

# Ontologizing Senses

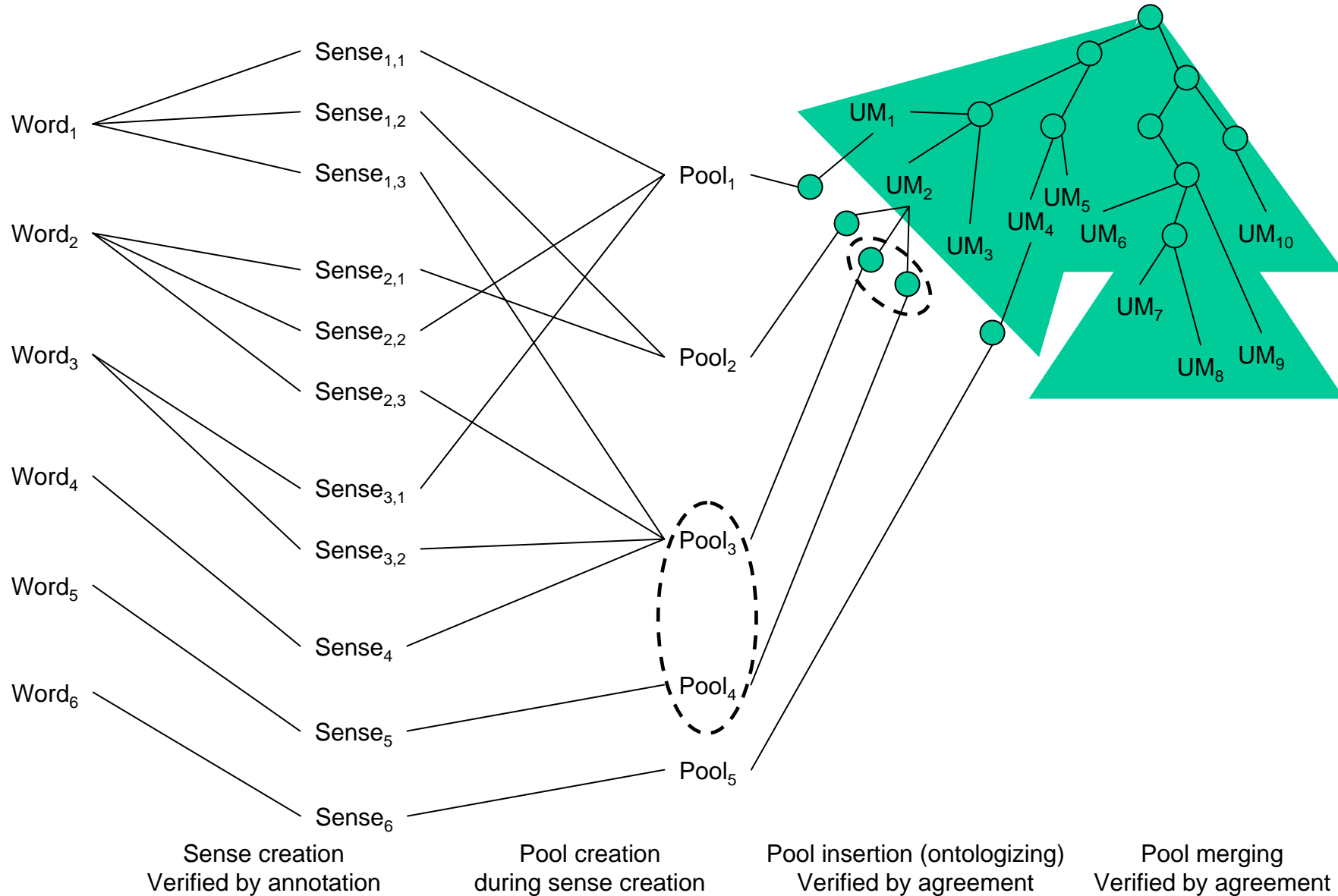
2009-02-24

Ed Hovy





# Idea and Approach





# Example Sense Pools



<pre> &lt;SENSEPOOL SPID="Weapon - instrument used for combat"&gt;   &lt;SENSE&gt;     &lt;SENSEID&gt;weapon.a.n.1&lt;/SENSEID&gt;     &lt;SENSEID&gt;weapon.o.n.1&lt;/SENSEID&gt;     &lt;SENSEID&gt;weapon.o.n.3&lt;/SENSEID&gt;     &lt;SENSEID&gt;arsenal.o.n.1&lt;/SENSEID&gt;     &lt;SENSEID&gt;arm.a.n.2&lt;/SENSEID&gt;     &lt;SENSEID&gt;arm.o.n.7&lt;/SENSEID&gt;   &lt;/SENSE&gt;   &lt;SUBTO&gt; ARTIFACT &lt;/SUBTO&gt;   &lt;FEATURE&gt;     &lt;FEATURETAG&gt;+instrument&lt;/FEATURETAG&gt;     &lt;FEATURETAG&gt;+concrete&lt;/FEATURETAG&gt;     &lt;FEATURETAG&gt;+fighting&lt;/FEATURETAG&gt;     &lt;FEATURETAG&gt;+capability&lt;/FEATURETAG&gt;   &lt;/FEATURE&gt;   &lt;RELATION&gt; &lt;/RELATION&gt;   &lt;COMMENTARY&gt; An instrument of fighting or combat.   &lt;/COMMENTARY&gt; &lt;/SENSEPOOL&gt; </pre>	<pre> &lt;SENSEPOOL SPID="Barrier - artifact situated to prevent access"&gt;   &lt;SENSE&gt;     &lt;SENSEID&gt;barrier.a.n.1&lt;/SENSEID&gt;     &lt;SENSEID&gt;barrier.o.n.1&lt;/SENSEID&gt;     &lt;SENSEID&gt;barrier.o.n.2&lt;/SENSEID&gt;     &lt;SENSEID&gt;obstruction.a.n.1&lt;/SENSEID&gt;     &lt;SENSEID&gt;obstruction.o.n.3&lt;/SENSEID&gt;     &lt;SENSEID&gt;impediment.o.n.4&lt;/SENSEID&gt;   &lt;/SENSE&gt;   &lt;SUBTO&gt; ARTIFACT &lt;/SUBTO&gt;   &lt;FEATURE&gt;     &lt;FEATURETAG&gt;+PHYSICAL&lt;/FEATURETAG&gt;     &lt;FEATURETAG&gt;+ARTIFACT&lt;/FEATURETAG&gt;     &lt;FEATURETAG&gt;-MOVEMENT&lt;/FEATURETAG&gt;     &lt;FEATURETAG&gt;-ACCESS&lt;/FEATURETAG&gt;     &lt;FEATURETAG&gt;+INTENTIONAL&lt;/FEATURETAG&gt;   &lt;/FEATURE&gt;   &lt;RELATION&gt;     &lt;RELATIONTAG&gt;impediment.a.n.1&lt;/RELATIONTAG&gt;   &lt;/RELATION&gt;   &lt;COMMENTARY&gt; A physical artifact that impedes or restricts     free movement, or denies access to some area.   &lt;/COMMENTARY&gt; &lt;/SENSEPOOL&gt; </pre>
--	---



# Possible Uses of Ontology-Based Information



- **Pool contains multiple words' senses:**
  - e.g., *barrier.a.n.1*, *obstruction.a.n.1*, *impediment.o.n.3*
  - (Quasi-)synonyms extends training data for WSD, etc.
- **Pool contains Chinese and Arabic wordsenses:**
  - e.g., *xin1hua2she4*, *da4xue2*
  - Ditto, plus may help MT at sense-aligned level
- **Pools arranged under Upper Model:**
  - e.g., different kinds of Artifacts
  - Closely similar yet different pools still share some type characteristics; perhaps useful for metonymy or gross WSD
- **Pool contains features:**
  - e.g., *+PHYSICAL*, *+CONCRETE*, *-MOVEMENT*
  - Eventually may be used for semantic role labeling/assignment
  - Features will be inherited



- **Objects (for noun pools): Stable at about 128 concepts**
- **Events (for verb pools): Still growing, now about 20 concepts**
- **Pool attachment:**
  - Each UM concept is potential ‘attachment point’ for sense pools
  - Most pools attach at UM fringe: about 70 for Objects and 15 for Events
- **Planned eventual structure:**
  - Fairly dense clusters of pools under UM attachments, no deep taxonomies
  - Many uplinks for event pools, fewer for objects
  - Pools within a cluster share some features, and have some unique ones
  - Features: *+/-Animate, +/-Concrete*, etc.



# Pool Ontologizing (English)



- Each pool seen by between 3 and 6 people independently
- Weekly or biweekly telecons and online discussion
- 200 hard cases done by experts
- Accepted only UM attachments with majority agreement

Noun sense	UM choice 1			# people	ITA	UM choice 2	# people	ITA	UM choice3	# people	ITA
channel.y.n.7	3	4	Artifact	3	100%	NaturalNonLivingObjec	3	100%	LocationOf*NaturalNor	3	100%
club.y.n.7	3	2	LocationOf*Artifact	3	100%	CommercialOrganizati	3	100%			
colony.y.n.2	3	2	SocialGathering	3	100%	MeasurementQuantity	3	100%			
colony.y.n.5	3	2	SpatialLocation	3	100%	SocialAbstraction	3	100%			
colony.y.n.3	3	2	SocialGathering	3	100%	MeasurementQuantity	3	100%			
column.y.n.7	3	2	ShapeAndStructureA	3	100%	GeneralizedLanguageC	3	100%			
compartment.	3	3	Artifact	3	100%	PartOf*Artifact	3	100%	ShapeAndStructureAb	3	100%
core.y.n.8	3	2	Human	3	100%	NonProfessionRole	3	100%			
country.y.n.6	3	2	LocationOf*NaturalN	3	100%	SpatialLocation	3	100%			
crash.y.n.5	3	2	ThoughtProcess	3	100%	EventAsObject	3	100%			
cubicle.y.n.5	3	2	Artifact	3	100%	ShapeAndStructureAb	3	100%			
deal.y.n.8	3	2	PsychoSocialAbstrac	3	100%	EventAsObject	3	100%			
deal.y.n.5	3	2	Artifact	3	100%	SetOf*Artifact	3	100%			
deal.a.n.7	3	2	PsychoSocialAbstrac	3	100%	QualityAsObject	3	100%			
debate.y.n.1	3	2	EventAsObject	3	100%	GeneralizedLanguageC	3	100%			
negotiation.y.r	3	2	EventAsObject	3	100%	GeneralizedLanguageC	3	100%			
decline.y.n.5	3	3	Artifact	3	100%	NaturalNonLivingObjec	3	100%	ShapeAndStructureAb	3	100%
department.y.	3	5	PoliticoGovernmentC	3	100%	EducationalOrganizati	3	100%	MilitaryOrganization	3	100%
department.y.	3	3	PoliticoGovernmentC	3	100%	EducationalOrganizati	3	100%	CommercialOrganizati	3	100%
design.y.n.6	3	3	Artifact	3	100%	ShapeAndStructureAb	3	100%	GeneralizedLanguageC	3	100%
door.y.n.4	3	3	PartOf*Artifact	3	100%	LocationOf*Artifact	3	100%	SpatialLocation	3	100%
plunge.y.n.2	3	2	ShapeAndStructureA	3	100%	EventAsObject	3	100%			
drop.y.n.5	3	2	NaturalNonLivingObj	3	100%	ShapeAndStructureAb	3	100%			



- **Task:** Merge incoming pool (derived from English monosemous or Arabic or Chinese noun) into appropriate pool, or create new pool
- **Approach:**
  - Reduce candidates by automated initial assignment to likely pools (heuristics: definition overlaps, WordNet links, etc.)
  - Interface: CAT display tool too cumbersome
  - Developed dedicated interface
  - Speed: fast, over 1 per minute
- **Results:**
  - English monosemous: 4 annotators, working on initial 2800 terms (covers current corpus); another 7500 in reserve to be done later
  - Chinese: 2 annotators, completed (but not checked) about 300 Chinese pools
  - Arabic: 2 annotators, work underway



# Pool Insertion/Merging Interface



**ONTONOTES Pool Insertion Interface**  
version 1.4

**chinese**

**Noun Pool: Z1341**  
Lemma: di4wei4\_1-n / 地位  
noun  
Pool Sense(s):  
Sense: di4wei4.r.n.1 (di4wei4 (1 of 3))  
Name: position 地位, 身  
Group: 1 N: 1  
Examples: 他正处于有地位。  
He is in a favorable position.  
他是个有地位的人。  
He is a man of position.

**english**

**Noun Pool P2582**  
??? (= Sumnum Genus)  
• Sumnum Genus  
• The root of the ontology: everything  
Definition(s):  
1. Type - printed character symbols  
2. printed characters, written symbols with specific print traits, e.g., font, size, darkness.  
Local features(s):  
[+entity] [+symbol] [+character] [+printed] [+WRITING]  
Pool sense(s):  
print.e.a.n.1  
• written typographical text in a document  
• PRINT[+entity][+physical][+form][+written][+typographical][+text]  
print.e.o.n.5 (= print-fine print)  
• the result of the printing process  
type.e.a.n.4  
• printed characters  
• TYPE[+entity][+symbol][+character][+printed] printed characters, written symbols of language  
type.e.o.n.5 (= type-face)  
• printed characters  
Other names for this word:

**Logged in as: philpot**

**insertion choices and memos**  
Submit all decisions

P2022		
P2023		
P2025		
P2582		
P2939		
Artifact		
GeneralizedLanguageObject		
Human		
MathematicalObject		
NonProfessionRole		
PartOfArtifact		

**GESTURES:** Tab=Next Pool || Shift-Tab=Previous Pool || Right Arrow=Write Memo || Left Arrow=Return to Pool || Mouse Click= Revisit Pool  
**CODES:** I=identical || C=close but not exact || <= LH column less general than RH column || >= LH column more general than RH column || 0=none || Z=problem

Candidate pools for insertion or merging

Details of each candidate pool

Pool to be inserted or merged

Insertion or merging operations



- **Upper Model**
  - **Objects (from nouns): Stable at about 128 concepts**
  - **Events (from verbs): Some changes, now about 25 concepts**
  
- **Ontologizing English sense pools under Upper Model**
  - **Pools from polysemous nouns:**
    - **About 2100 pools linked so far**
    - **Pools cover about 5300 senses, which (partly) cover about 970 nouns**
    - **Each pool derived from polysemous English nouns**
    - **200 hard cases done by experts; all other linking decisions done by 3 or more people**
  - **Pools from polysemous verbs (at Columbia):**
    - **Over 670 pools linked so far (over 700 created)**



- **Pools from English monosemous nouns:**
  - WordNet senses to UM attachments: Unambig links: 26.2K; ambig links: 895; no link: 10.8K
  - Manual linking/merging of 2800 corpus mono nouns underway
- **Chinese noun pools:**
  - Pool creation and ontologizing: About 300 pools completed
  - Merging with English-derived pools:
    - Approx. 1200 links created
    - 87 links with perfect agreement, 1137 links with no agreement
  - Chinese verb pools: about 100
- **Arabic noun pools:**
  - Pool creation and ontologizing: About 320 pools completed
  - Merging with English-derived pools:
    - Approx. 200 links created



# Longer-Range Plans

---



- **Continue ontologizing and merging noun- and verb-derived pools for English, Chinese, Arabic**
- **Finalize taxonomization of pools under each UM node**
- **Regularize features within pools:**
  - Currently 5000 features, about 2000 high-freq
- **Finalize links from pools back to WordNet**
- **Add information from other resources into UM (funded by DoD):**
  - Domain labels from NSA
  - Axioms from SUMO
  - Modalities



# BACKUP



# Feb 09 Noun Annotation Status — Internal



	English	Chinese
Total noun types double-annotated	1311	200
Total noun types adjudicated	623	64
Total polysemous noun instances	141890	52284
Total noun instances double-annotated	81024 (57.10%)	9557 (18.28%)
Total noun instances adjudicated	44361	5514
Average agreement	0.93	0.95

Arabic
98
61
141890
10201 (7.19%)
6411
0.89



# Upper Model for Objects

(Work done at ISI)



## TangibleObject

Animal

NonVolitionalBiologicalObject

VolitionalNonBiologicalObject

NonVolitionalNonBiologicalObject

- **Animals: 7 leaf nodes**
- **Other living things: 5 leaf nodes**
- **Androids: 1 leaf node**
- **Artifacts and natural nonliving things: 6 leaf nodes**

## IntangibleObject

MentalObject

Collection

Relation

ImmeasurableObject

MeasurableAbstraction

- **Thoughts, emotions, etc.: 7 leaf nodes**
- **Sets, groups, organizations: 11 leaf nodes**
- **Relations between things: 11 leaf nodes**
- **Abstractions (information, procedures, functions, etc.): 12 leaf nodes**
- **Space, time, wealth, etc.: 12 leaf nodes**



# Upper Model for Events

(Work done at Colorado)



Eventuality1

Change

ChangeIntegralProperties

- **Change of core defining properties**

ChangeIncidentalProperties

- **Change of other properties: location, possession, etc.**

Motion

- **Motions**

MentalEvent

- **Thoughts**

PhysicalExperience

- **Bodily experiences**

SocialInteraction

- **Communication, interaction, judgment, etc.**

Eventuality2

Aspect

- **Aspect: completeness, continuity, etc.**

Modality

- **Modality: desires, wishes, obligations, etc.**



# Pool Ongtologizing Interface (CAT)



[Coding Analysis Toolkit] - Code Dataset

http://cat.ucsur.pitt.edu/app/codeDataset.aspx?id=307

Google MeetingMaker Wikipedia ISI home Hovy OntoNotes IE-summarization-... CKID DG MT Conferences Travel USC ISI Powerset

Sign in to Meeting Maker Dipdive.com [Coding Analysis Toolkit] - Cod...

## Coding Analysis Toolkit

A Service of the University of Pittsburgh's Qualitative Data Analysis Program

Main Menu Datasets Analysis Validations Reports Account Logout

Select one or more:

☒ (1) Human ☐ (2) NonHumanMammal ☐ (3) Amphibian  
☐ (4) Bird ☐ (5) Reptile ☐ (6) Fish  
☐ (7) Invertebrate ☐ (8) RoleOf\*Animal ☐ (0) None

Code Item

« Previous Next »

### Noun pool P2787

Definition(s):

1. Hunter - person who hunts game for food or recreation
2. a person who hunts game animals - for food or sports.

Local features(s):

[+PERSON] [+GAME\_KILLER] [+FOOD/+RECREATION]

Pool sense(s):

**hunter.o.n.1 (= |huntsman|)**

- someone who hunts game

16 paragraph(s) left to code

Done

- Choices from UM (defs at bottom)
- Item to annotate (def, features) with link to details