

# Configuring Domain Knowledge for Natural Language Understanding

Matt Selway, Wolfgang Mayer, Markus Stumptner

Advanced Computing Research Centre (ACRC)  
School of IT & Mathematical Sciences  
University of South Australia

ConfWS 2013, Vienna, AT  
29 August, 2013

# Introduction

## Goals:

- Translate natural language specifications and requirements into formal models
- Support software development process (MDE)
  - improve quality of requirements
  - reduce manual effort and ambiguities
  - identify and correct inconsistencies and errors

## Problems:

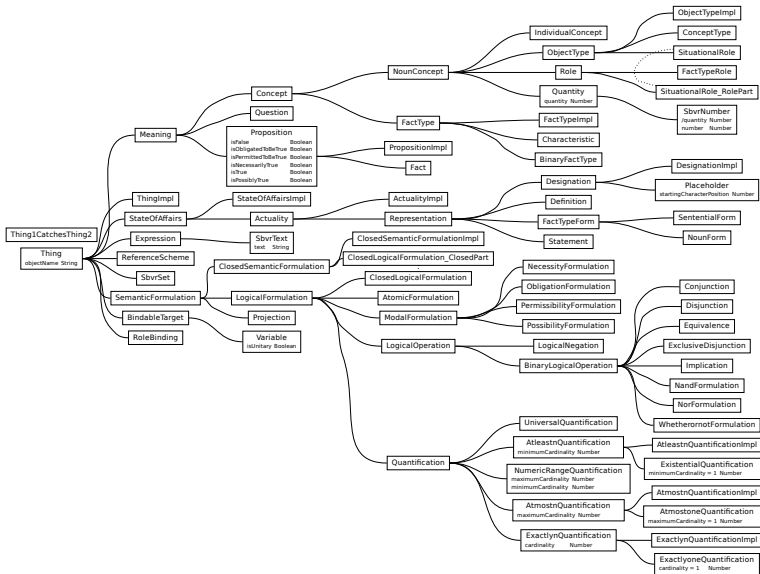
- Requires semantic processing or *understanding* of texts

## Our Solution:

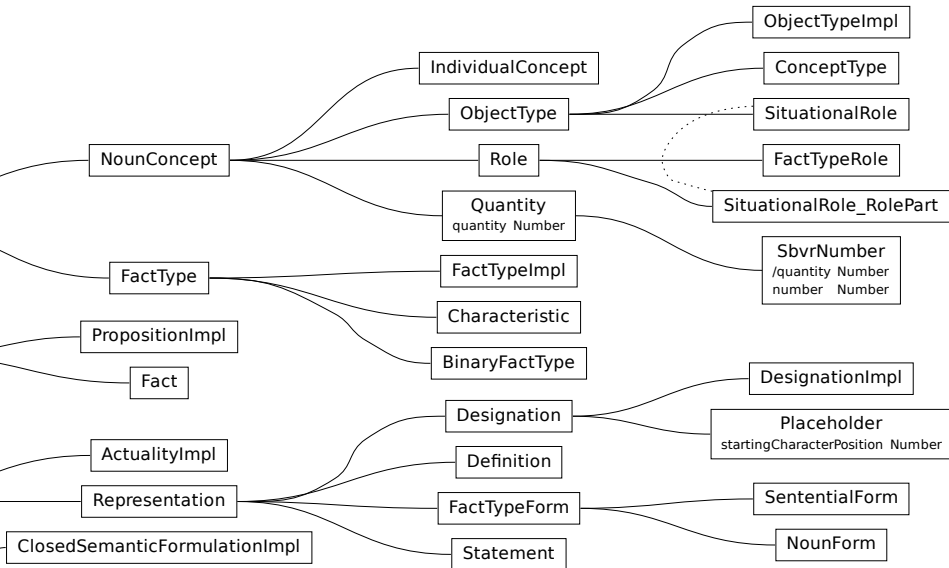
- Apply knowledge-based configuration to fragments of a semantic model comprising the domain knowledge

## Case Study (extract)

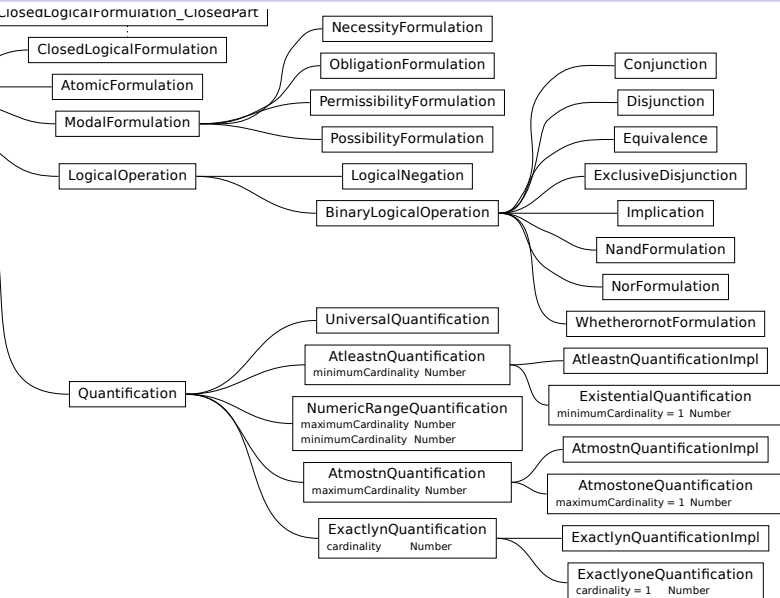
## SBVR Core Concepts



# SBVR Core Concepts



# SBVR Core Concepts



## SBVR-based Specification (1)

### EU-Rent Vocabulary

#### rental organisation unit

Definition: organisational unit that operates part of EU-Rent's car rental business

#### rental organisation unit *having rental responsibility*

Definition: **the** rental organisation unit is responsible for the operation of customer-facing rental business

#### rental organisation unit *having area responsibility*

Definition: **the** rental organisation unit includes organisation units for which it has the responsibility to coordinate operations and ensure resources

## SBVR-based Specification (2)

### EU-Rent Vocabulary (cont.)

#### local area

Definition: rental organisation unit **that** *has area responsibility*

#### branch

Definition: rental organisation unit **that** *has rental responsibility*

branch *is included in local area*

Synonymous Form: local area *includes* branch

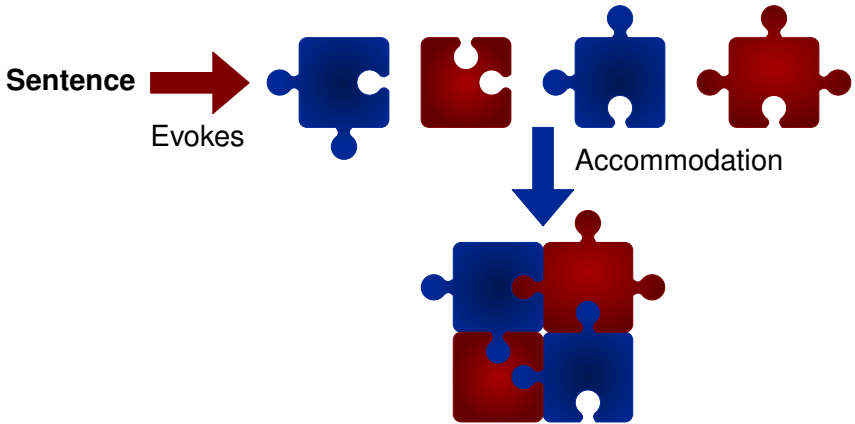
### EU-Rent Rules

**Each** branch *is included in exactly one* local area.

**The** country *of a* branch *is the* country **that** *includes the* operating company **that** *includes the* local area **that** *includes the* branch.

# Cognitive Grammar

## Semantic Structures





# Syntactic Analysis (1)

[Each]*	[branch]	*[is included in]*	[exactly one]*	[local area]	BE	LC	GC	
██████████	0	→	0	██████████	0	1	1	
	██████████	0	→	0	██████████	0	1	4/5
██████████	0	→	0	██████████	0	1	4/5	
██████████	0	→	1	██████████	1	4/5	4/5	
██████████	0	→			0	1	3/5	
	██████████	0	→	0	██████████	0	1	3/5
	██████████	0	→	1	██████████	1	3/4	3/5
	██████████	0	→		0	1	2/5	
			→	0	██████████	0	1	2/5
→	0	██████████			0	1	2/5	

## Syntactic Analysis (2)

- Process is iterative
- As catches occur, the combination is sent to the accommodation process
- If accommodation is successful, the resulting partial configuration can be re-used in larger combinations
- If unsuccessful, the combination is discarded

## Syntactic Analysis (3)

Traditional lexicons include:

- word
- syntactic categories (*noun, verb, etc.*)
- plurality
- voice (*passive, active*)
- transitivity

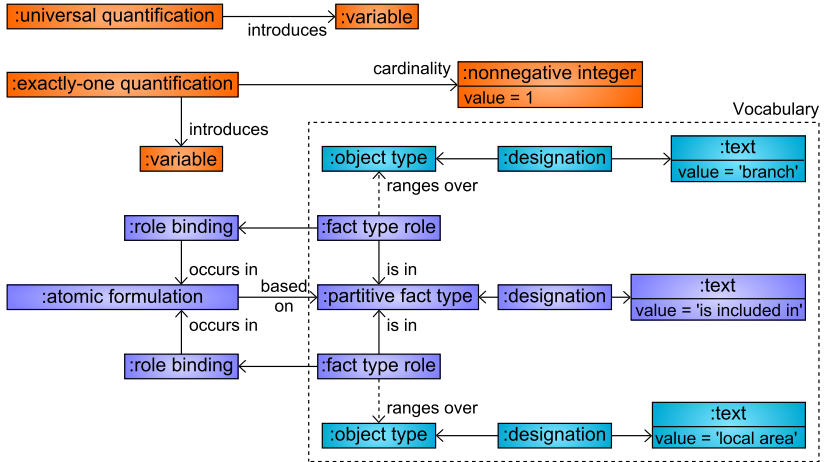
Our Lexicon includes:

- word (link to semantic model)
- expectations (linked to sites of semantic structure)
- expectation direction (possibly derived from model)

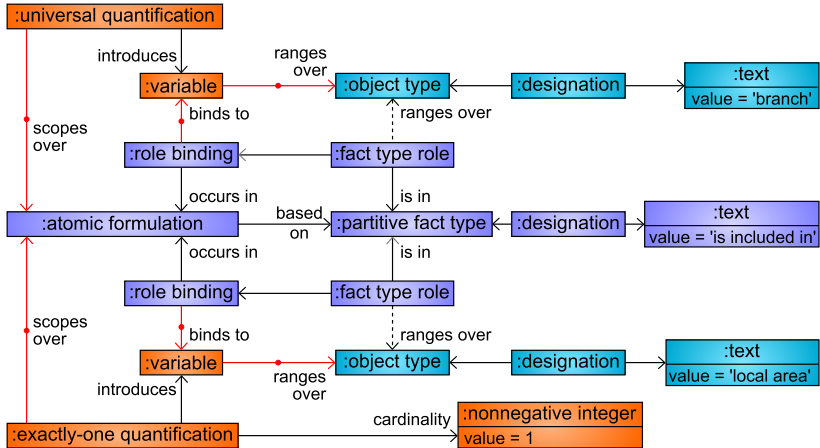
## Semantic Accommodation

- Performed using component-based configuration
- SBVR Model to configuration mapping:
  - Class hierarchy  $\mapsto$  Component type taxonomy
  - Objects  $\mapsto$  Components
  - Attributes and Data Types  $\mapsto$  Attributes and Attribute Types
  - Relationships  $\mapsto$  Ports and Port Types
  - Constraints  $\mapsto$  Constraints
- Syntactic information encoded as a Port Type and a pair of Ports (*catches* and *caughtby*)
- A constraint forces the configurator to try to connect two components if they are connected through *catches–caughtby*

# Partial Configuration



# Complete Configuration



## Configuration Results

- Using a slightly different KB, so numbers differ

	<b>R1</b>	<b>R1 Iterative</b>	<b>R2</b>	<b>R2 Iterative</b>
Ave. Time (seconds)	1.8	4.5	11.8	128
# Constraints	107	259	405	4467
# Variables Assigned	2090	5824	6627	82782
# Assigned Unconn.	2078	5812	6579	82734
Min. # Backtracks	56	56	442	434
Max. # Backtracks	59	58	462	454
Ave. # Backtracks	57.6	57.0	452	443
# Comps. Created	6	6	24	24

## Changes and New Results

- Added new port type based on constrained set variables
- Multiple ports representing a relation  $\rightarrow$  a single port

	<b>R1</b>	<b>R1 Iterative</b>	<b>R2</b>	<b>R2 Iterative</b>
Ave. Time (seconds)	0.16	0.41	0.67	6.11
# Constraints	102	244	390	4263
# Variables Assigned	258	699	824	9830
# Assigned Unconn.	246	687	776	9782
Min. # Backtracks	10	10	97	82
Max. # Backtracks	11	11	114	108
Ave. # Backtracks	10.6	10.5	105	94
# Comps. Created	6	6	24	24



## Contribution

- Combined Cognitive Grammar and Configuration to perform NL Processing/Understanding
- Semantic analysis of text
  - improves mappings between NL and the semantic model
  - allows semantic disambiguation of terms
- Iterative construction of a domain model
  - ensures consistency
  - improves identification of erroneous, ambiguous, and inconsistent statements
- Simplifies lexicon for our target application

## Future Work

- Optimise the Configuration process:
  - Additional ordering heuristics to minimise backtracks for more complex sentences
  - Implement better handling of correct partial configurations
  
- Perform analysis of entire specification to evaluate scalability as domain knowledge (i.e the model) grows