

# **Semantic Analysis in the Mikrokosmos Machine Translation Project**

**Stephen Beale, Sergei Nirenburg and Kavi Mahesh  
Computing Research Laboratory  
New Mexico State University**

**Seminar : Knowledge Representation and Domain Ontologies  
Presented by: Praharsha Sirsi**

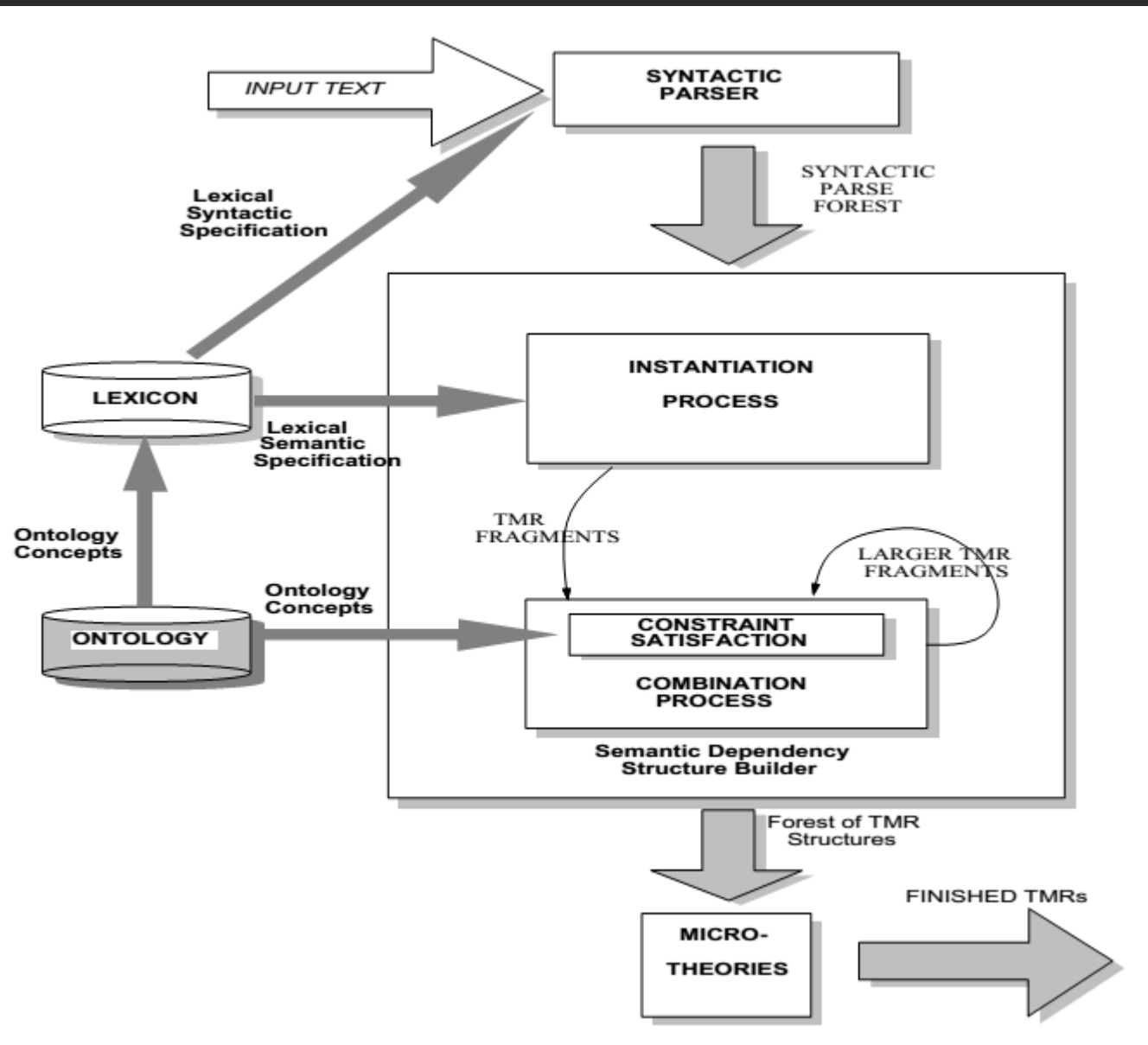
# Introduction

- Mikrokosmos ( $\mu$ K)
- Knowledge-based machine translation (KBMT) system.
- Semantic analysis
- Spanish to English
- Language independent formats
- Microtheories
- Lexical-Semantic dependency

# Different components in $\mu K$

- ⦿ Lexicon
  - Syntactic Parser
  - Semantic Specification
- ⦿ Ontology
  - Ontological Concepts
- ⦿ Semantic Analyzer
  - Text Meaning Representation (TMR)
  - Constraint Analyzer
  - Microtheories

# NLP Architecture

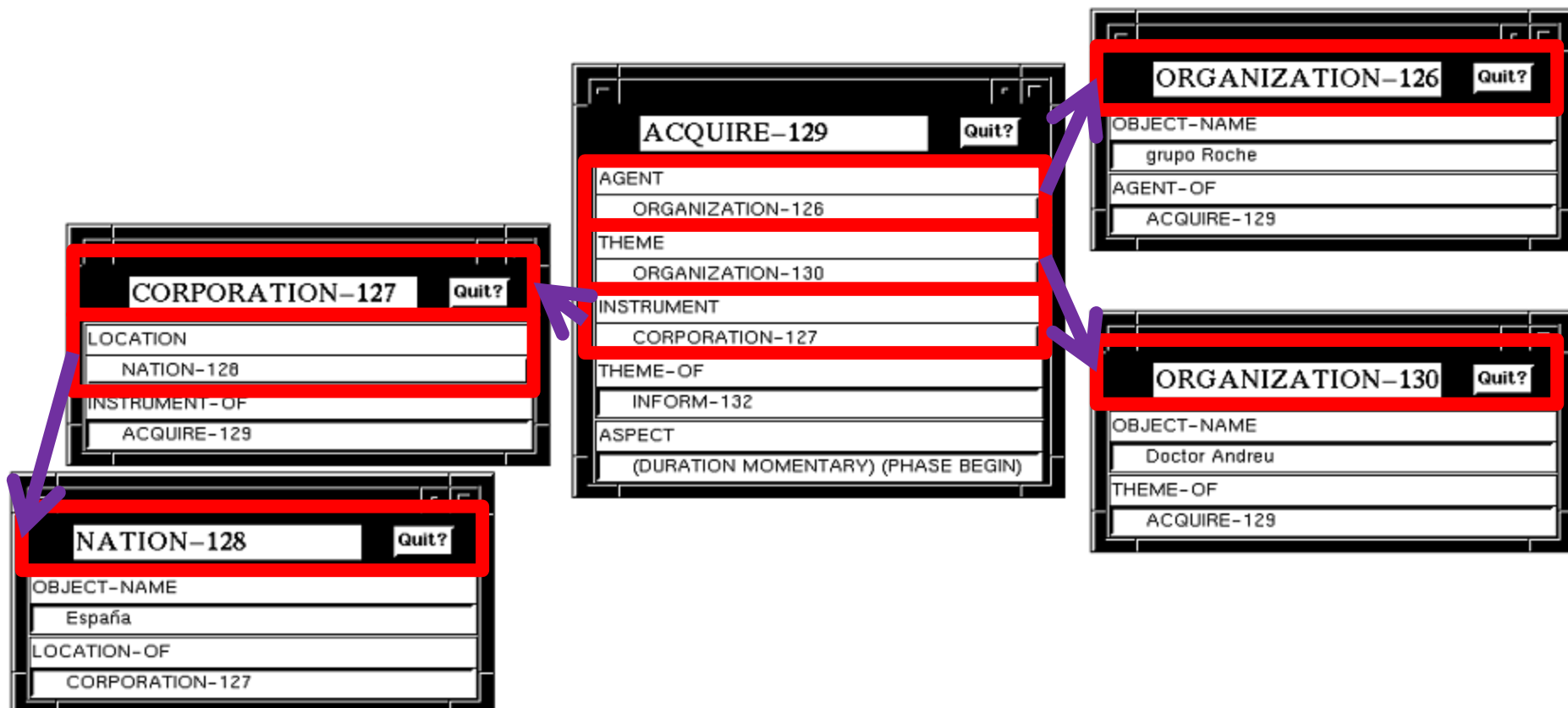


# Text Meaning Representations

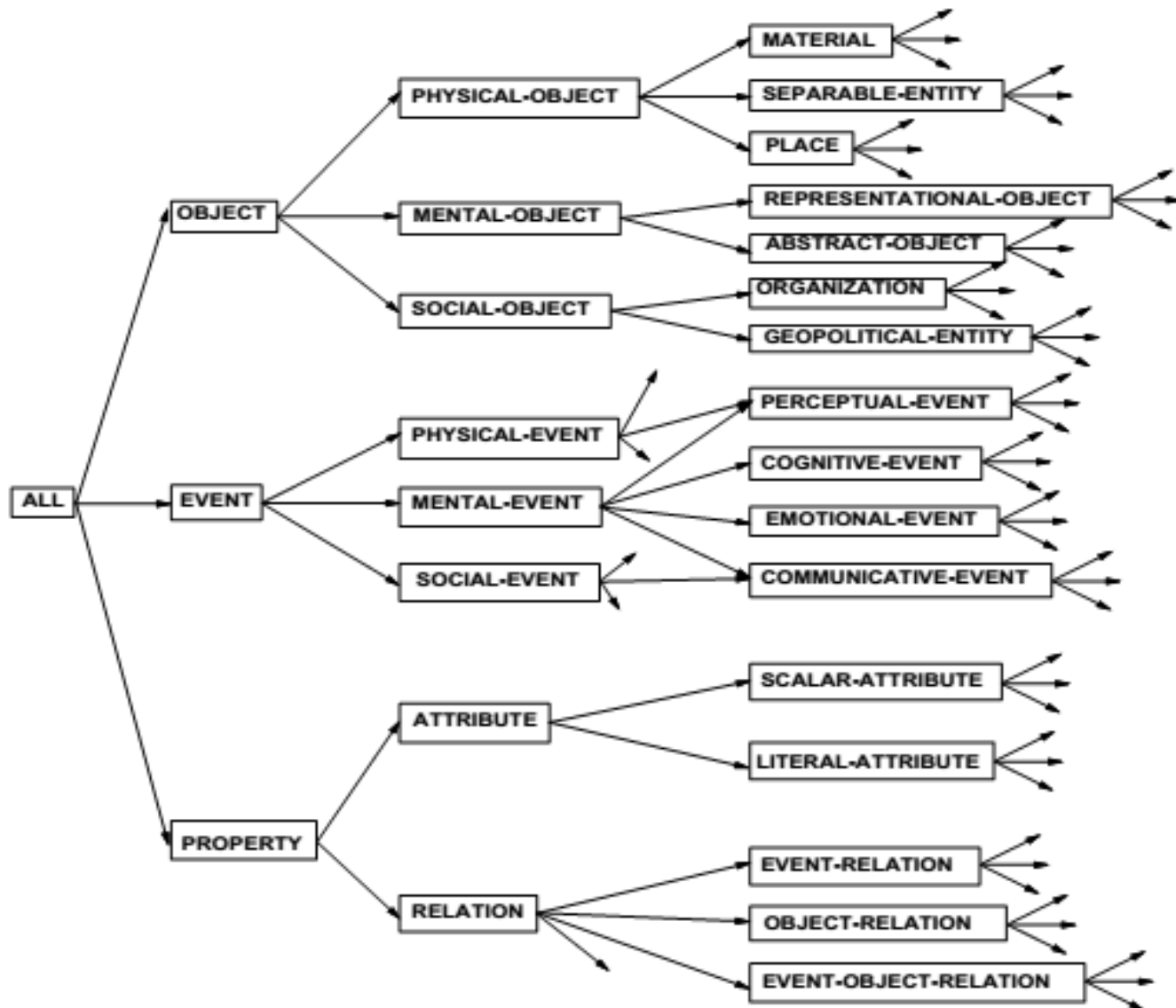
- ⦿ Language-neutral description
- ⦿ Deliberately syntax neutral
- ⦿ Lexical-semantic dependencies
- ⦿ Information provided
  - > Stylistic factors
  - > Discourse relations
  - > Speaker attitudes
  - > Pragmatic factors
- ⦿ Special notations for
  - > Attitudes
  - > Relations
  - > Quantities
  - > Time

# Example of a TMR

- 1a. *El grupo Roche, a traves de su compania en Espana, adquirio Doctor Andreu, se informo hoy aqui.*  
1b. *The Roche group, through its company in Spain, acquired Doctor Andrew, it was announced today.*



# Top level Ontology



# Concept entries in Ontology

### Ontology Concept Display.

Enter Concept Name or Keyword:

Concept Name: **ACQUIRE**

DEFINITION	
VALUE	The transfer of possession event where
TIME-STAMP	
VALUE	created by mahesh at 17:36:26 on 03/1
IS-A	
VALUE	TRANSFER-POSSESSION
SUBCLASSES	
VALUE	INHERIT
THEME	
SEM	OBJECT (NOT HUMAN)
AGENT	
SEM	HUMAN
PRECONDITION-OF	
SEM	OWN
SOURCE	
SEM	HUMAN
PURPOSE-OF	
SEM	BID WIN
INSTRUMENT	
SEM	PHYSICAL-OBJECT EVENT
CAUSED-BY	SEM TRY
LOCATION	SEM PLACE
HAS-PARTS	VALUE TRANSFER-C

### Ontology Concept Disp

### Ontology Concept I

Enter Concept Name or Keyword:

AGENT	
SEM	HUMAN
INSTRUMENT	SEM PHYSICAL-OBJECT



# Semantic Lexicon

## Lexical Entry Display.

Word:

[Display Another?](#)

[Lex Complain](#)

[\[button\]](#)

Word:

Entry:

Category:

Definition:

Example:

### Syn Struc:

```
(1 ((root $var0)
    (cat v)
    (subj ((root $var1)
            (cat n)))
    (obj ((root $var2)
          (cat n)))))
```

### Lex Map:

```
(1 (acquire
    (agent (value ^$var1))
    (theme (value ^$var2))
    (aspect ((phase begin)
              (duration momentar
              (iteration single)
y)
```

[MORE Entries?](#)

## Lexical Entry Display

## Lexical Entry Display.

Word:

[Display Another?](#)

[Lex Complain](#)

[\[button\]](#)

Word:

Entry:

Category:

Definition:

Example:

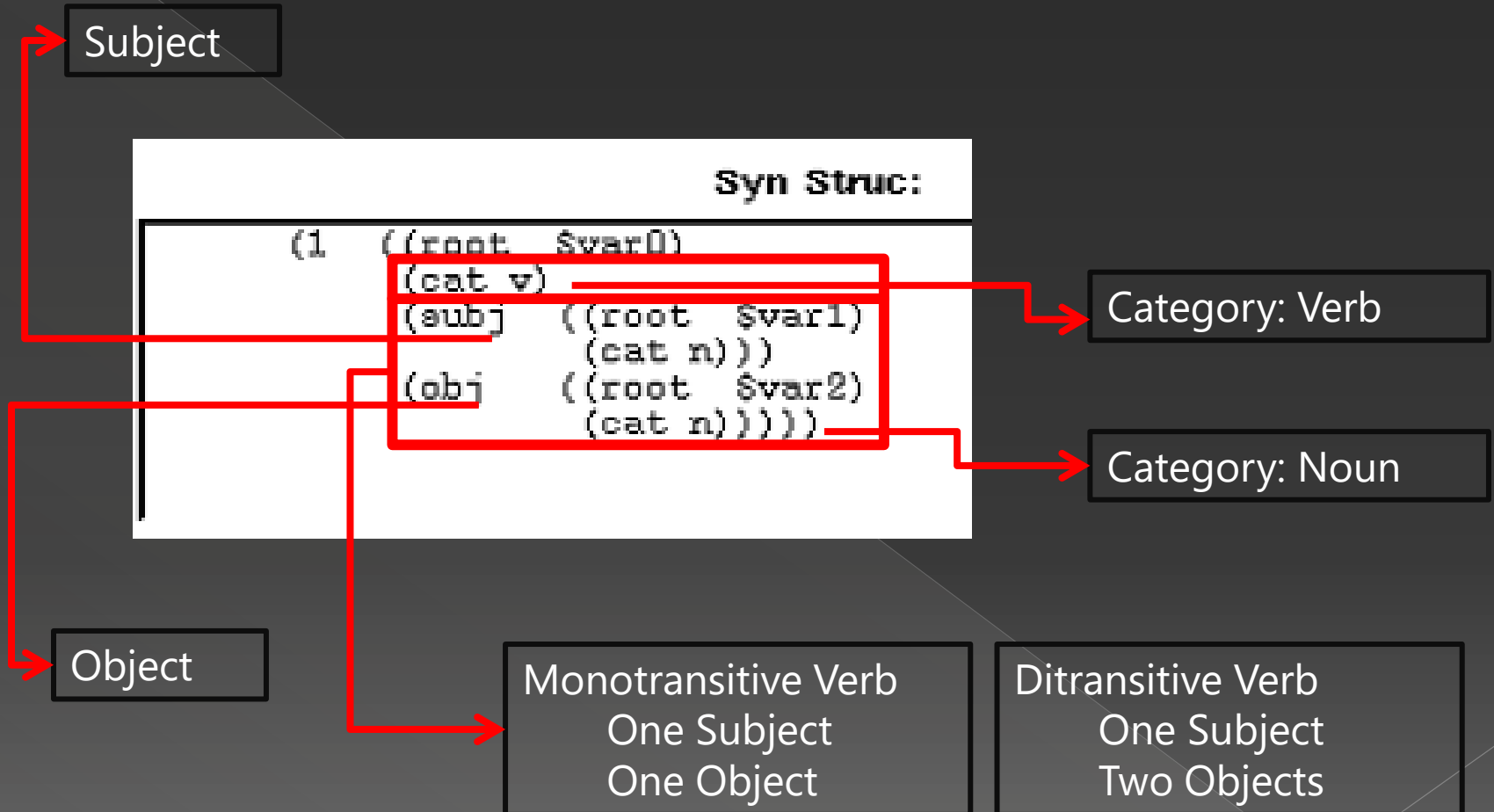
### Syn Struc:

```
(1 ((root $var0)
    (cat v)
    (subj ((root $var1)
            (cat n)))
    (obj ((root $var2)
          (cat n)))))
```

### Lex Map:

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

# Lexicon: SYN-STRUC Zone



- 1a. *El grupo Roche, a traves de su compania en Espana, adquirio Doctor Andreu, se informo hoy aqui.*  
1b. *The Roche group, through its company in Spain, acquired Doctor Andrew, it was announced today.*

# Lexicon: SEM Zone

## Lex Map:

```
(1 (acquire  
  (agent (value ^$var1))  
  (theme (value ^$var2))  
  (aspect  
    ((phase begin)  
     (duration momentar  
    (iteration single)
```

A red arrow originates from the 'agent' slot of the 'acquire' Lex Map and points to a text box on the left. Another red arrow originates from the 'aspect' slot of the 'acquire' Lex Map and points to a text box on the right.

Filled by TMR relating to  
AGENT: 'grupo Roche' (var1)  
THEME: 'Dr. Andreu' (var2)

Additional information from ACQUIRE TMR  
'a traves de' will add INSTRUMENT slot

## Lex Map:

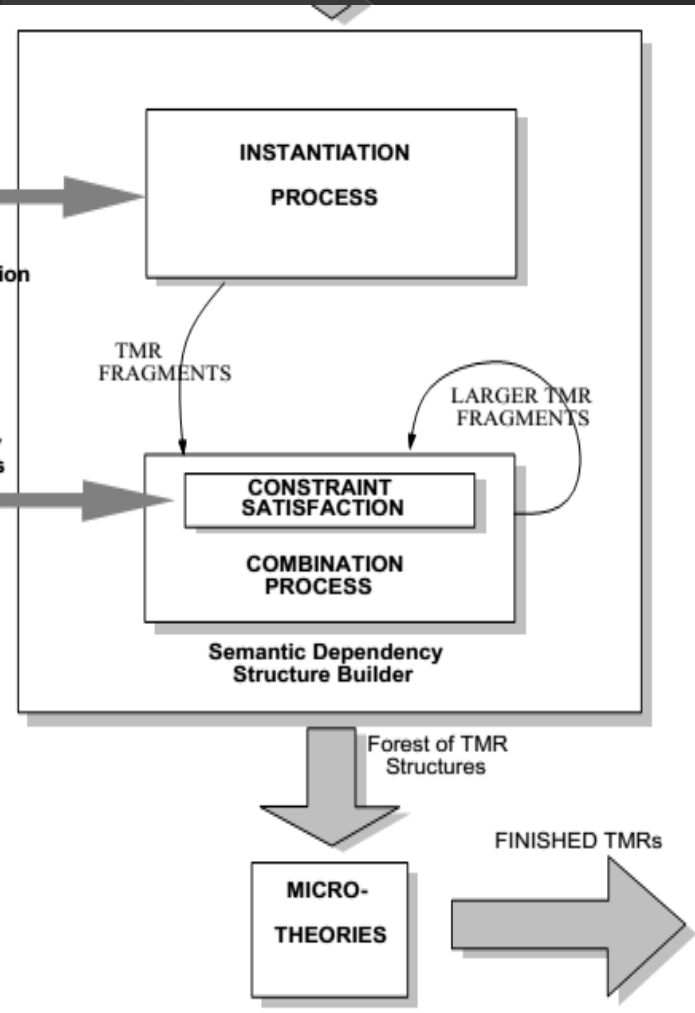
```
(1 (learn  
  (agent (value ^$var1))  
  (theme (value ^$var2))
```

A red arrow originates from the 'agent' slot of the 'learn' Lex Map and points to the text box on the left.

# Lexicon

- ◉ Variable bindings for interaction
- ◉ Principle of Compositionality
- ◉ SEM zone for language-specific constraints
- ◉ Example for Language Specific Semantic:
  - > 'to taxi' as in 'the jet taxied to the gate'
  - > Maps to GROUND-CONTACT-MOTION
  - > Specifies INSTRUMENT must be AIRCRAFT
- ◉ But, cannot define every conceivable mapping

# The Semantic Analyzer



- Combine ontology and lexicon
- Apply them to get TMRs
- Find semantic constraints
- Test each constraint in context
- Further check the SEM zones
- Apply Microtheories for Core TMRs
- Construct the output TMRs

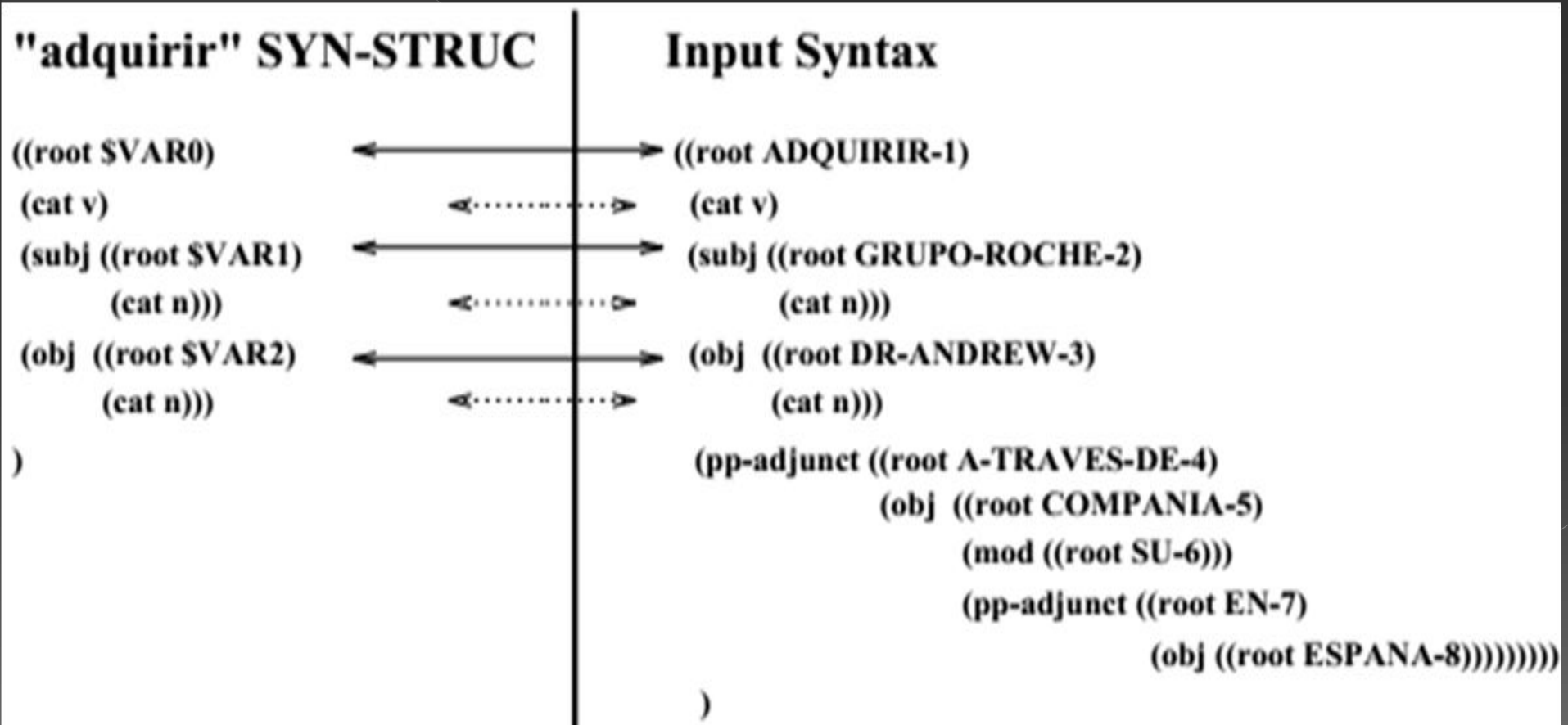
# Generating Constraints

- Gather all possible word sense using lexicon entries

Grupo-Roche	a-traves-de	su	compania	en	espana	adquirir	Dr. Andrew
ORGANIZATION	LOCATION	OWNER	CORPORATION	LOCATION	NATION	ACQUIRE	HUMAN
	INSTRUMENT		SOCIAL-EVENT	TEMPORAL		LEARN	ORGANIZATION

# Generating Constraints

- For each word sense, examine the SYN-STRUC zone



# Generating Constraints

- Now examine the SEM zone of each word sense to construct a list of constraints

## Lex Map:

```
(1 (acquire
    (agent (value ^$var1))
    (theme (value ^$var2))
    (aspect
        ((phase begin)
         (duration momentar
          (iteration single))
```

## Lex Map:

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

- Constraints can arise from 5 sources
  - Semantics in Ontological Concept
  - Fillers in Ontological Definition
  - Slots in Ontological Definition
  - Lexicon Entries
  - Other structures in the sentence



# 1 - Semantics in Ontological Concept

### Ontology Concept Display.

Enter Concept Name or Keyword:

---

**Concept Name:**

DEFINITION	
VALUE	The transfer of possession event where
TIME-STAMP	
VALUE	created by mahesh at 17:36:26 on 03/1
IS-A	
VALUE	TRANSFER-POSSESSION
SUBCLASSES	
VALUE	INHERIT
THEME	
SEM	OBJECT (NOT HUMAN)
AGENT	
SEM	HUMAN
PRECONDITION-OF	
SEM	OWN
SOURCE	
SEM	HUMAN
PURPOSE-OF	
SEM	BID WIN
INSTRUMENT	
SEM	PHYSICAL-OBJECT EVENT
CAUSED-BY	SEM TRY
LOCATION	SEM PLACE
HAS-PARTS	VALUE TRANSFER-C

### Ontology Concept Disp

### Ontology Concept I

Enter Concept Name or Keyword:

---

**Concept Name:**

AGENT	
SEM	HUMAN
INSTRUMENT	SEM PHYSICAL-OBJECT

```
graph LR; A[ACQUIRE] --> B[LEARN]; A[ACQUIRE] --> C[LEARN];
```

# 2 – Fillers in Ontological Definition

- These constraints ask about the fillers
- Type 1 Constraint :
  - > What kind of fillers do I allow?
- Type 2 Constraint :
  - > What kind of concepts can this filler modify with slot?
- Example: HAMMER
  - > filler for INSTRUMENT slot
  - > Modifies BUILD event

# 3 – Slots in Ontological Definition

- ⦿ These constraints depend on the definition of slots
- ⦿ Each slot has
  - > DOMAIN
  - > RANGE
- ⦿ Example:
  - > AGENT slot requires
    - DOMAIN = EVENT
    - RANGE = HUMAN
  - > THEME slot requires
    - DOMAIN = EVENT
    - RANGE = OBJECT or EVENT
- ⦿ Default values exist for the slots
- ⦿ Although general constraints, but they still help

# 4 – Lexicon Entries

- They can have explicit constraints
  - But in our example they don't

**Entry:** `adquirir-V1`

**Lex Map:**

```
(1 (acquire
    (agent (value ^$var1))
    (theme (value ^$var2))
    (aspect
        ((phase begin)
         (duration momentar
          (iteration single))
```

**Entry:** `adquirir-v2`

**Lex Map:**

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

# 5 – Other structures in the sentence

Grupo-Roche	a-traves-de	su	compania	en	espana	adquirir	Dr. Andrew
ORGANIZATION	LOCATION INSTRUMENT	OWNER	CORPORATION SOCIAL-EVENT	LOCATION TEMPORAL	NATION	ACQUIRE LEARN	HUMAN ORGANIZATION

'a-traves-de'  
will choose  
LOCATION  
or  
INSTRUMENT  
slot for the TMR

Location Slot

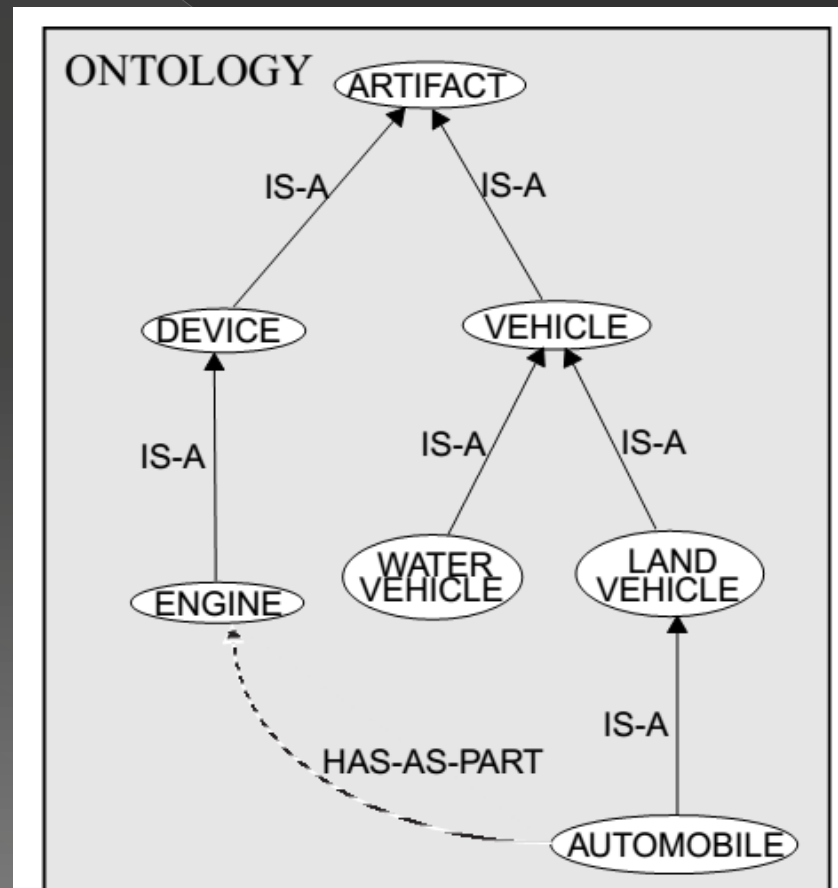
Instrument Slot

Slot filled by  
'compania' TMR :  
CORPORATION  
or  
SOCIAL-EVENT

- These slots and TMRs will influence other words in the Sentence

# Applying the constraints

- Ontological graph search function
- Determines relevant paths between concepts
- Each path will have a cost between 0 to 1



# Applying the constraints

- Returns a score based on degree of closeness
- Example:
  - > checkonto-con(ACQUIRE EVENT)
    - Returns score of 1.0 (out of 1.0)
    - Because ACQUIRE is a type of EVENT.
  - > check-onto-con(ORGANIZATION HUMAN)
    - returns a score of 0.9
    - Along with path (ORGANIZATION HAS MEMBER HUMAN).
    - Indicates ORGANIZATION can be in place of HUMAN because of HUMAN members
- Metonyms and Metaphorical meanings are derived

# Determining the Best Combination

- ⦿ Each combination
  - After applying the constraints
  - Combined to get total score
- ⦿ Combination with the best total score is chosen
- ⦿ This gives the Core TMR
  - Further Microtheories are applied



# Example Sentence

Grupo-Roche	a-traves-de	su	compania	en	espana	adquirir	Dr. Andrew
ORGANIZATION	LOCATION INSTRUMENT	OWNER	CORPORATION SOCIAL-EVENT	LOCATION TEMPORAL	NATION	ACQUIRE LEARN	HUMAN ORGANIZATION

Will choose INSTRUMENT  
As LOCATION ->  
'adquirir' = physical object  
But  
'adquirir' ≠ physical object

Will choose LOCATION  
As TEMPORAL ->  
'espana' = temporal object  
But  
'espana' ≠ temporal object

Will choose ACQUIRE  
As LEARN ->  
'Dr.Andrew' = information  
But  
'Dr.Andrew' ≠ information

The choice is not yet defined  
Additional Ontological Information  
Statistical Information

Will choose ORGANIZATION  
As HUMAN cannot be the theme of  
'ACQUIRE'

THEME	
SEM	OBJECT (NOT HUMAN)

# Advanced Computational Methods

- Finding Constraints = Lot of processing
- 'bulletin-board' processing scheme
- The scheme includes
  - > Dependency Analysis
  - > Best First Processing
  - > Failure Recovery Techniques
  - > Ambiguity Resolution

# Dependency Analysis

- ⦿ Complex interplay of constraints.
- ⦿ Choosing one sense may be optimal, but it may create problems elsewhere
- ⦿ Dependency-directed analysis systematically tracks dependencies
  - Propagates related constraints forward
  - Detects inconsistent solutions
  - Failure processing
- ⦿ Developed by Authors

# Best First Processing

- ⦿ Statistical data
- ⦿ Word senses that are close enough
- ⦿ Processing stops on satisfactory results
- ⦿ Used in every aspect of processing

# Failure Recovery Techniques

- Failures from
  - > Spelling errors
  - > Syntactic analysis
  - > Lexicon, Ontology errors
  - > Lack of needed information
  - > Analyzer makes incorrect decisions

# Failure Recovery Techniques

- $\mu$ K deals with these by:
  - > Using dependency analysis
  - > Checking for metonymic/metaphoric language
  - > Check for missing slot fillers
  - > Changing syntactic analysis, different attachments
  - > Relaxing thresholds
  - > Using sophisticated 'best first' approach

# Ambiguity Resolution

- ◎ If basic semantic constraints are not enough
  - > Preferences in lexicon
  - > Statistical methods
  - > Subsequent clauses combined with co-references
  - > Attachment rules
  - > Expectations in senses
    - Example: 'adquirir' sense expects INSTRUMENT slot
    - Which 'a-traves-de' adds

# Conclusion



Thank you!