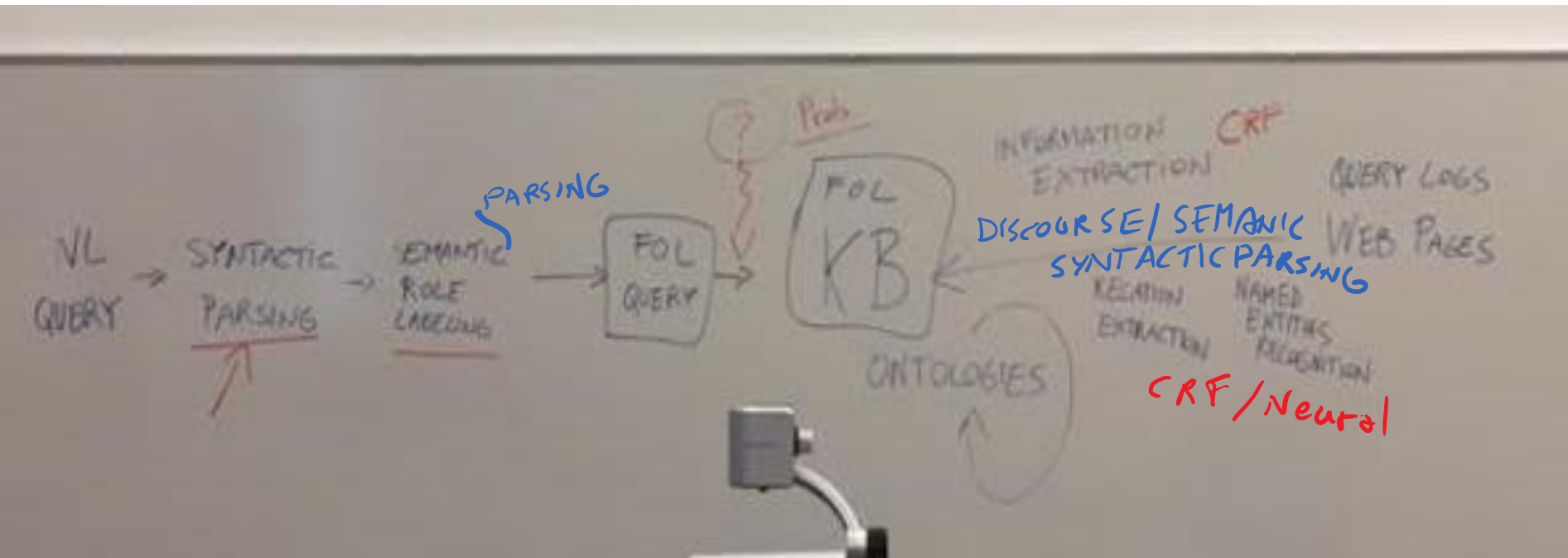
$\exists c \text{ Country}(c) \wedge \neg \text{Borders}(c, \text{Med.Sea}) \wedge \text{In}(c, \text{Africa})$

- Was 2007 the first El Nino year after 2001?

$\text{ElNino}(2007) \wedge \neg \exists y \text{ Year}(y) \wedge \text{After}(y, 2001) \wedge \text{Before}(y, 2007) \wedge \text{ElNino}(y)$



Just a sketch: to provide some context for some concepts / techniques covered in 422



Logics in AI: Similar slide to the one for planning

Propositional Definite Clause Logics

Semantics and Proof Theory

BU ^{Sound}
TD ^{& Complete}

Propositional Logics

First-Order Logics

Data log

Satisfiability Testing (SAT)

Hardware Verification

Ontologies

Production Systems

Product Configuration

you will know
you will know a little

Cognitive Architectures

Video Games

Some Applications

Semantic Web

Summarization

Tutoring Systems

Information Extraction

Lecture Overview

- **Ontologies** – what objects/individuals should we represent? what relations (unary, binary,..)?
- Inspiration from **Natural Language**: WordNet and FrameNet
- Extensions based on Wikipedia and mining the Web (YAGO, ProBase, Freebase)
- Domain Specific Ontologies (e.g., Medicine: MeSH, UMLS)

Links to Web Interfaces on course webpage

Each can be downloaded

Ontologies

Given a logical representation (e.g., FOL)

What individuals and relations are there and we need to model?

In **AI** an **Ontology** is a specification of what **individuals** and **relationships** are **assumed to exist** and what terminology is used for them

- What **types** of individuals
- What **properties** of the individuals
-

Ontologies: inspiration from Natural Language

:

How do we refer to individuals and relationships in the world in Natural Languages e.g., English?

Where do we find definitions for words?

Most of the definitions are circular? They are descriptions.

Fortunately, there is still some useful semantic info (**Lexical Relations**):

w_1 w_2 same Form and Sound, different Meaning *Homonymy*

w_1 w_2 same Meaning, different Form *Synonymy*

w_1 w_2 "opposite" Meaning *Antonymy*

w_1 w_2 Meaning₁ subclass of Meaning₂ *Hyponymy*

Polysemy

Def. The case where we have a set of words with the same form and **multiple related meanings**.

Consider the homonym:

bank → commercial **bank**₁ vs. river **bank**₂

- **Now consider:** "VGH is the hospital with the largest blood **bank** in BC" or
- *"A PCFG can be trained using derivation trees from a tree **bank** annotated by human experts"*
- **Are these a new independent senses of bank?**

Synonyms

Def. Different words with the same meaning.

Substitutability- if they can be substituted for one another in **some** environment without changing meaning or acceptability.

Would I be flying on a **large/big** plane?

?... became kind of a **large/big** sister to...

? You made a **large/big** mistake



Hyponymy/Hypernym

Def. Pairings where one word denotes a sub/super class of the other

- Since dogs are canids
 - ✓ Dog is a *hyponym* of canid and
 - ✓ Canid is a *hypernym* of dog

car/vehicle

doctor/human

.....

Lexical Resources

Databases containing all lexical relations among all words

- **Development:**
 - Mining info from dictionaries and thesauri
 - Handcrafting it from scratch
- **WordNet:** first developed with reasonable coverage and widely used, started with [Fellbaum... 1998]
 - for English (versions for other languages have been developed - see MultiWordNet)

WordNet 3.0

Part Of Speech	Unique Strings	Word-Sense Pairs	Synsets
Noun	117798	146312	82115
Verb	11529	25047	13767
Adjective	21479	30002	18156
Adverb	4481	5580	3621
Totals	155287	206941	117659

- For each word: all possible senses (no distinction between homonymy and polysemy)
- For each sense: a set of synonyms (**synset**) and a gloss

WordNet: entry for “table”

The **noun** "table" has 6 senses in WordNet. *gloss*

1. **table, tabular array** -- (*a set of data ...*)
2. **table** -- (*a piece of furniture ...*)
3. **table** -- (*a piece of furniture with tableware...*)
4. **mesa, table** -- (*flat tableland ...*)
5. **table** -- (*a company of people ...*)
6. **board, table** -- (*food or meals ...*)

The **verb** "table" has 1 sense in WordNet.

1. **postpone, prorogue, hold over, put over, table, shelve, set back, defer, remit, put off** – (*hold back to a later time; "let's postpone the exam"*)

WordNet Relations (between synsets!)

NOUNS

Relation	Definition	Example
Hypernym	From concepts to superordinates	<i>breakfast</i> → <i>meal</i>
Hyponym	From concepts to subtypes	<i>meal</i> → <i>lunch</i>
Has-Member	From groups to their members	<i>faculty</i> → <i>professor</i>
Member-Of	From members to their groups	<i>copilot</i> → <i>crew</i>
Has-Part	From wholes to parts	<i>table</i> → <i>leg</i>
Part-Of	From parts to wholes	<i>course</i> → <i>meal</i>
Antonym	Opposites	<i>leader</i> → <i>follower</i>

VERBS

Relation	Definition	Example
Hypernym	From events to superordinate events	<i>fly</i> → <i>travel</i>
Troponym	From events to their subtypes	<i>walk</i> → <i>stroll</i>
Entails	From events to the events they entail	<i>snore</i> → <i>sleep</i>
Antonym	Opposites	<i>increase</i> ↔ <i>decrease</i>

WordNet Hierarchies: “Vancouver”

WordNet: example from ver1.7.1

For the three senses of “Vancouver”

⇒ (city, metropolis, urban center)

⇒ (municipality)

⇒ (urban area)

⇒ (geographical area)

⇒ (region)

⇒ (location)

⇒ (entity, physical thing)

⇒ (administrative district, territorial division)

⇒ (district, territory)

⇒ (region)

⇒ (location)

⇒ (entity, physical thing)

⇒ (port)

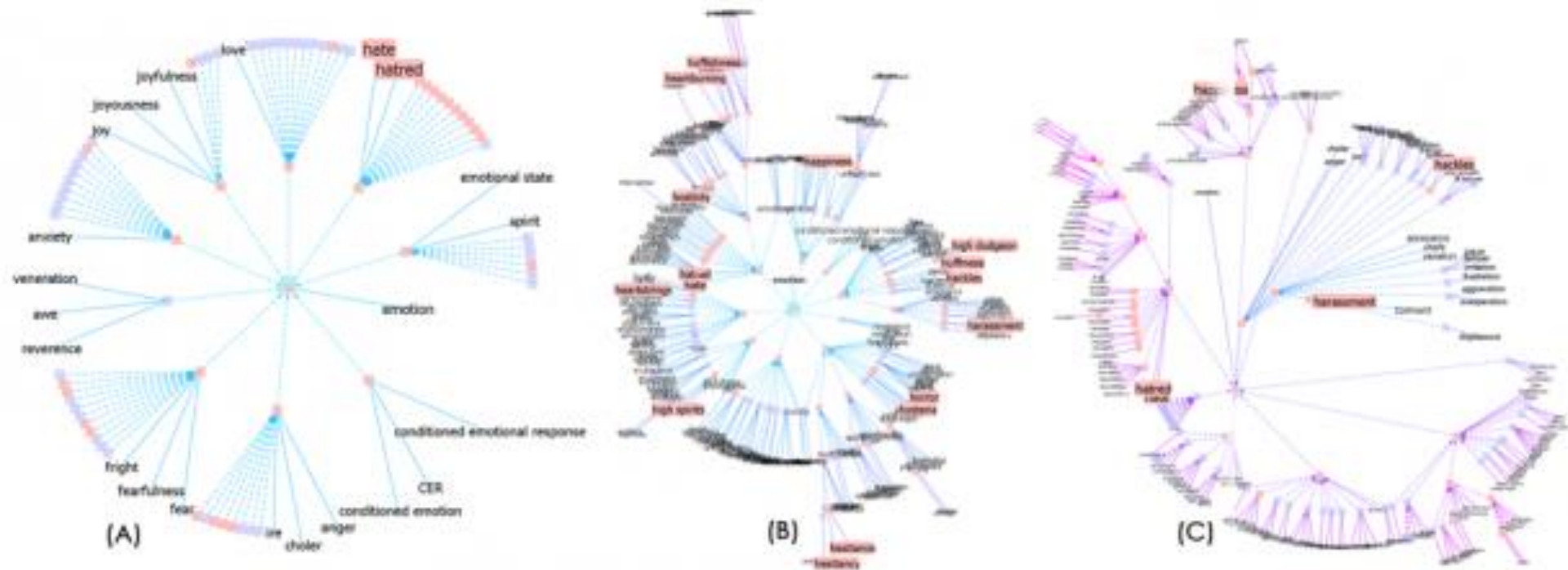
⇒ (geographic point)

⇒ (point)

⇒ (location)

⇒ (entity, physical thing)

Visualizing Wordnet Relations



C. Collins, "WordNet Explorer: Applying visualization principles to lexical semantics," University of Toronto, Technical Report km-di 2007-2, 2007.

Web interface & API

WordNet Search - 3.1

- [WordNet home page](#) - [Glossary](#) - [Help](#)

Word to search for:

Display Options:

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Display options for sense: (gloss) "an example sentence"

Noun

- [S:](#) (n) **bass** (the lowest part of the musical range)
- [S:](#) (n) **bass**, [bass part](#) (the lowest part in polyphonic music)
- [S:](#) (n) **bass**, [basso](#) (an adult male singer with the lowest voice)
 - [direct hypernym](#) / [inherited hypernym](#) / [sister term](#)
 - [S:](#) (n) [singer](#), [vocalist](#), [vocalizer](#), [vocaliser](#) (a person who sings)
- [S:](#) (n) [sea bass](#), **bass** (the lean flesh of a saltwater fish of the family Serranidae)
- [S:](#) (n) [freshwater bass](#), **bass** (any of various North American freshwater fish with lean flesh (especially of the genus Micropterus))
- [S:](#) (n) **bass**, [bass voice](#), [basso](#) (the lowest adult male singing voice)
- [S:](#) (n) **bass** (the member with the lowest range of a family of musical instruments)
- [S:](#) (n) **bass** (nontechnical name for any of numerous edible marine and freshwater spiny-finned fishes)

Adjective

- [S:](#) (adj) **bass**, [deep](#) (having or denoting a low vocal or instrumental range) "a deep voice"; "a bass voice is lower than a baritone voice"; "a bass clarinet"

Wordnet: NLP Tasks

- First success in “obscure” task for Probabilistic Parsing (PP-attachments): words + word-classes extracted from the hypernym hierarchy increase accuracy from 84% to 88% [Stetina and Nagao, 1997]
 - Word sense disambiguation
 - Lexical Chains (summarization)
 - and *many others* !
- 

More importantly starting point for larger Ontologies!

More ideas from NLP....

Relations among words and their meanings
(*paradigmatic*)

Internal structure of individual words
(*syntagmatic*)

Predicate-Argument Structure

- Represent relationships among **concepts**, **events** and their **participants**

"I ate a turkey sandwich for lunch"

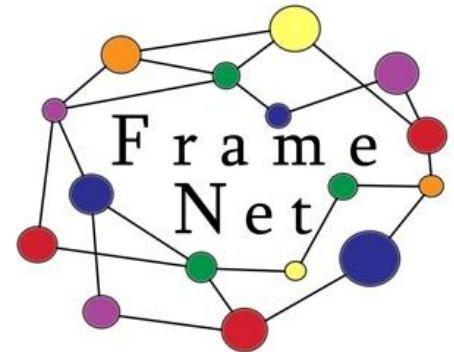
$\exists w: Isa(w, Eating) \wedge \text{Eater}(w, Speaker) \wedge$
 $\text{Eaten}(w, TurkeySandwich) \wedge \text{MealEaten}(w, Lunch)$

"Nam does not serve meat"

$\exists w: Isa(w, Serving) \wedge \text{Server}(w, Nam) \wedge$
 $\neg \text{Served}(w, Meat)$

Semantic Roles: Resources

- Move beyond inferences about single verbs
 - " IBM hired John as a CEO "*
 - " John is the new IBM hire "*
 - " IBM signed John for 2M\$ "*
- **FrameNet**: Databases containing **frames** and their syntactic and semantic argument structures
- book online Version 1.5-update (revised in 2016)
 - for English (versions for other languages are
- FrameNet Tutorial at NAACL/HLT 2015!



FrameNet Entry

Hiring

- **Definition:** An **Employer** hires an **Employee**, promising the **Employee** a certain **Compensation** in exchange for the performance of a job. The job may be described either in terms of a **Task** or a **Position** in a **Field**.
- **Inherits From:** Intentionally affect
- **Lexical Units:** *commission.n, commission.v, give job.v, hire.n, hire.v, retain.v, sign.v, take on.v*

FrameNet : Semantic Role Labeling

Some roles..

Employer Employee Task Position

- np-vpto
 - In 1979 , singer Nancy Wilson HIRED him to open her nightclub act .
 -
- np-ppas
 - Castro has swallowed his doubts and HIRED Valenzuela as a cook in his small restaurant .

Lecture Overview

- **Ontologies** – what objects/individuals should we represent? what relations (unary, binary,..)?
 - Inspiration from **Natural Language**: WordNet and FrameNet
 - Extensions based on Wikipedia and mining the Web & Web search logs (YAGO, ProBase, Freebase,.....)
 - Domain Specific Ontologies (e.g., Medicine: MeSH, UMLS)
- 

YAGO2: huge semantic knowledge base

Derived from **Wikipedia**, **WordNet** and **GeoNames**.
(started in 2007, paper in www conference)

10^6 entities (persons, organizations, cities, etc.)

$>120 * 10^6$ facts about these entities.

- **YAGO accuracy of 95%. has been manually evaluated.**
- **Anchored in *time* and *space*. YAGO attaches a *temporal* dimension and a *spatial* dimension to many of its facts and entities.**

Freebase

- “Collaboratively constructed database.”
- Freebase contains **tens of millions** of topics, **thousands of types**, and **tens of thousands of properties** and over **a billion of facts**
- Automatically extracted from a number of resources including **Wikipedia**, **MusicBrainz**, and Notable Names Database (**NNDB**)
- as well as the **knowledge contributed by the human volunteers.**
- All **was available for free** through the APIs or to download from weekly data dumps

Fast Changing Landscape....

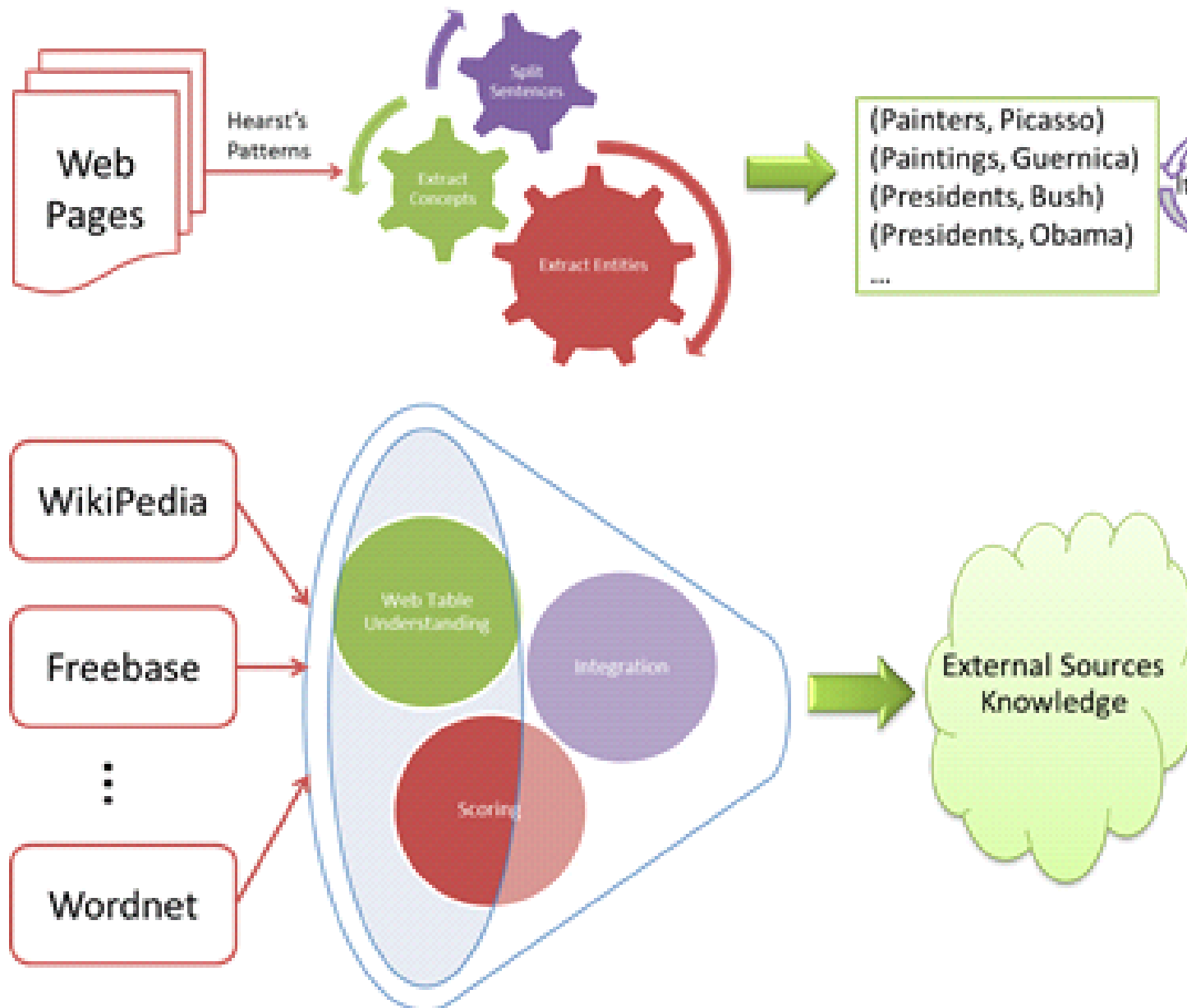
On 16 December 2015, Google officially announced the Knowledge Graph API, which is meant to be a replacement to the Freebase API.

Freebase.com was officially shut down on 2 May 2016.^[6]

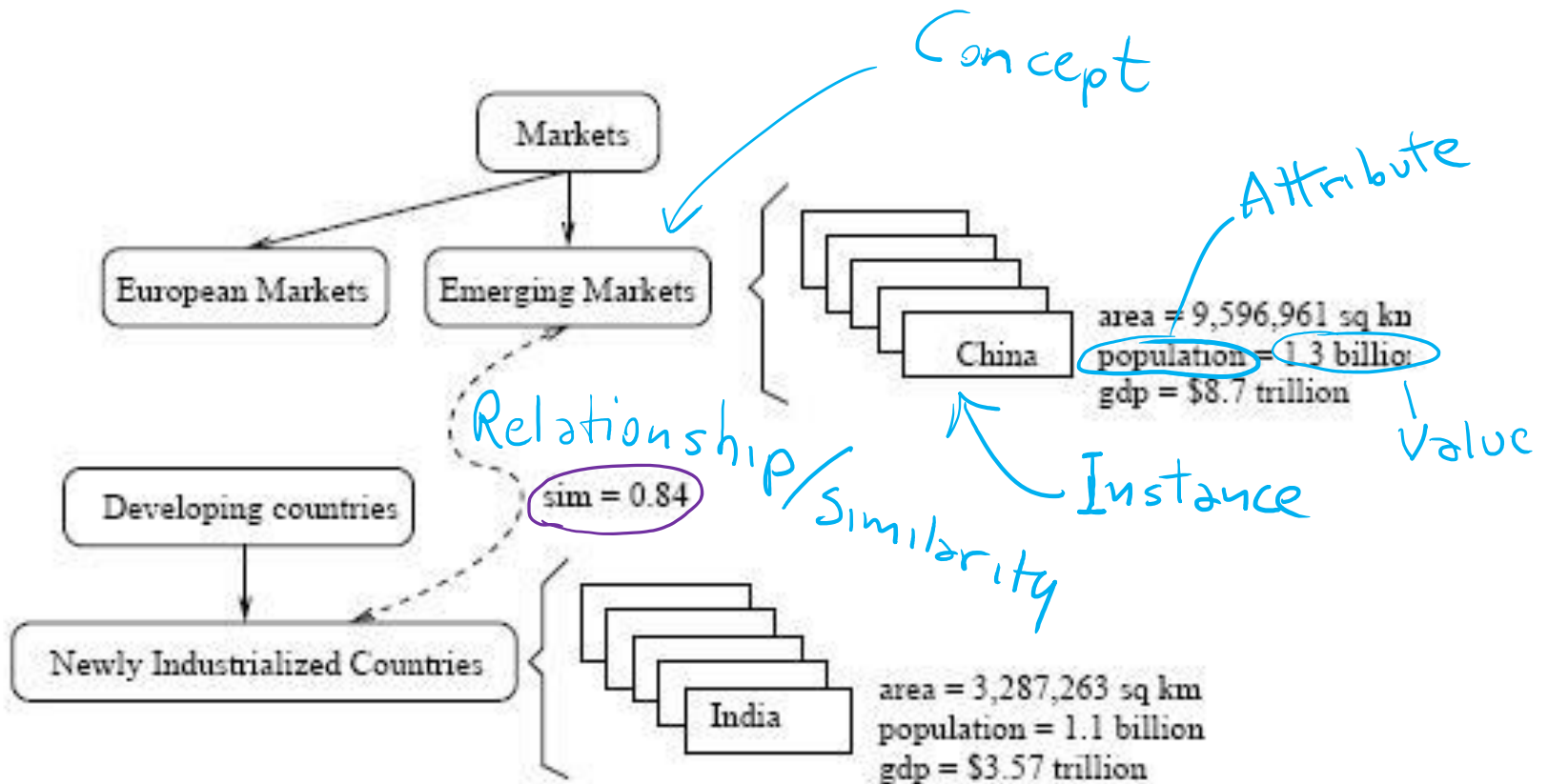
Probase (MS Research)

- Harnessed from billions of **web pages** and years worth of **search logs**
- Extremely large concept/category space (2.7 million categories).
- Probabilistic model for correctness, typicality (e.g., between concept and instance)

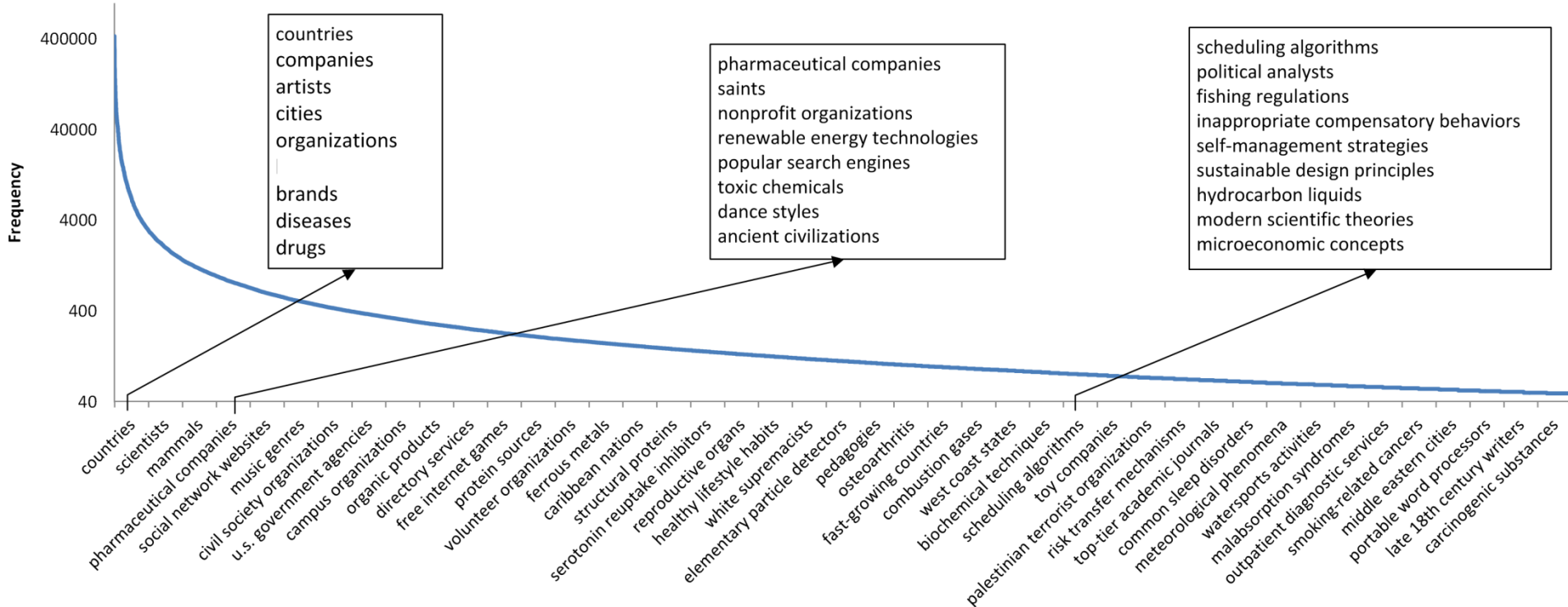
Infrastructure



A snippet of Probase's core taxonomy



Frequency distribution of the 2.7 million concepts



The **Y axis** is the **number of instances each concept contains** and on the **X axis** are the 2.7 million concepts ordered by their size.

besides popular concepts such as “cities” and “musicians”, which are included by almost every general purpose taxonomy, Probase has millions of long tail concepts such as “[anti-parkinson treatments](#)”, “[celebrity wedding dress designers](#)” and “[basic watercolor techniques](#)”,

Fast Changing Landscape....

From Probase page.....

[Sept. **2016**] Please visit our Microsoft Concept Graph release for up-to-date information of this project!

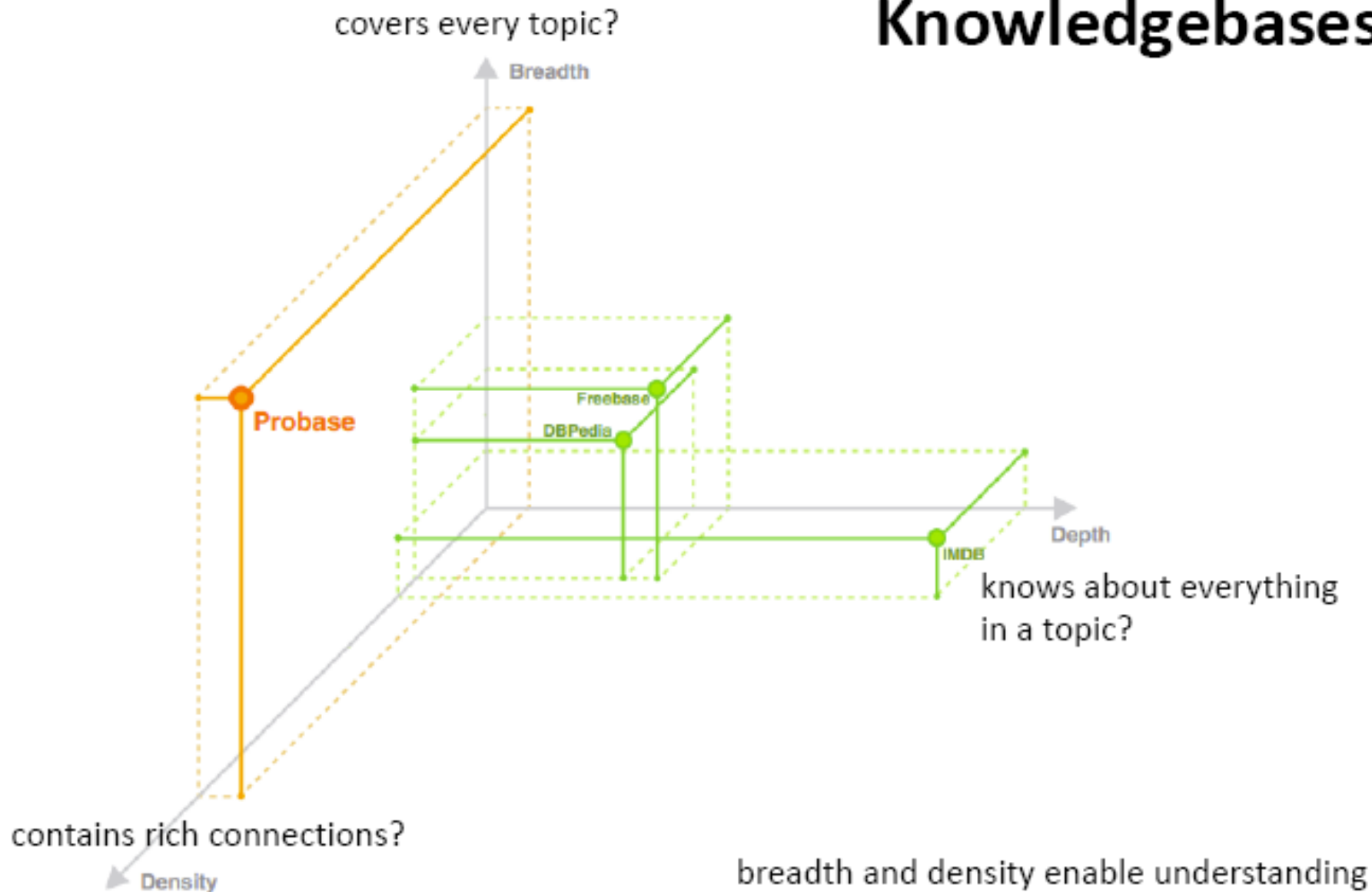
Another one DBpedia

DBpedia is a crowd-sourced community effort to extract structured information from Wikipedia and make this information available on the Web.

- allows you to ask sophisticated queries against Wikipedia,
- link the different data sets on the Web to Wikipedia data

Interesting dimensions to compare Ontologies (but form Probase so possibly biased)

Knowledgebases



Lecture Overview

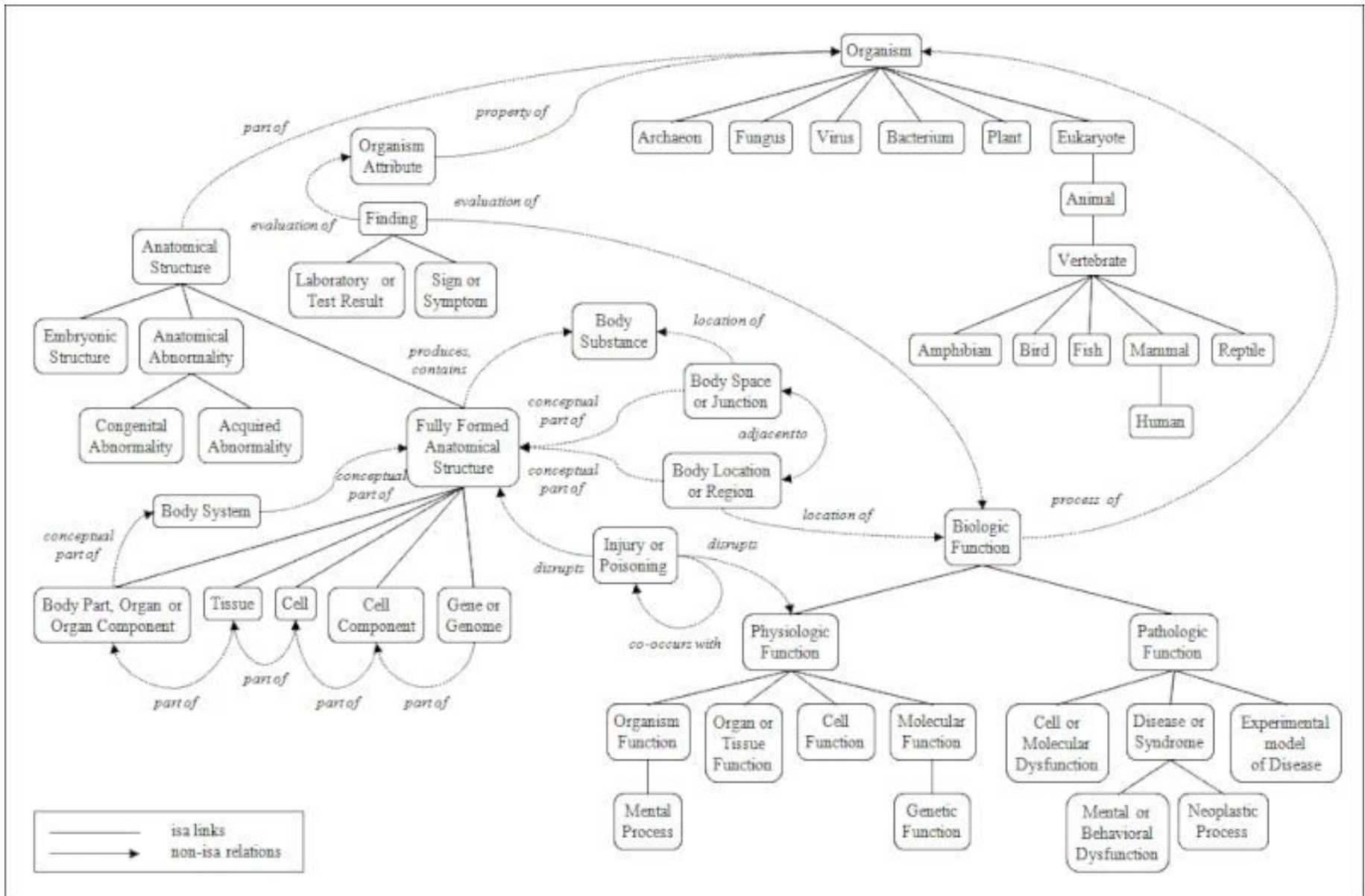
- **Ontologies** – what objects/individuals should we represent? what relations (unary, binary,..)?
- Inspiration from **Natural Language**: WordNet and FrameNet
- Extensions based on Wikipedia and mining the Web (YAGO, ProBase, Freebase)
- Domain Specific Ontologies (e.g., Medicine: MeSH, UMLS)

Domain Specific Ontologies: UMLS, MeSH

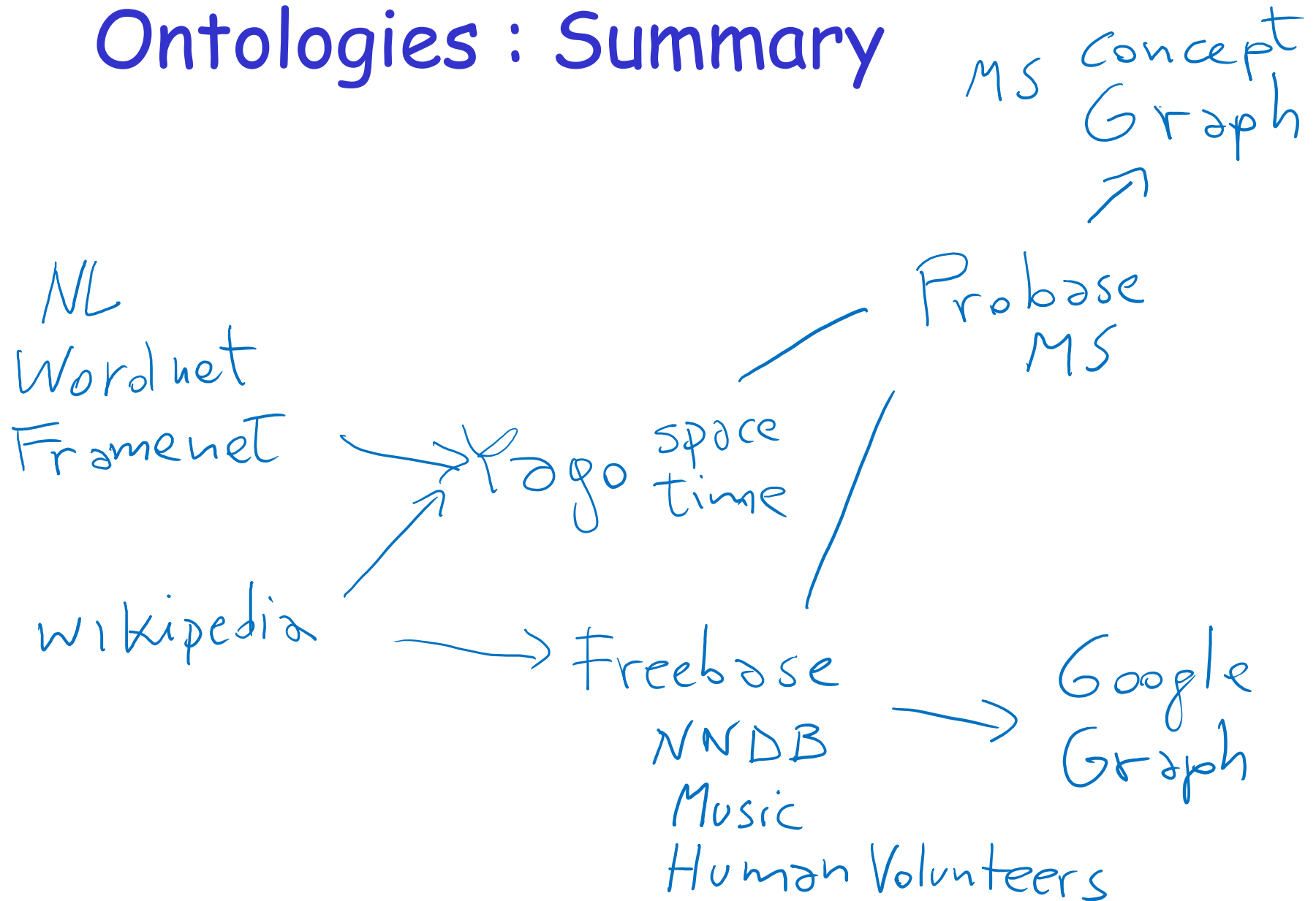
Unified Medical Language System: brings together many health and biomedical vocabularies

- Enable interoperability (linking medical terms, drug names)
- Develop electronic health records, classification tools
- Search engines, data mining

Portion of the UMLS Semantic Net



Ontologies : Summary



Learning Goals for today's class

You can:

- Define an Ontology
- Describe and Justify the information represented in Wordnet and Framenet
- Describe and Justify the three dimensions for comparing ontologies

Assignment-3 out - due Nov 18
(8-18 hours - working in pairs on programming
parts is strongly advised)

Next class Fri

- **Similarity measures in ontologies (e.g., Wordnet)**

DBpedia is a structured twin of Wikipedia. Currently it describes more than 3.4 million entities. DBpedia resources bear the names of the Wikipedia pages, from which they have been extracted.

YAGO is an automatically created ontology, with taxonomy structure derived from WordNet, and knowledge about individuals extracted from Wikipedia. Therefore, the identifiers of resources describing individuals in YAGO are named as the corresponding Wikipedia pages. YAGO contains knowledge about more than 2 million entities and 20 million facts about them.

Freebase is a collaboratively constructed database. It contains knowledge automatically extracted from a number of resources including Wikipedia, MusicBrainz,² and NNDB,³ as well as the knowledge contributed by the human volunteers. Freebase describes more than 12 million interconnected entities. Each Freebase entity is assigned a set of human-readable unique keys, which are assembled of a value and a namespace. One of the namespaces is the Wikipedia namespace, in which a value is the name of the Wikipedia page describing an entity.

Summary

- Relations among words and their meanings

Wordnet

YAGO

Probase

- Internal structure of individual words

PropBank

VerbNet

FrameNet

Table 1: Scale of concept dimension

<i>name</i>	<i># of concepts</i>
SenticNet	14,244
Freebase	1,450
WordNet	25,229
WikiTaxonomy	< 127,325
YAGO	149,162
DBPedia	259
ResearchCyc	≈ 120,000
KnowItAll	N/A
TextRunner	N/A
OMCS	23,365
NELL	123
Probase	2,653,872

Today 12 Feb

Syntax-Driven Semantic Analysis

Meaning of words

- Relations among words and their meanings (**Paradigmatic**)
- Internal structure of individual words (**Syntagmatic**)

Practical Goal for (Syntax-driven) Semantic Analysis

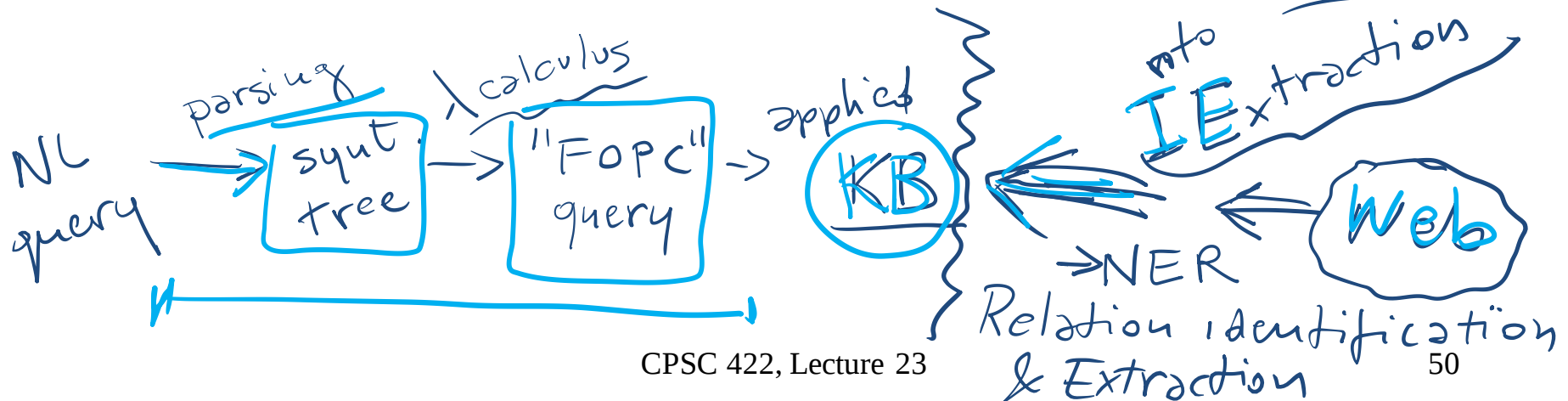
Map NL queries into FOPC so that answers can be effectively computed

- What African countries are not on the Mediterranean Sea?

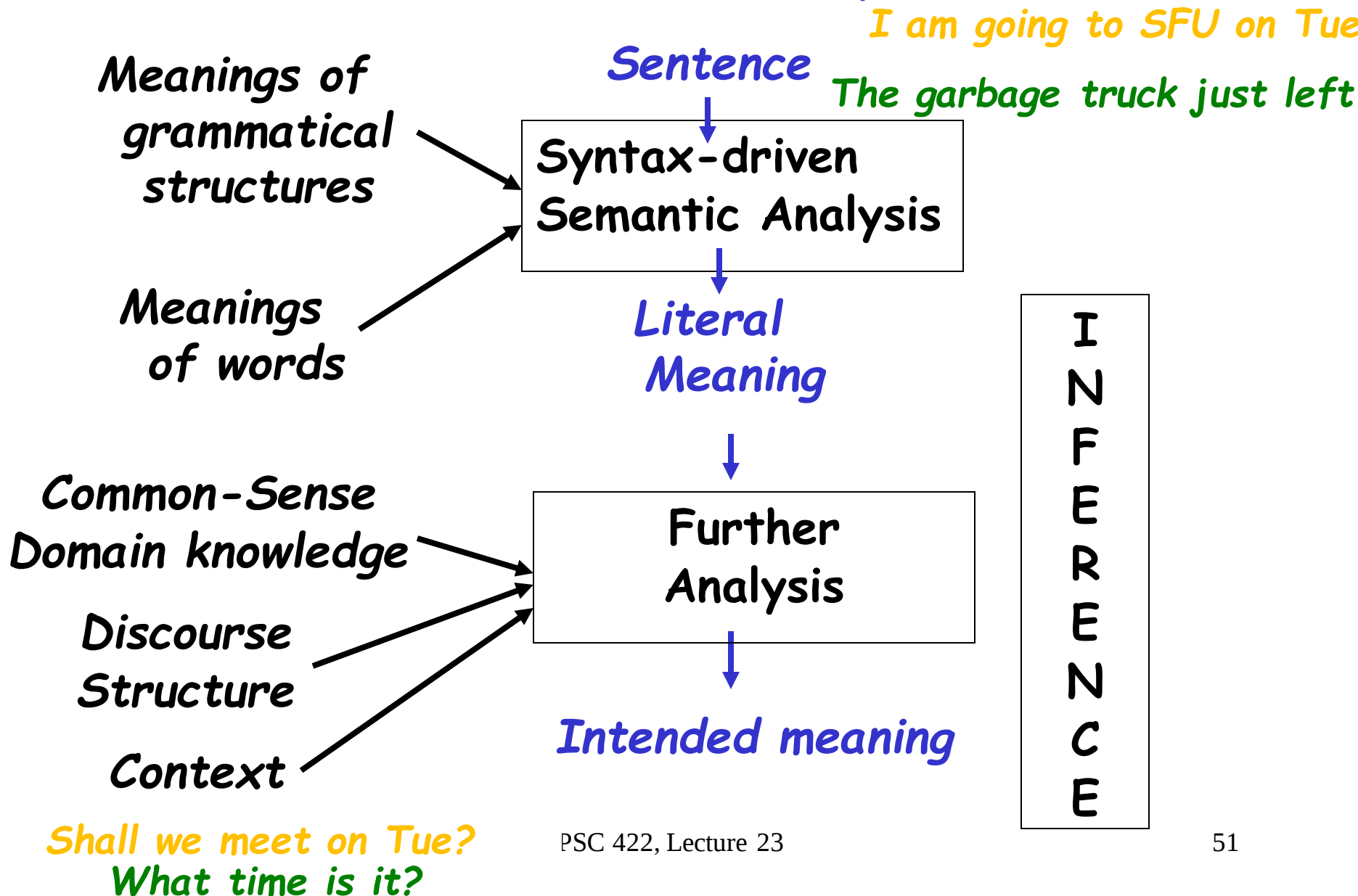
$\exists c \text{ Country}(c) \wedge \neg \text{Borders}(c, \text{Med.Sea}) \wedge \text{In}(c, \text{Africa})$

- Was 2007 the first El Nino year after 2001?

$\text{ElNino}(2007) \wedge \neg \exists y \text{ Year}(y) \wedge \text{After}(y, 2001) \wedge$
 $\text{Before}(y, 2007) \wedge \text{ElNino}(y)$



Semantic Analysis



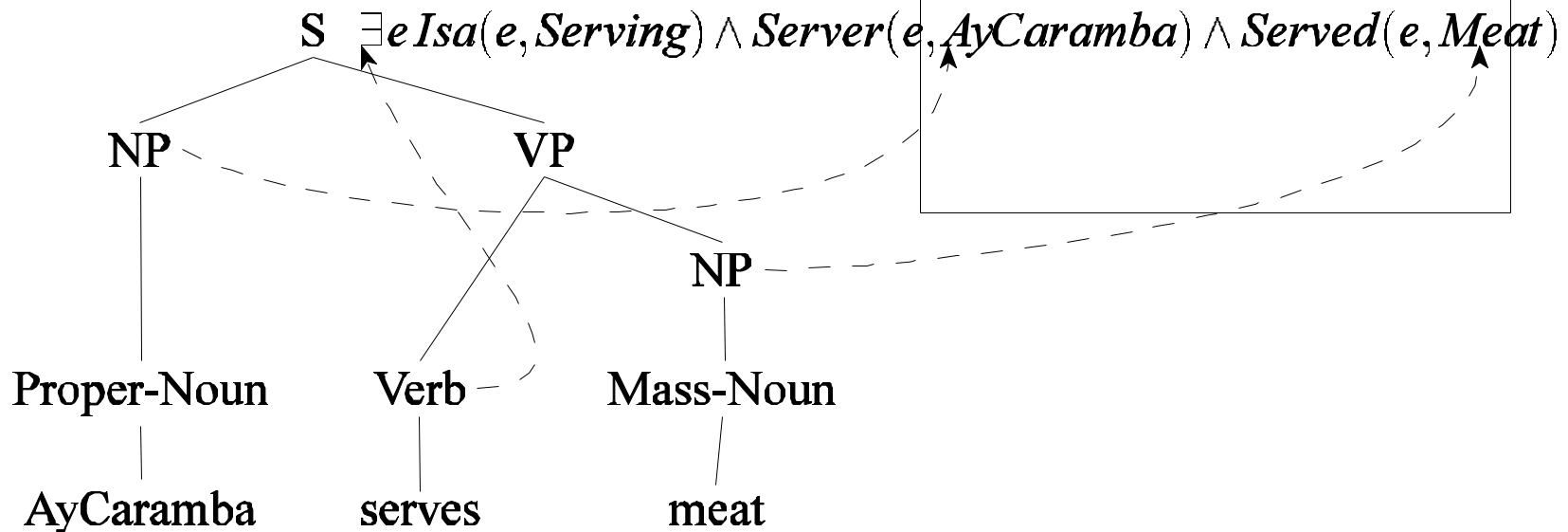
Compositional Analysis

- **Principle of Compositionality**
 - The meaning of a whole is derived from the meanings of the parts
- **What parts?**
 - The constituents of the syntactic parse of the input

Compositional Analysis: Example

- *AyCaramba serves meat*

$\exists e \text{ Serving}(e) \wedge \text{Server}(e, \text{AyCaramba}) \wedge \text{Served}(e, \text{Meat})$



Augmented Rules

- Augment each syntactic CFG rule with a semantic formation rule
- Abstractly

$$A \rightarrow \alpha_1 \dots \alpha_n \quad \{ f(\alpha_1.sem, \dots \alpha_n.sem) \}$$

- i.e., The semantics of A can be computed from some function applied to the semantics of its parts.
- The class of actions performed by f will be quite restricted.

Simple Extension of FOL: Lambda Forms

- A FOL sentence with variables in it that are to be bound.
- **Lambda-reduction**: variables are bound by treating the lambda form as a function with formal arguments

$$\lambda x P(x)$$



$$\lambda x P(x)(Sally) \\ P(Sally)$$

$$\lambda x \lambda y In(x, y) \wedge Country(y) \rightarrow \lambda x \lambda y In(x, y) \wedge Country(y)(BC) \\ \lambda y In(BC, y) \wedge Country(y)$$

$$\lambda y In(BC, y) \wedge Country(y)$$

$$\lambda y In(BC, y) \wedge Country(y)(CANADA) \\ In(BC, CANADA) \wedge Country(CANADA)$$

Augmented Rules: Example

- Concrete entities

assigning FOL constants

- Attachments
 - PropNoun -> AyCaramba {AyCaramba}
 - MassNoun -> meat {MEAT}

- Simple non-terminals

copying from daughters up to mothers.

- Attachments
 - NP -> PropNoun {PropNoun.sem}
 - NP -> MassNoun {MassNoun.sem}

Augmented Rules: Example

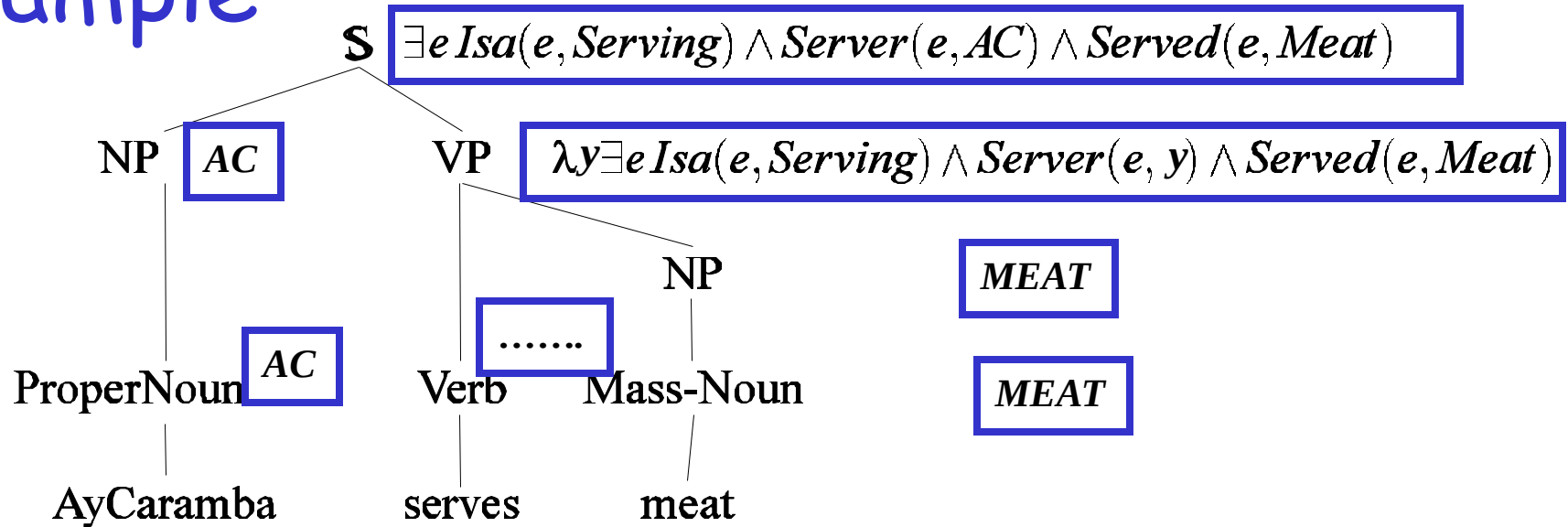
Semantics attached to one daughter is applied to semantics of the other daughter(s).

- $S \rightarrow NP VP$ *daughter(s)* • $\{VP.sem(NP.sem)\}$
- $VP \rightarrow Verb NP$ • $\{Verb.sem(NP.sem)\}$

lambda-form

- **Verb** -> *serves* $\lambda x \lambda y \exists e \text{Serving}(e) \wedge \text{Server}(e, y) \wedge \text{Served}(e, x)$

Example



- $S \rightarrow NP VP$
- $VP \rightarrow Verb NP$
- $Verb \rightarrow serves$
- $NP \rightarrow PropNoun$
- $NP \rightarrow MassNoun$
- $PropNoun \rightarrow AyCaramba$
- $MassNoun \rightarrow meat$
- $\{VP.sem(NP.sem)\}$
- $\{Verb.sem(NP.sem)\}$
- $\lambda x \lambda y \exists e Serving(e) \wedge Server(e, y) \wedge Served(e, x)$
- $\{PropNoun.sem\}$
- $\{MassNoun.sem\}$
- $\{AC\}$
- $\{MEAT\}$

References (Project?)

- **Text Book: Representation and Inference for Natural Language : A First Course in Computational Semantics** *Patrick Blackburn and Johan Bos, 2005, CSLI*
- *J. Bos (2011): A Survey of Computational Semantics: Representation, Inference and Knowledge in Wide-Coverage Text Understanding. Language and Linguistics Compass 5(6): 336–366.*

Next Time

- **Read Chp. 19 (Lexical Semantics)**

Next Time

Read Chp. 20

Computational Lexical Semantics

- **Word Sense Disambiguation**
- **Word Similarity**
- **Semantic Role Labeling**

~~Stem?~~

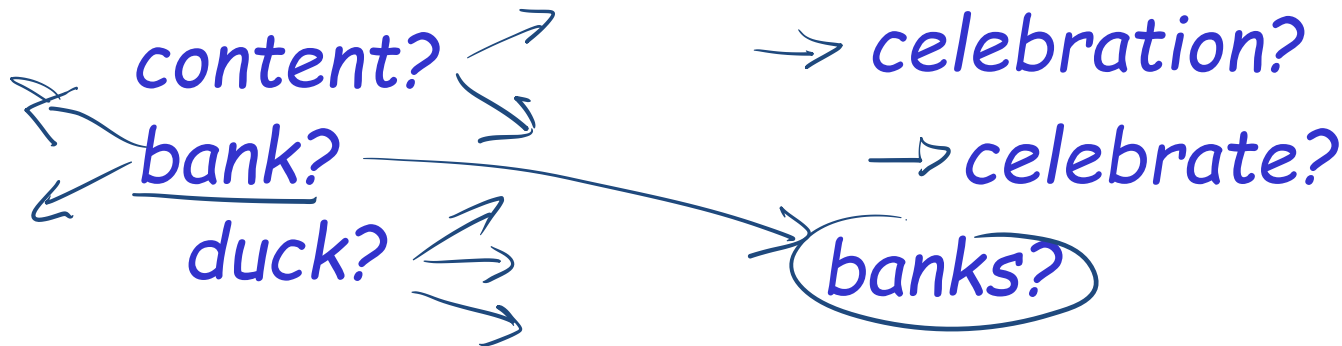
~~Word?~~

Lemma?

Lexeme:

- Orthographic form +
- Phonological form +
- Meaning (sense)

[Modulo inflectional morphology]



- Lexicon: A collection of lexemes

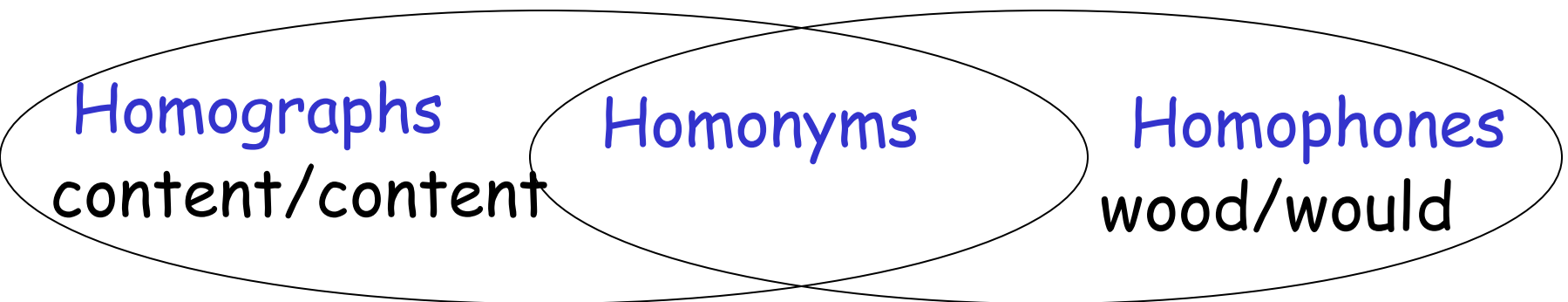
Homonymy

Def. Lexemes that have the same
“forms” but unrelated meanings

- **Examples:** Bat (wooden stick-like thing)
vs. Bat (flying scary mammal thing)

Plant (.....) vs.

Plant (.....)



Relevance to NLP Tasks

Information retrieval (homonymy):

✓ QUERY: bat

Spelling correction: homophones can lead to real-word spelling errors

Text-to-Speech: homographs (which are not homophones)

Polysemy

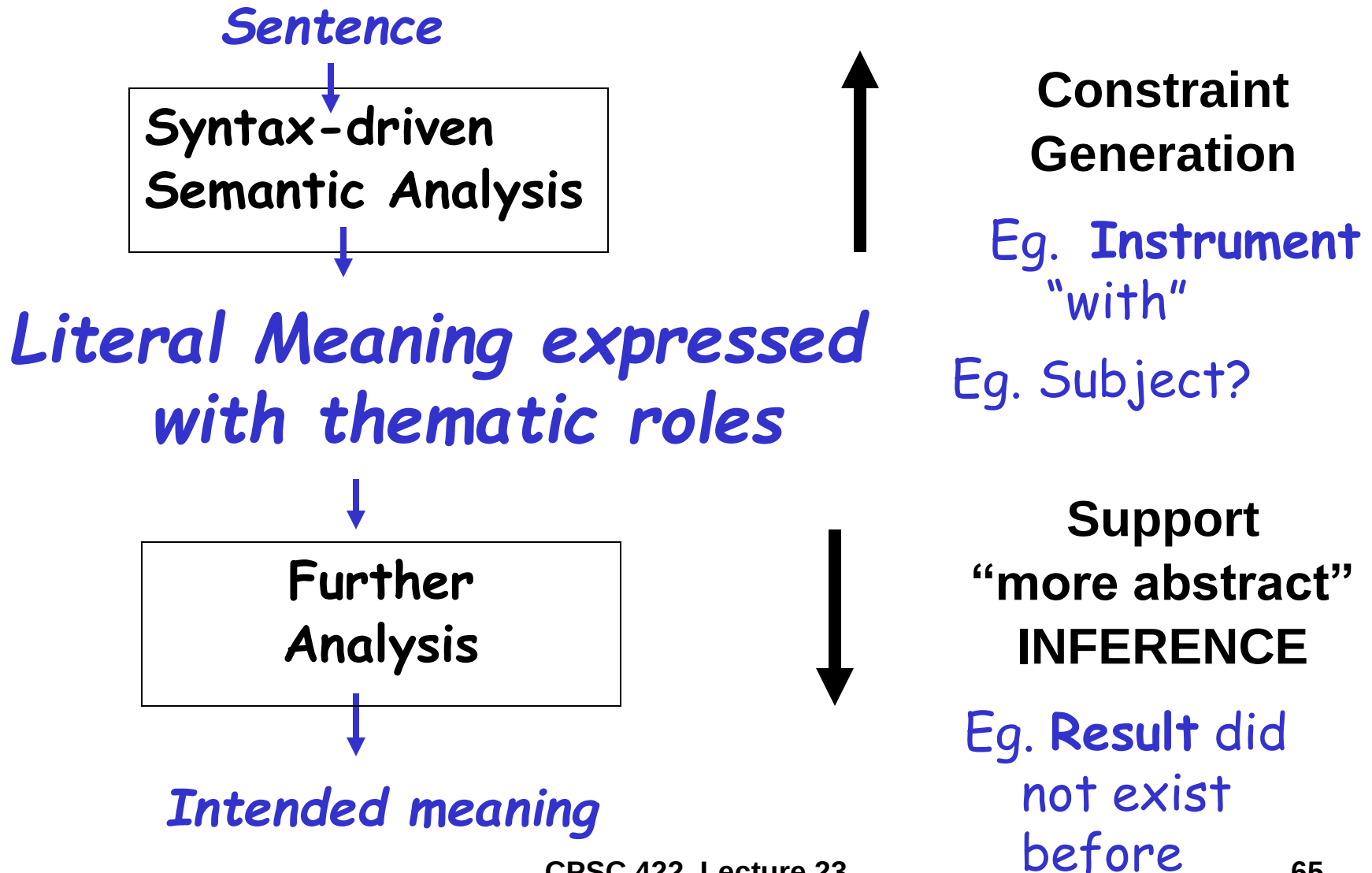
Lexeme (new def.):

- Orthographic form + Phonological form +
- Set of related senses

How many distinct (but related) senses?

- They **serve** meat...
 - He **served** as Dept. Head...
 - She **served** her time....
 - Does AC **serve** vegetarian food?
 - Does AC **serve** Rome?
 - (?)Does AC **serve** vegetarian food and Rome?
- ← Different subcat
- ← Intuition (prison)
- Zeugma

Thematic Roles: Usage



Semantic Roles

- Def. Semantic generalizations over the specific roles that occur with specific verbs.
- I.e. eaters, servers, takers, givers, makers, doers, killers, all have something in common
- We can generalize (or try to) across other roles as well

Thematic Role Examples

Thematic Role	Example
AGENT	<i>The waiter</i> spilled the soup.
EXPERIENCER	<i>John</i> has a headache.
FORCE	<i>The wind</i> blows debris from the mall into our yards.
THEME	Only after Benjamin Franklin broke <i>the ice</i> ...
RESULT	The French government has built a <i>regulation-size baseball diamond</i> ...
CONTENT	Mona asked “ <i>You met Mary Ann at a supermarket</i> ”?
INSTRUMENT	He turned to poaching catfish, stunning them <i>with a shocking device</i> ...
BENEFICIARY	Whenever Ann Callahan makes hotel reservations <i>for her boss</i> ...
SOURCE	I f lew in <i>from Boston</i> .
GOAL	I drove <i>to Portland</i> .

Thematic Roles

Thematic Role	Definition
AGENT	The volitional causer of an event
EXPERIENCER	The experiencer of an event
FORCE	The non-volitional causer of the event
THEME	The participant most directly affected by an event
RESULT	The end product of an event
CONTENT	The proposition or content of a propositional event
INSTRUMENT	An instrument used in an event
BENEFICIARY	The beneficiary of an event
SOURCE	The origin of the object of a transfer event
GOAL	The destination of an object of a transfer event

- Not definitive, not from a single theory!

Problem with Thematic Roles

- NO agreement of what should be the standard set
- NO agreement on formal definition
- **Fragmentation problem:** when you try to formally define a role you end up creating more specific sub-roles

Two solutions

- Generalized semantic roles
- Define verb (or class of verbs) specific semantic roles

Generalized Semantic Roles

- Very abstract roles are defined heuristically as a set of conditions
- The more conditions are satisfied the more likely an argument fulfills that role

- **Proto-Agent**

- Volitional involvement in event or state
- Sentience (and/or perception)
- Causing an event or change of state in another participant
- Movement (relative to position of another participant)
- (exists independently of event named)

- **Proto-Patient**

- Undergoes change of state
- Incremental theme
- Causally affected by another participant
- Stationary relative to movement of another participant
- (does not exist independently of the event, or at all)

Semantic Roles: Resources

- Databases containing for each **verb** its syntactic and thematic argument structures
- **PropBank**: sentences in the Penn Treebank annotated with semantic roles
- Roles are verb-sense specific
- Arg0 (PROTO-AGENT), Arg1(PROTO-PATIENT), Arg2,
- (see also VerbNet)

PropBank Example

- Increase "go up incrementally"

→ Arg0: causer of increase

→ Arg1: thing increasing

~ Arg2: amount increase by

~ Arg3: start point

~ Arg4: end point

Glosses for human reader. Not formally defined

- PropBank semantic role labeling would identify common aspects among these three examples

"[Y performance] ^{Arg1}increased ^{Arg2}[by 3%]"

"[Y performance] was increased [by the new X technique] ^{Arg0}"

"[The new X technique] ^{Arg0}increased [performance of Y] ^{Arg1}"