

Global WordNet and The Suggested Upper Merged Ontology (SUMO): Ontologies, lexicons and their Relationships

Presented at PANL10n

Adam Pease

Articulate Software

`apease at articulatesoftware dot com`

`http://www.ontologyportal.org/`

`http://home.earthlink.net/~adampease/professional/`

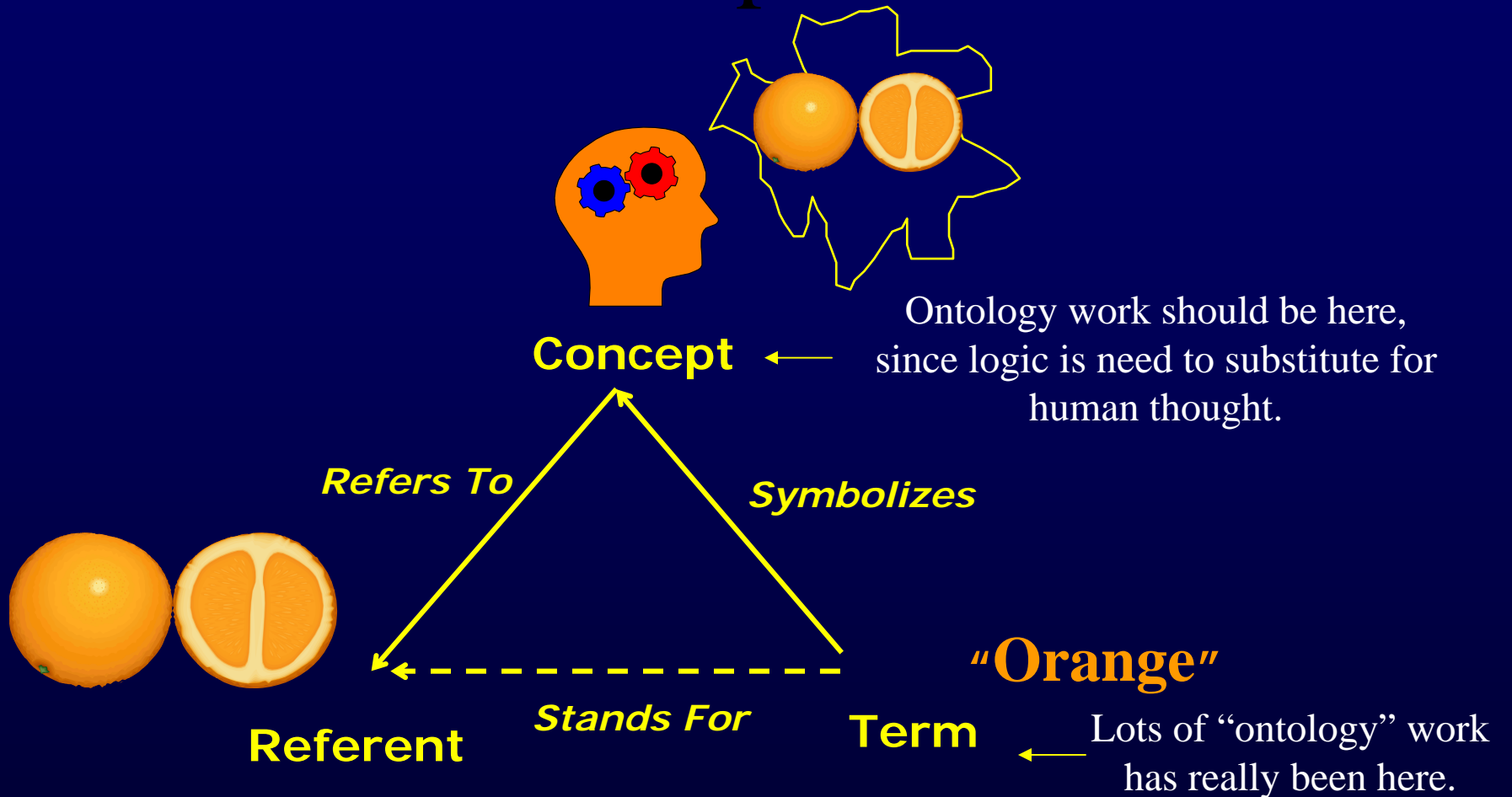
Presentation Contents

- **Ontology**
- Suggested Upper Merged Ontology
- WordNet

Language Topics

- **Pragmatics**
- **Semantics**
- Syntax
- Phonology/Phonetics

Terms and Concepts



Slide adapted from (c) Key-Sun Choi for Pan
Localization 2005

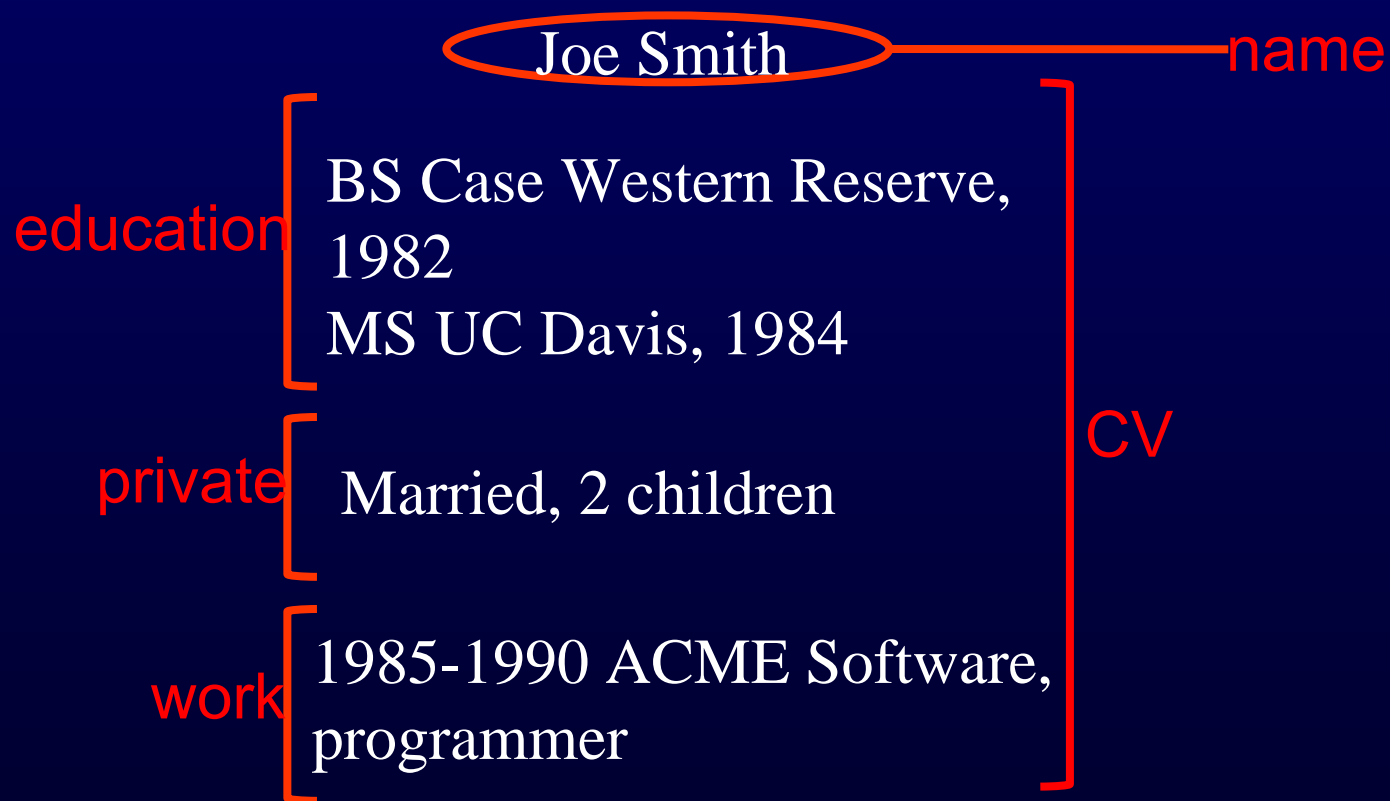
from the slide of [Bargmeyer, Bruce, Open Metadata Forum,
Berlin, 2005]

C.K. Ogden/I.A. Richards, *The Meaning of Meaning*
A Study in the Influence of Language upon Thought and The Science of Symbolism
London 1923, 10th edition 1969

Application

- Semantic word sense disambiguation
 - “The board approved the pay raise.”
 - Piece of wood, or corporate government?
- Anaphoric resolution
 - “Betty saw Susan asleep on the couch. She put her to bed.”
 - Sleeping people do not perform intentional actions

Imagine...your view of the web



...and the Computer's View

(assuming you don't read Chinese)

岳飛，字鵬舉，相州湯陰人也

name

education

少負氣節，沈厚寡言。天資敏悟，強記書傳，尤好左氏春秋及孫吳兵法。家貧，拾薪為燭，誦習達旦不寐。學射於周同。同射三矢，皆中的，以示飛；飛引弓一發，破其筈；再發，又中。同大驚，以所愛良弓贈之。飛由是益自練習，盡得同術。

private

生時，有大禽若鵠，飛鳴室上，因以為名。未彌月，河決內黃，水暴至，母姚氏，抱飛坐巨甕中，衝濤乘流而下，及岸，得不死。未幾，同死，飛悲慟不已。每值朔望，必具酒肉，詣同墓，奠而泣；又引同所贈弓，發三矢，乃酹。

work

父知而義之，撫其背曰：「使汝異日得為時用，其殉國死義乎？」應曰：「惟大人許兒以身報國家，何事不可為？」

CV

But wait, we've got XML -

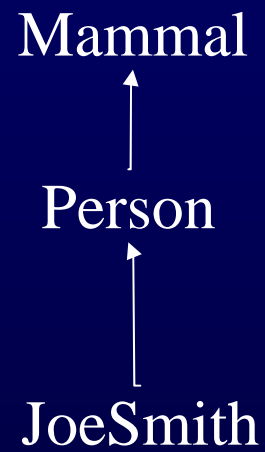
```
<job name="Joe Smith" title="Programmer">
```

But wait, we've got XML -

```
<job name="Joe Smith" title="Programmer">
```

```
<x83 m92="|||||||" title=".....">
```

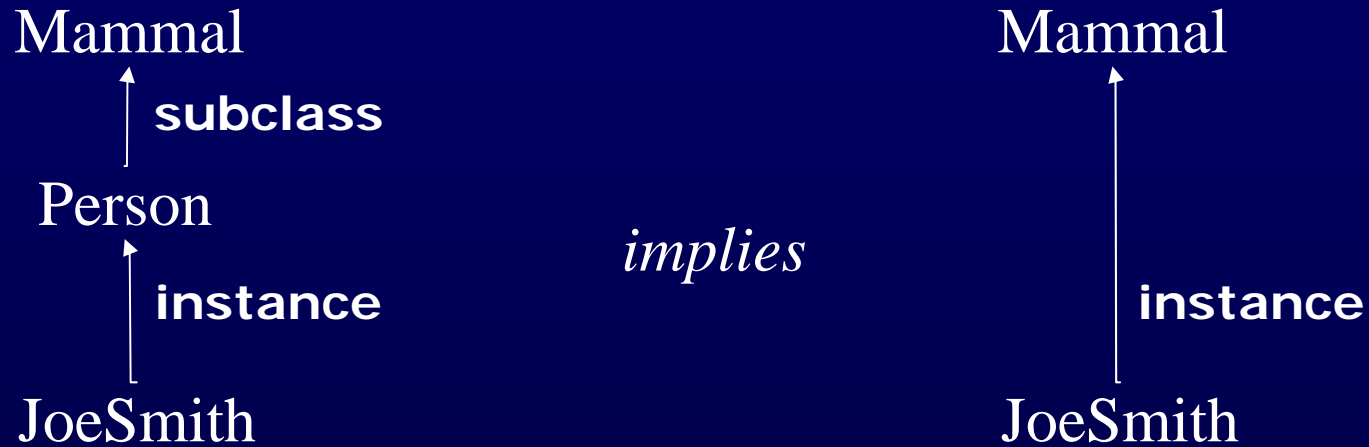
But wait, we've got Taxonomies -



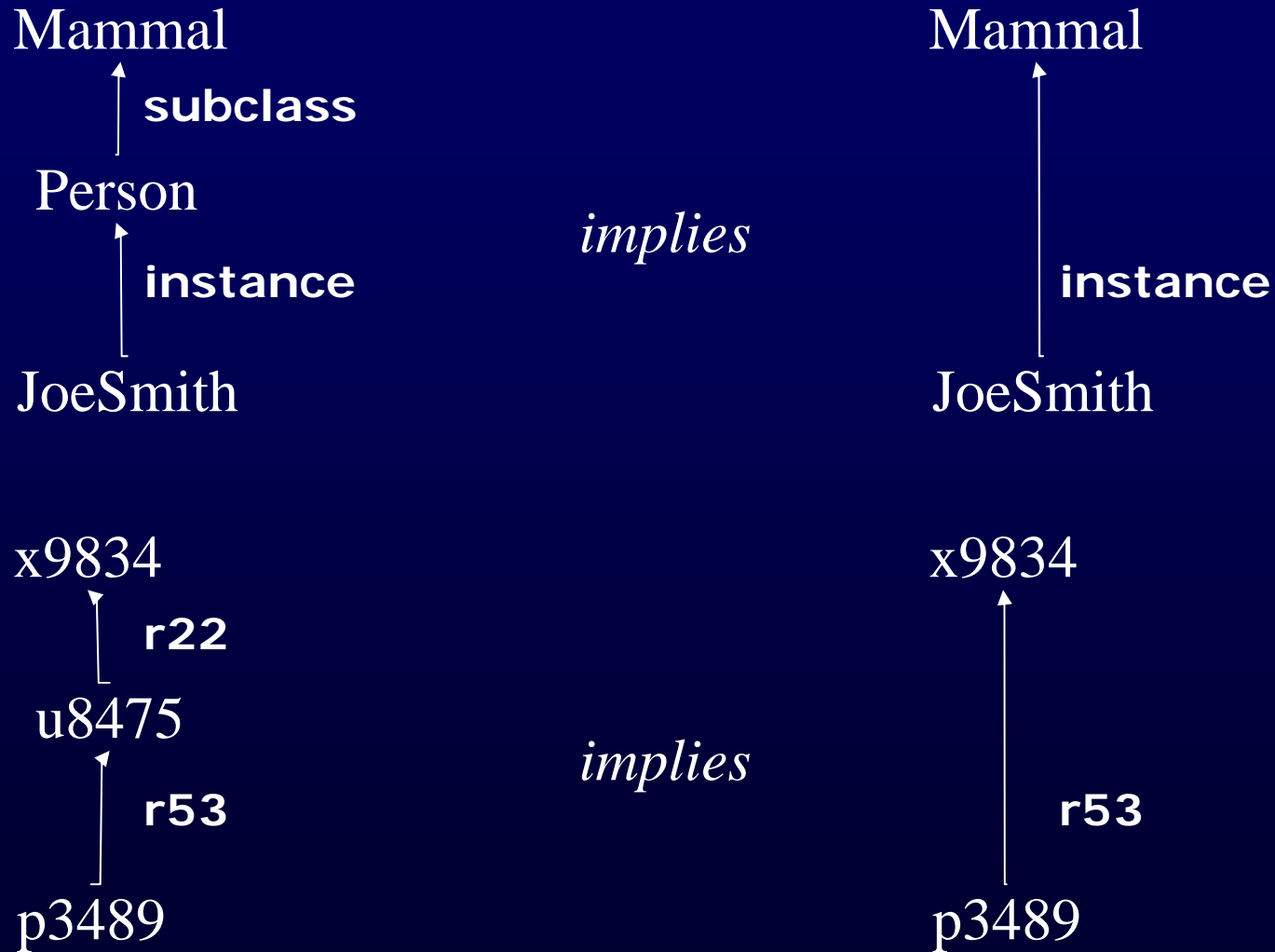
But wait, we've got Taxonomies -



Wait, we've got semantics -



Wait, we've got semantics -



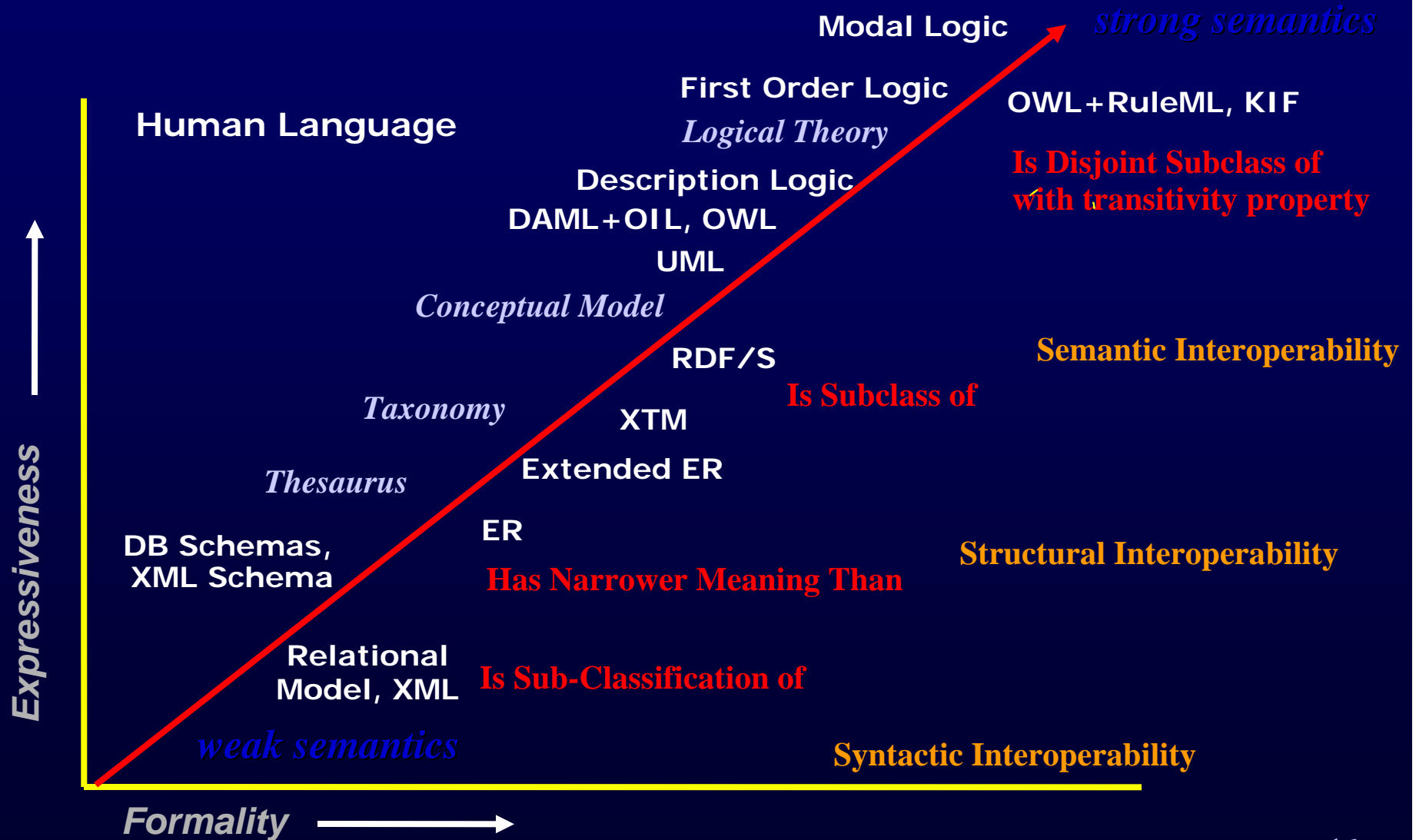
Semantics Helps a Machine Appear Smart

- A “smart” machine should be able to make the same inferences we do
- (let's not debate the AI philosophy about whether it would actually be smart)

Definitions

- An ontology is a shared conceptualization of a domain
- An ontology is a set of definitions in a formal language for terms describing the world

Language Formality & Expressiveness



Upper Ontology

- An attempt to capture the most general and reusable terms and definitions

Ontology vs Language and Knowledge

Ontology

- Expandable
- language independent
- machine understandable

Language

- understood by humans
- ambiguous

Knowledge

- changes rapidly
- may be local to an entity

Reuse

- “Cathedral and bazaar”
 - Structure needed to make new work coherent
- The world doesn't need yet another ontology of time, unconnected to any other ontology
- Many possible upper ontologies that are good enough
 - Network effects encourage convergence on one or small number
- Few people write their own operating system kernel or SQL engine

Open Source

- There's too much knowledge for any one entity to capture and code it
- Network effort – the more people that use the ontology, the more valuable it is
 - Needed to remove barrier to adoption
- Can't anticipate how it will be used
 - Testing theories of linguistic analogy!

Example: Why Use Upper Ontology?

- "1-800-Flowers" and "Harley-Davidson Motorcycles"
- Both need to send information to FedEx
- Even these entities have some general information in common
 - Weight, dates, dimensions, etc
- The alternative is mapping
 - But mapping ontologies is as hard or harder than authoring ontologies
 - Unless differences are glossed over, as with mapping vocabularies

Presentation Contents

- Ontology
- **Suggested Upper Merged Ontology**
- WordNet

Suggested Upper Merged Ontology

- 1000 terms, 4000 axioms, 750 rules
- Mapped by hand to all of WordNet 1.6
 - then ported to 2.0
- A “starter document” in the IEEE SUO group
- Associated domain ontologies totalling 20,000 terms and 60,000 axioms
- Free
 - SUMO is owned by IEEE but basically public domain
 - Domain ontologies are released under GNU
 - www.ontologyportal.org

SUMO (continued)

- Formally defined, not dependent on a particular implementation
- Open source toolset for browsing and inference
 - <https://sourceforge.net/projects/sigmakee/>
- Many uses of SUMO (independent of the SUMO authors and funders)
 - <http://www.ontologyportal.org/Pubs.html>

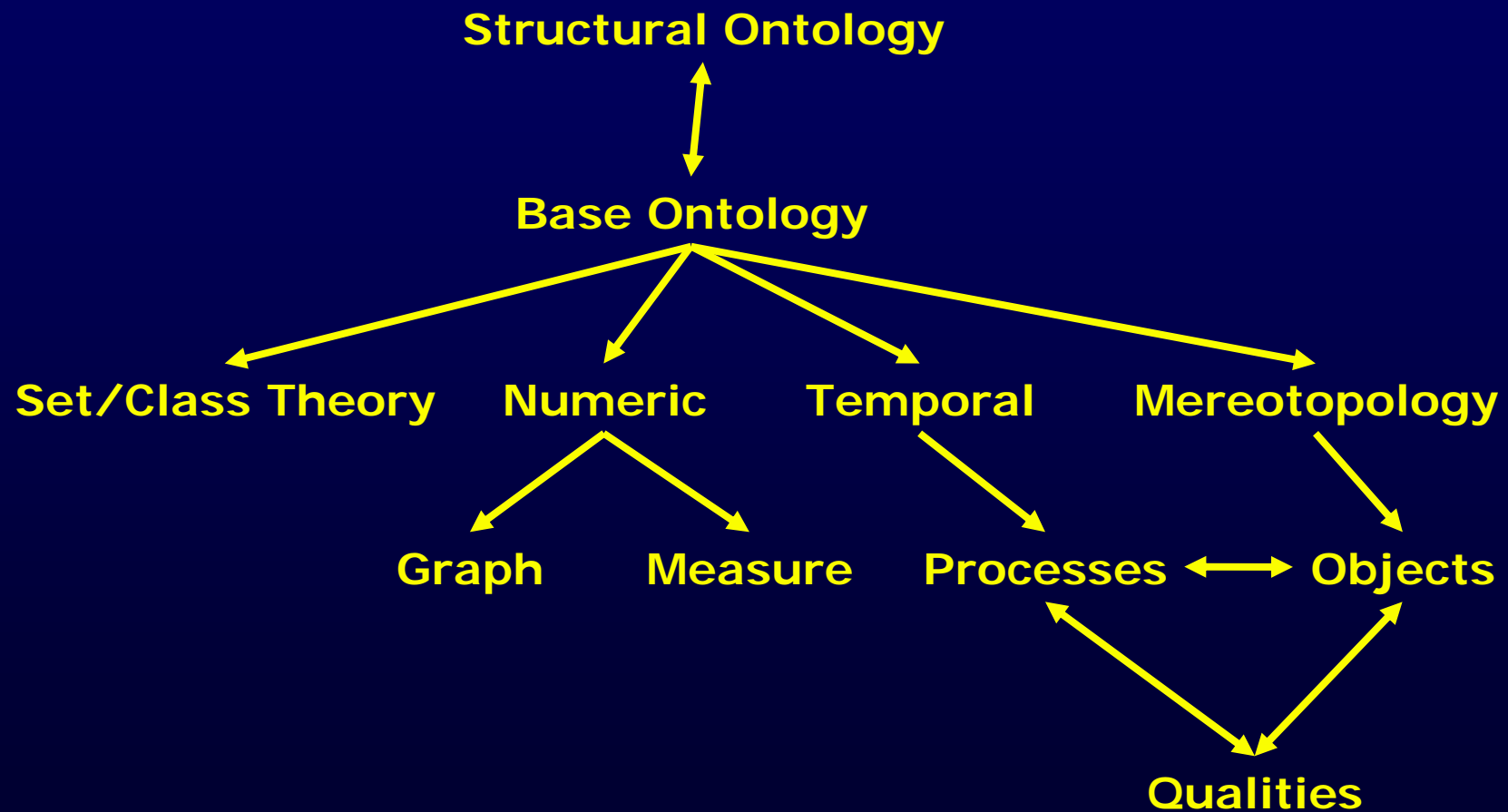
SUMO Validation

- Mapping to all of WordNet lexicon
 - A check on coverage and completeness (at a given level of generality)
- Peer review
 - Open source since its inception
- Formal validation with a theorem prover
 - Free of contradictions (within a generous time bound for search)
- Application to dozens of domain ontologies
- Translation of SUMO paraphrases to diverse multiple languages
 - Some confidence there's no cultural or linguistic bias
 - Chinese, Hindi, Tagalog, Czech, German, Italian

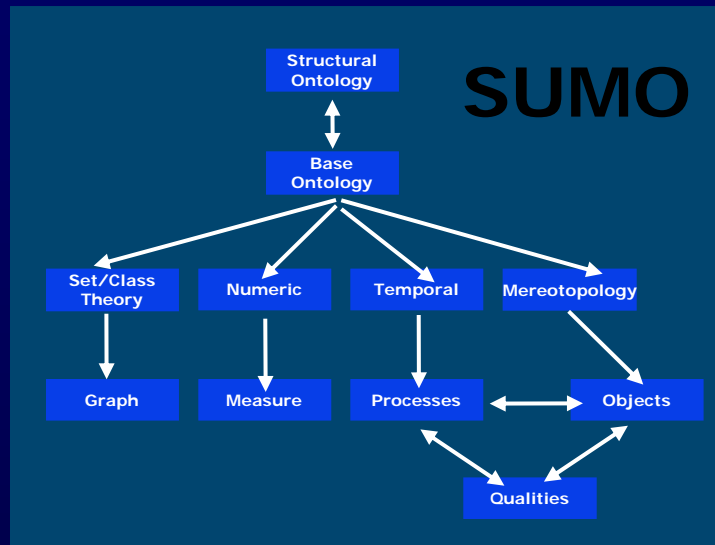
WordNet

- Lexical database
- 100,000 word senses – synsets
- Created by George Miller's group at Princeton
- Free
- De facto standard in the linguistics world

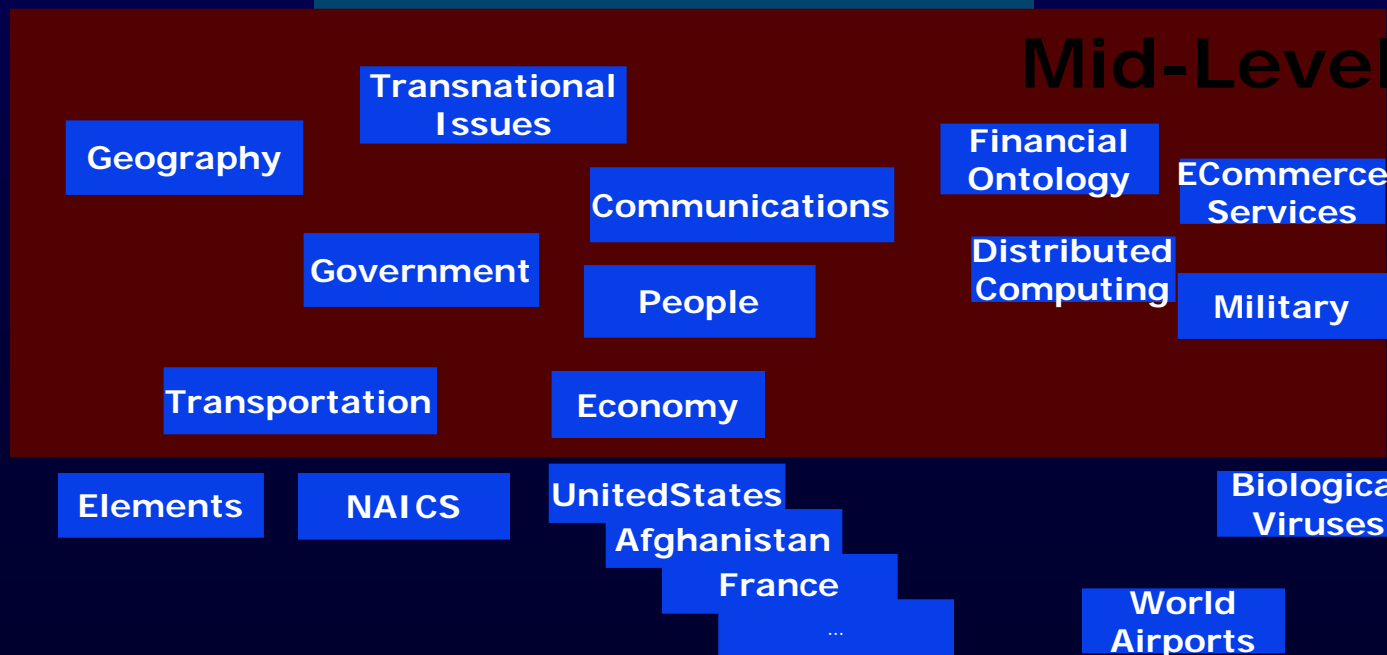
SUMO Structure



SUMO+Domain Ontology



Total Terms 20399
Total Axioms 67108
Rules 2500



Compliance

- “legal” compliance doesn’t realize full advantages of a technology
- Be careful about “Writing C in C++.”

```
<owl:Class rdf:ID="Entity">  
    <rdfs:subClassOf rdf:resource="PipeWrench"  
/>  
</owl:Class>
```

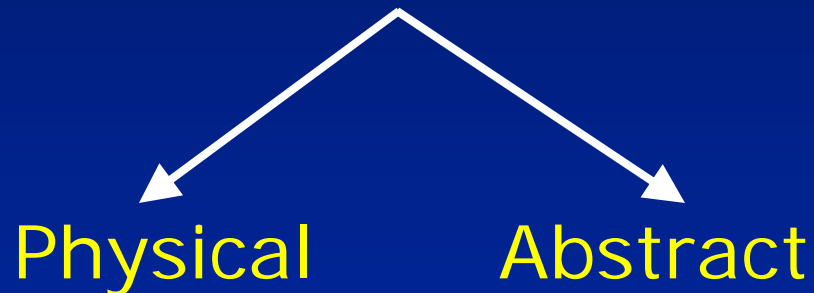
- If the only relation reused is subClassOf and everything inherits from Entity it’s not much of an ontology
- Always possible to use a good technology badly

Are SUMO Terms Directly Usable?

- Yes.
- Study – 1/3 of upper ontology terms directly appear in answers on large test
 - Cohen, P., Chaudhri, V., Pease A., and Schrag, R. (1999), Does Prior Knowledge Facilitate the Development of Knowledge Based Systems, In Proceedings of the Sixteenth National Conference on Artificial Intelligence (AAAI-1999). Menlo Park, Calif.: AAAI Press.
<http://home.earthlink.net/~adampease/professional/cohen-aaai99.ps>
- before (in time), agent (of a process), etc.

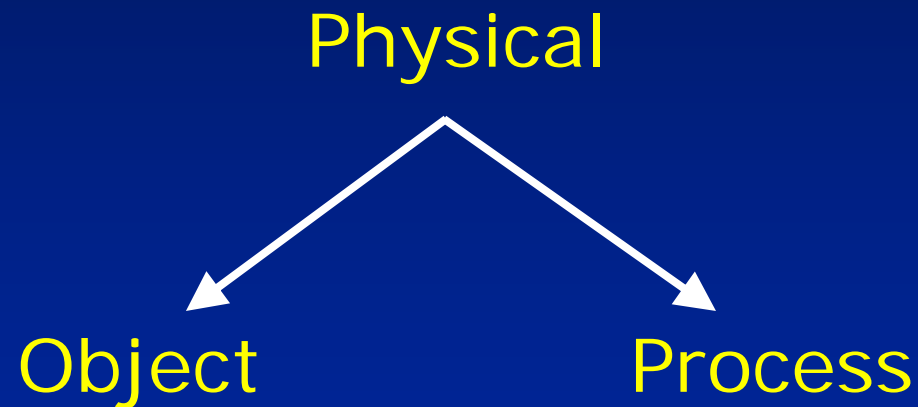
High Level Distinctions

The first fundamental distinction is that between 'Physical' (things which have a position in space/time) and 'Abstract' (things which don't)



High Level Distinctions

Partition of 'Physical' into 'Objects' and 'Processes'



Objects

Object

SelfConnectedObject

Substance

CorpuscularObject

Region

Collection

Processes

DualObjectProcess

Substituting
Transaction
Comparing
Attaching
Detaching
Combining
Separating

InternalChange

BiologicalProcess
QuantityChange
Damaging
ChemicalProcess
SurfaceChange
Creation
StateChange

ShapeChange

IntentionalProcess

IntentionalPsychologicalProcess
RecreationOrExercise
OrganizationalProcess
Guiding
Keeping
Maintaining
Repairing
Poking
ContentDevelopment
Making
Searching
SocialInteraction
Maneuver

Motion

BodyMotion
DirectionChange
Transfer
Transportation
Radiating

Abstract

SetOrClass

Relation

Proposition

Quantity

Number

PhysicalQuantity

Attribute

Graph

GraphElement

Simple Language Generation

- Term translation
- Relation templates
 - Use C-like printf statements
- Result is awkward but usually grammatical
- Preserves deep meaning
- English (Sevcenko), German (Wulf), Czech (Sevcenko), Italian (Ulivieri & Molino), Hindi (IIT Bombay), Chinese (Academica Sinica)

Generation Examples

- water – wasser – acqua
- Atlantic ocean – Atlantik - oceano Atlantico
- (material Water AtlanticOcean)
- %2 is %n &%made of %1
- %2 è %n &%fatto di %1
- %1 ist von %2 %n{nicht} gebildet
- Atlantic ocean is made of water.
- Oceano Atlantico è fatto di acqua.
- Atlantik ist von Wasser gebildet.

Student project idea

- Translations for Korean, Oriya, Dzongkha, Khmer, Urdu, Pashto etc.
- 1000 terms, 400 relation templates
- Requires a truly bilingual translator
- Can be done in a few days, but a few weeks is more typical

Presentation Contents

- Ontology
- Suggested Upper Merged Ontology
- **WordNet and GlobalWordNet**

Global WordNet

- Language resources that use same tools and file format as Princeton WordNet
- Synonyms in different languages are linked together
- Benefits
 - Leverage tools
 - Build cross-linguistic resources
 - Leverage linkage to language-independent ontology
- www.globalwordnet.org

WordNet to SUMO Mapping

- WordNet synset "plant, flora, plant_life" is equivalent to the formal SUMO term 'Plant'
 - 00008864 03 n 03 plant 0 flora 0 plant_life 0 027@ . . . | a living organism lacking the power of locomotion &%Plant=
 - SUMO has axioms that explain formally what a plant is

```
(=>
  (and
    (instance ?SUBSTANCE PlantSubstance)
    (instance ?PLANT Organism)
    (part ?SUBSTANCE ?PLANT))
    (instance ?PLANT Plant))
```

WordNet to SUMO Mapping

- Many highly specific words map to general formal terms
- Several word senses may map to one SUMO term and vice versa
 - 00128951 04 n 02 substitution 0 exchange 1 004 @ 00125689 n 0000 ~ 00129213 n 0000 ~ 00129804 n 0000 ~ 00129915 n 0000 | the act of putting one one thing or person in the place of another: "he sent Smith in for Jones but the substitution came too late to help" &%Removing+ &%Putting+

WordNet to SUMO Mapping

- Most nouns map to classes
- Most verbs map to subclasses of Process
- Most adjectives map to a SubjectiveAssessmentAttribute
- Most adverbs map to relations of &%manner

Controlled English to Logic Translation (CELT)

- Specify an unambiguous language that is as close to English as possible
- Keep it completely general purpose

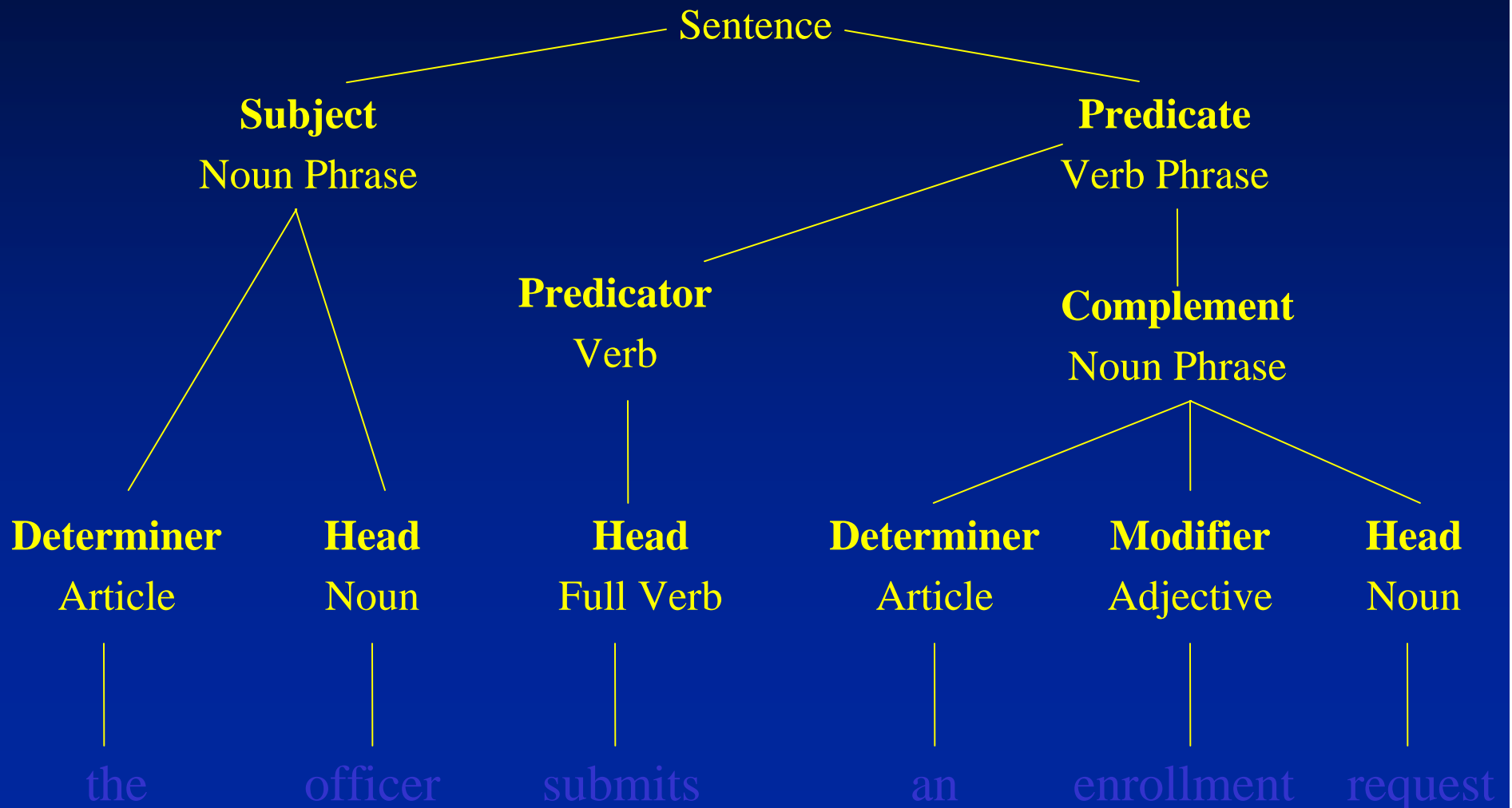
Language Understanding – Phrase Mapping (proposed work)

- *“John takes a walk.”*
- *[John, subject][takes a walk, VP template 547]*

```
(exists (?walk <subject>)  
  (and  
    (instance ?walk Walking)  
    (agent ?walk <subject>)))
```

```
(exists (?walk ?john)  
  (and  
    (instance ?walk Walking)  
    (instance ?john Human)  
    (names "John" ?john)  
    (agent ?walk ?john)))
```

CELT: Grammar and an Example



CELT: Examples of Sentences

- Simple sentence
 - * The student enrolls in a class.
- Composite sentence
 - * The student walks to class and opens a book.
- if-then sentence
 - * If the student is late then he fails the assignment.
- Possessives
 - * John's class is difficult.
- Anaphora
 - * John enrolls in the class. He studies diligently.
- Quantifiers
 - * Every farmer owns a horse.

CELT: What is not Allowed?

•Restrictions

- * active voice
- * indicative mood
- * simple present tense
- * 3rd person singular
- * no plural verbs
- * no modals

•Not allowed

- * passive voice
- * imperatives, subjunctives
- * past, future, ongoing
- * 1st or 2nd person
- * plural verbs or nouns
- * may, can, must

Proposed Work – Detail

- Possessives
 - John's arm...
 - (part John Arm1)
 - John's car...
 - (possesses John Car1)
- Prepositional phrases
 - John gets in the car.
 - (destination Transfer1 Car1)
 - John gets on the bus.
 - (destination Transfer1 Bus1)

More Examples

- Copula template forms

- A dog is a mammal

```
(=>
  (instance ?X Canine)
  (instance ?X Mammal))
(subclass Canine Mammal)
```

better



- A dog is eating

```
(exists (?X ?E)
  (and
    (instance ?X Canine)
    (instance ?E Eating)
    (agent ?E ?X)))
```

More Examples

–A dog is brown




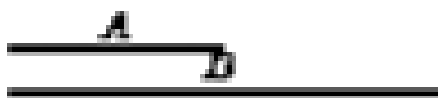
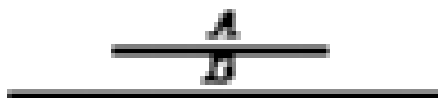
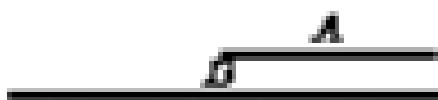
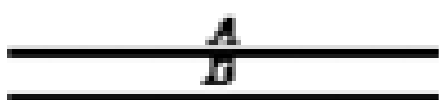
```
(exists (?X)
  (and
    (instance ?X Canine)
    (attribute ?X BrownColor)))
```

Backup Slides

Theory details

Temporal Relations

- Enables one to talk about two events without assigning specific times to the endpoints

$A < B$ $B > A$		A before B B after A
$A m B$ $B m i A$		A meets B B met by A
$A o B$ $B o i A$		A overlaps B B overlapped by A
$A s B$ $B s i A$		A starts B B started by A
$A d B$ $B d i A$		A during B B contains A
$A f B$ $B f i A$		A finishes B B finished by A
$A = B$		A equals B

“3d” vs “4d” representation

- Addressing change and identity
 - Also called “endurantism vs. perdurantism”
- Am I the same person I was yesterday?
 - Yes and no
 - What are essential parts of my identity?
- 3d represents things with attributes that may change over time
- 4d represents temporal parts with unchanging attributes

3d vs 4d example

- Joe loses an arm at a point in time which we'll call "t"
- JoeBeforeT, JoeAfterT, JoePartOtherThanHisArmBeforeT
- Before amputation we have both Joe and the part of Joe without his arm
- Endurantists also believe that Joe survives the loss of his arm with his identity intact
- After the amputation we have only a Joe without his arm
- Joe is the same as both Joe before the amputation
- Joe after the amputation is the same as the spatial part of Joe minus his arm before the amputation
- Both can't be true

Semiotics

- The study of signs and symbols
- Important for issues like intellectual property law
- An edition of the printed play Hamlet
- A copy of an edition of Hamlet
- A performance of Hamlet
- A performance of Hamlet captured on video
- The timeless informational content of the play Hamlet

...formally

- (refers ?OBJECT1 ?OBJECT2)
 - (represents ?THING ?ENTITY)
 - (realization ?PROCESS ?PROP)
 - (containsInformation ?OBJECT ?PROP)
- (subsumesContentInstance ?OBJECT1 ?OBJECT2)
- (subsumesContentClass ?OBJECT1 ?OBJECT2)
- (equivalentContentInstance ?OBJECT1 ?OBJECT2)
- (equivalentContentClass ?OBJECT1 ?OBJECT2)

Implementation is Different from Representation

- Why lose meaning at design time just because of runtime issues?
 - We can't reason with English definitions, but that doesn't mean we shouldn't document our terms
- Many different implementations may be done from the same representation
- This does not mean that run time issues should be ignored at design time
 - If you represent information you know can't be reasoned with, it better not be essential in most conceivable applications

Many Ways to Use Ontology

- As an information engineering tool
 - Create a database schema
 - Map the schema to an upper ontology
 - Use the ontology as a set of reminders for additional information that should be included
- As more formal comments
 - Define an ontology that is used to create a DB or OO system
 - Use a theorem prover at design time to check for inconsistencies
- For taxonomic reasoning
 - Do limited run-time inference in Prolog, a description logic, or even Java
- For first order logical inference
 - Full-blown use of all the axioms at run time

Technology Deployment

	Design Time	Run Time
<ul style="list-style-type: none">• As an information engineering tool	Text Editor Ontology Browser	Java, Oracle etc.
<ul style="list-style-type: none">• As more formal comments	+Theorem prover/ Inference engine	
<ul style="list-style-type: none">• For taxonomic reasoning		+DL or Prolog
<ul style="list-style-type: none">• For first order logical inference		+Theorem prover/ Inference Engine