



Московский гуманитарный  
институт

ИМЕНИ  
Е.Р. Дашковой

1992

1885

Е.Р. Дашковой

# Lexicon Management in Ontological Semantics

**Max Petrenko**

**Princess Ekaterina R. Dashkova**

**Moscow Humanities Institute**

**RiverGlass, Inc.**



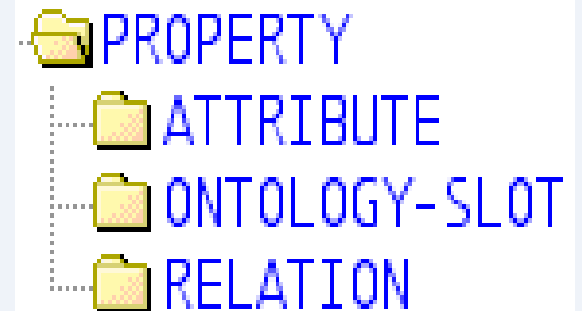
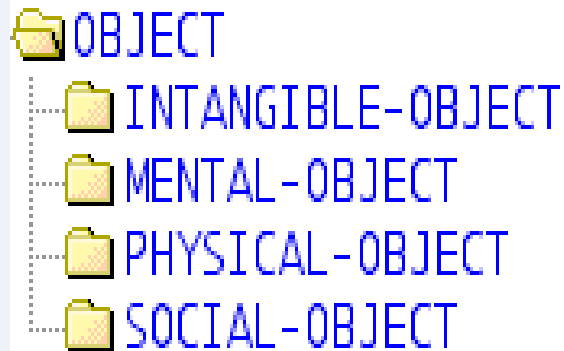
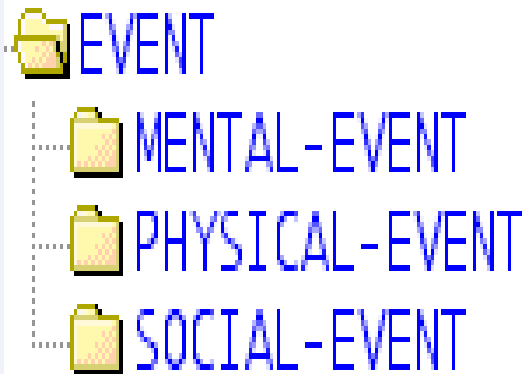
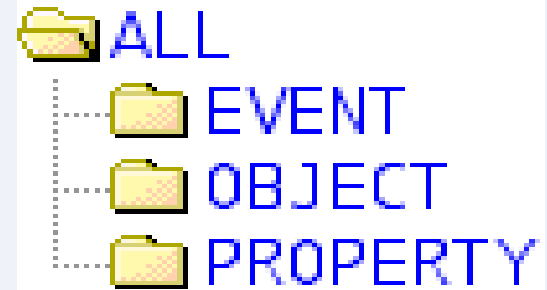
**RiverGlass**

# Structure

- The lexical and ontological knowledge resources
- Lexicon acquisition strategies and techniques
- Recalcitrant cases in lexicon acquisition

# Ontology: concept hierarchy

- ❑ language-independent concepts;
- ❑ universal and natural hierarchy;
- ❑ rich system of interrelations;
- ❑ properties inheritance mechanism:



# Ontology management

## Concept structure: BOX

```
(box  
  (definition (value("a rectangular container")))  
  (is-a (hier(container)))  
  (made-of  
    (sem(ceramic metal paper plastic wood)))  
  (shape(value(rectangular square)))  
)
```

# Ontology

(made-of

(definition (value("the relation  
between a thing and things made out of  
it"))))

(is-a (hier(physical-object-relation)))

(inverse (value(material-of)))

(domain (sem(**physical-object**)))

(range (sem(**material**)))

)

# Ontology: concept structure

**(physical-object**

**(definition (value("objects that physically  
exist"))))**

**(is-a (hier(object)))**

**(subclasses (hier( surface-feature  
landscape-object  
animate  
animate-part  
animal-artifact  
material  
artifact  
celestial-object))))**

**PHYSICAL-OBJECT**



*IS-A*

**ARTIFACT**



*IS-A*

**CONTAINER**



*IS-A*

**BOX**

**CERAMIC**

**METAL**

**PAPER**

**PLASTIC**

**WOOD**

***IS-A***

**MATERIAL**



# Lexicon management

## Lexicon

### Core features

- cannot transgress ontological restrictions
- can only further restrict ontological fillers
- carries syntactic and semantic info
- covers closed classes (prepositions, pronouns, etc)
- represents synonymy
- covers derivatives
- covers alternations

# Lexicon management

## Lexical entry template

(head-entry

(sense-1, 2, 3...

(cat(n/v/adj/pro/prep))

(synonyms "")

(anno(def ""))

(comments "Acquired by <acquirer name>  
on <date> at <time>.")

(ex ""))

(syn-struct((root(\$var0))(cat(n/v/adj/pro/prep))))

(sem-struct(CONCEPT))

))

# Lexicon management

(box

(box-n1

(cat(n))

(synonyms ""

(anno(def "a rectangular container")

(comments ""

(ex "the man put the book in the box"))

(syn-struct((root(\$var0))(cat(n))))

(sem-struct(box))

)

```
(tourist
  (tourist-n1
    (cat(n))
    (synonyms ""))
    (anno(def "a person who travels"))
    (comments ""))
    (ex "the tourists stayed at the hotel"))
    (syn-struct((root($var0))(cat(n))))
    (sem-struct(human(agent-of(sem(travel))))))
  )
```

# Lexicon management

(box-v1

```
(anno(def "to put in a box")(ex "he boxed the  
glassware")(comments ""))(synonyms "")(cat(v))
```

```
(syn-struct
```

```
(subject((root($var1))(cat(np))))(root($var0))(cat(v))  
(directobject((root($var2))(cat(np))))))
```

```
(sem-struct
```

```
(put(end-location(sem(box))))
```

```
(agent(value(^$var1(should-be-a(sem(human))))))
```

```
(theme(value(^$var2(should-be-a(sem(artifact material  
plant))))))
```

```
)
```

# Lexicon management

## Variables function:

<b>syn-struct</b>	\$var1	\$var0	\$var2	\$varx
<b>sem-struct</b>	^\$var1	concept	^\$var2	^\$varx

\$var0 stands for the ontological concept

- **Syntactic variables:**

- specify position in the sentence
- specify syntactic category

- **Semantic variables:**

- provide semantic information;
- constrain ontological properties;
- specify ontological fillers

# Lexicon management

## Constraining ontological properties

box-v1

...

(sem-struct

(**put**(end-location(sem(box))))

(**agent**(value(^\$var1(should-be-a(sem(human))))))

(**theme**(value(^\$var2(should-be-a(sem(artifact  
material plant))))))

# Lexicon management

(agent(value(^\$var1(should-be-a(sem(human))))))

Ontology:

(agent

(definition (value("the doer of an action")))

(is-a (hier(case-role)))

(inverse (value(agent-of)))

(domain (sem(event)))

(range (sem(animate))))

HUMAN → HOMINIDAE → <...> ANIMAL → ANIMATE

^\$var1 tells STAn to search for children of HUMAN



# Lexicon management

(**theme**

(value(^\$var2(should-be-a(sem(artifact material plant))))))

(**theme**

(definition (value("the object of an event")))

(is-a (hier(case-role)))

(inverse (value(theme-of)))

(domain (sem(event)))

(range (sem(event object property))))

ARTIFACT → PHYSICAL-OBJECT

MATERIAL → PHYSICAL-OBJECT

PLANT → ANIMATE → PHYSICAL-OBJECT

# Lexicon management

**Why are the restrictions necessary?**

(box-v1

(sem-struct

(**put**(end-location(sem(box))))

(**agent**(value(^\$var1)))

(**theme**(value(^\$var2))))))

Sentences now allowed:

\* **The fly** [Ag] **boxed the banana** [Th]

\* **The coral** [Ag] **boxed the bridge** [Th]

\* **The horse** [Ag] **boxed the star** [Th]

# Lexicon management

- **Lexicon acquisition:** machine-readable description of the senses in a corpus performed by an expert.
- A lexical entry is good if:
  - It is anchored in the appropriate concept;
  - It indicates all necessary syntactic positions and semantic roles;
  - It captures all syntactic variations of the sense;

# Acquisition strategies

- **Acquisition by rapid propagation**

One lexical entry is used as a template to acquire the whole class:

Adjectives:

```
(sense_name-adj1
  (cat(adj))
  (anno(def "a negative evaluation")(comments "")(ex ""))(synonyms ""))
(syn-struct((root($var0))(cat(adj))(syn-np((root($var1))(cat(np))))))
(syn-struct1((syn-np((root($var1))(cat(np))))(vp((root(be))(cat(v))))
  (adj-comp((root($var0))(cat(adj))))))
(sem-struct(^$var1(ATTRIBUTE(facet(value)))) )
```

# Acquisition strategies

- **Rule-based acquisition:**

Grammatical derivatives are acquired from their original forms based on semantic similarity:

murder-n1 → murder-v1

```
(sem-struct(murder(agent(value(^$var1)))(beneficiary(value(^$var2))))))
```

reconstruct-v1 → reconstruction-n1

```
(sem-struct(make(iteration(value(double))))
```

```
(agent(value(^$var1)))(theme(value(^$var2))))))
```

enjoy-v1 → enjoyable-adj1

```
(sem-struct(enjoy(experiencer(value(^$var1)))(theme(value(^$var2))))))
```

# Acquisition techniques

**Ontology-driven acquisition:** lexicalization of every EVENT and OBJECT in the ontology

- **pros:** time-efficient; produces a workable lexicon
- **cons:** incomplete lexicon, minimal sense coverage

**When to use:** early stages of ontological acquisition; helps overall ontology assessment

# Acquisition techniques

- **Parser-driven acquisition:** analysis of TMR's of domain-unrestricted corpora; acquisition, augmentation and adjustment of lexicon entries.
  - **pros:** immediate functionality check; quality control for lexical, ontological entries and parsing modules

A backbone of lexicon acquisition; beneficial as a testing/analysis tool

# Acquisition techniques

- **Domain-driven acquisition**: running OST text parser on a specific corpus; acquiring/adjusting lexical entries based on TMR's
  - **pros**: allows for the tweaking of the lexicon to the needs of a specific application;

Best used with the gap detector; a tool for determining missing entries with grammatical filters.



# Lexicon management

- **PROBLEM:** acquiring verbs with Subject-Instrument alternation (Beth Levin's index)

(1) The man [Agent] broke the window

(2) The hammer [Instrument] broke the window

(3) The asteroid [Instrument] broke the window

(4) The hurricane [Precondition] broke the window

**Solution:** expand the sem-struct to include more case roles mapping on the same syntactic variable:

# Lexicon management

(break-v1

(cat(v))(anno(def "to cause to break")

(ex ""))(comments ""))(senseprim(1)))(synonyms ""))

(syn-struct

((subject((root(**\$var1**))(cat(np))))(root(\$var0))(cat(v))

(directobject((root(\$var2))(cat(np))))))

(sem-struct (damage(**agent**(value(**^\$var1**))))

(**instrument**(value(**^\$var1**(should-be-a(sem(artifact  
animate-part material celestial-object))))

(**precondition**(value(**^\$var1**))))))

(theme(value(**^\$var2**(should-be-a(sem(artifact))))))))))