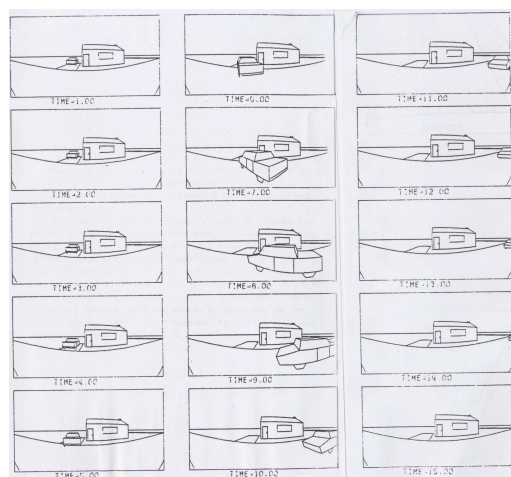


# Historical Examples for Scene Interpretation

1

## Early Traffic Scene Interpretation (Badler 75)



Norman I. Badler  
**Temporal Scene Analysis:  
Conceptual Descriptions of  
Object Movements**  
Dissertation, Report TR 80,  
Dep. of Computer Science,  
University of Toronto, 1975

**Task:**  
Describe motion in terms  
of changing spatial  
relations

**Input data:**  
15 "snapshots" of a car  
leaving the driveway of a  
house

2

## Directional Adverbials for Motion Description (Badler 75)

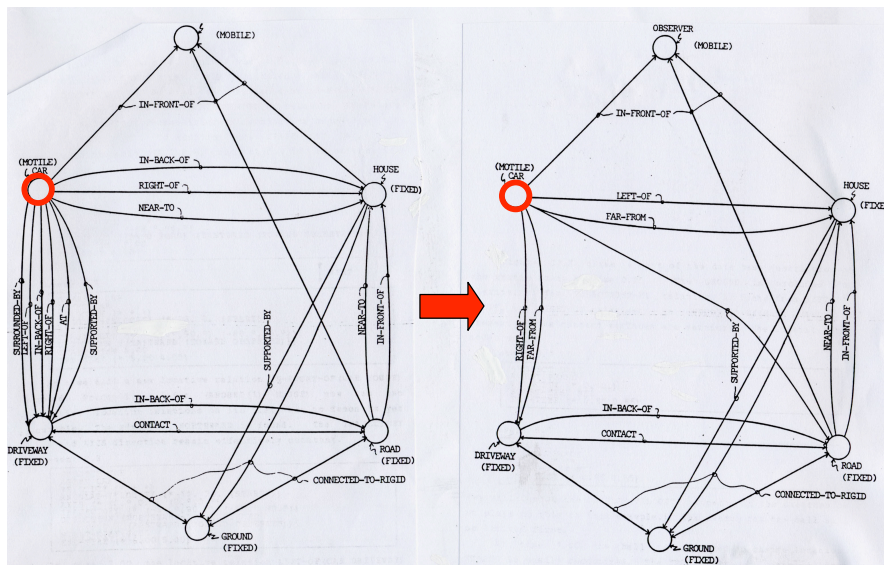
Scene interpretation is performed by recognising changing spatial relationships between objects and relating the changes to directional adverbials.

Directional adverbials used by Badler:

ACROSS	CLOCKWISE	OUT
AFTER	COUNTERCLOCKWISE	OUT-OF
AGAINST	DOWN	OUTWARD
AHEAD-OF	FORWARD	OVER
ALONG	FROM	SIDeways
APART	IN	THROUGH
AROUND	IN-THE-DIRECTION-OF	TO
AWAY	INTO	TO-AND-FRO
AWAY-FROM	INWARD	TOGETHER
BACK	OFF	TOWARD
BACK-AND-FORTH	OFF-OF	UNDER
BACKWARD	ON	UP
BEHIND	ONTO	UP-AND-DOWN
BY	ONWARD	UPWARD
		WITH

3

## Changing Scene Graph for Car Scene (Badler 75)



4

## Demon Representation of "ACROSS" Motion (Badler 75)

A NEAR-TO relation with one side of an object is broken and replaced by a similar relation with the other side. There is an implicit sense of passage ABOVE the object.

**Precondition 1**  
 NEAR-TO(X S1).  
 SUB-PART(Y S1) for some object Y and SUB-PART [chain] to object S1.  
 FRONT or BACK or LEFT-SIDE or RIGHT-SIDE(Y S1).  
 ACROSS remains active as long as NEAR-TO(X Y) and ABOVE(X Y) hold.

**Precondition 2**  
 NEAR-TO(X S2).  
 SUB-PART(Y S2) for a SUB-PART [chain] to object S2.  
 FRONT or BACK or LEFT-SIDE or RIGHT-SIDE(Y S2) where S1 ≠ S2 and at least one of the ORIENTATION relations to S1 (from Precondition 1) no longer holds.

**Postcondition**  
 SUBJECT X  
 DIRECTION PCONS((ACROSS Y), DIRECTION)

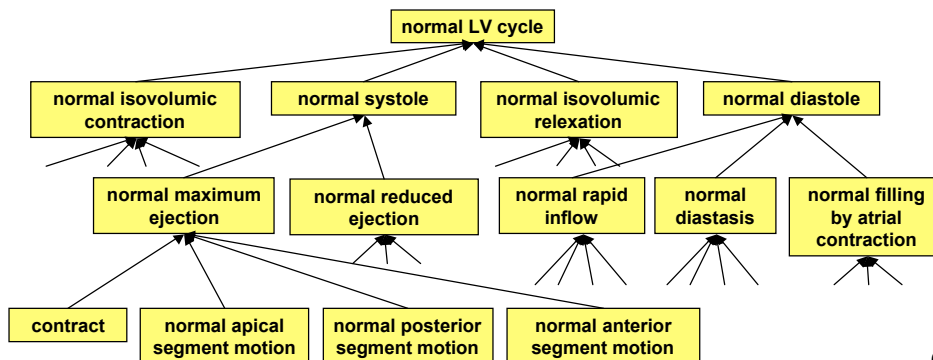
5

## Left-ventricular Motion PART-OF Hierarchy (Tsotsos et al. 79)

Tsotsos, J., Mylopoulos, J., Covvey, H.D., Zucker, S.W., **A Framework for Visual Motion Understanding**, Proc. Workshop on Computer Analysis of Time-Varying Imagery, Philadelphia, 1979.

**Task:** Recognize heart conditions from ultrasound heart images

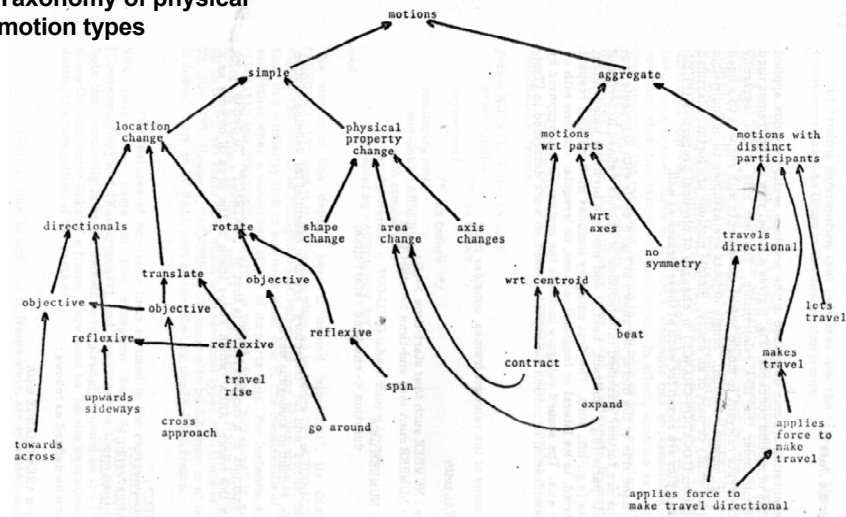
PART-OF structure supports part-whole reasoning in recognition processes



6

## Motion IS-A Hierarchy (Tsotsos et al. 79)

### Taxonomy of physical motion types



7

## Model-based Prediction for Tracking a Jointed Moving Object (Hogg 84)

D. Hogg, *Interpreting Images of a Known Moving Object*. PhD thesis, University of Sussex, Brighton, UK, 1984.

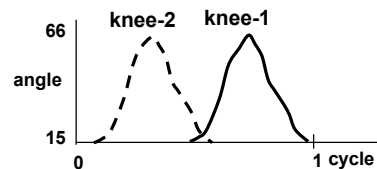
**Task:** Describe highly coordinated motion of parts.

Use of quantitative measures along time axis.



Posture curves + constraints represent coordinated motion of joints of walker.

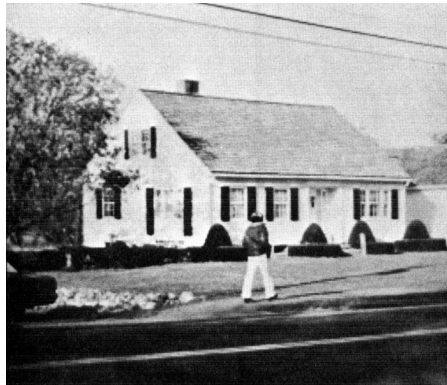
**Example:**



8

## The VISIONS Image Interpretation System (Hanson & Riseman 78)

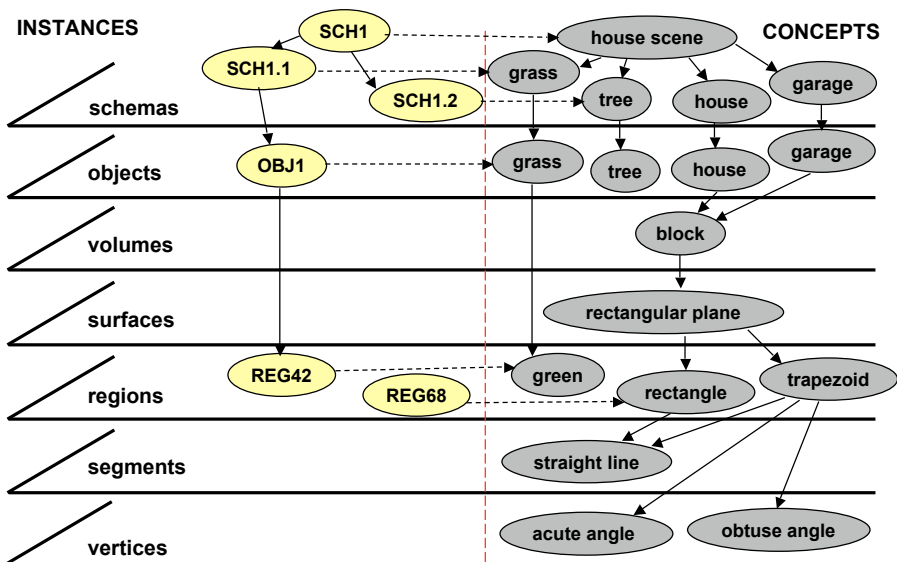
Long-term research about the interpretation of land-house scenes



(original in colour)

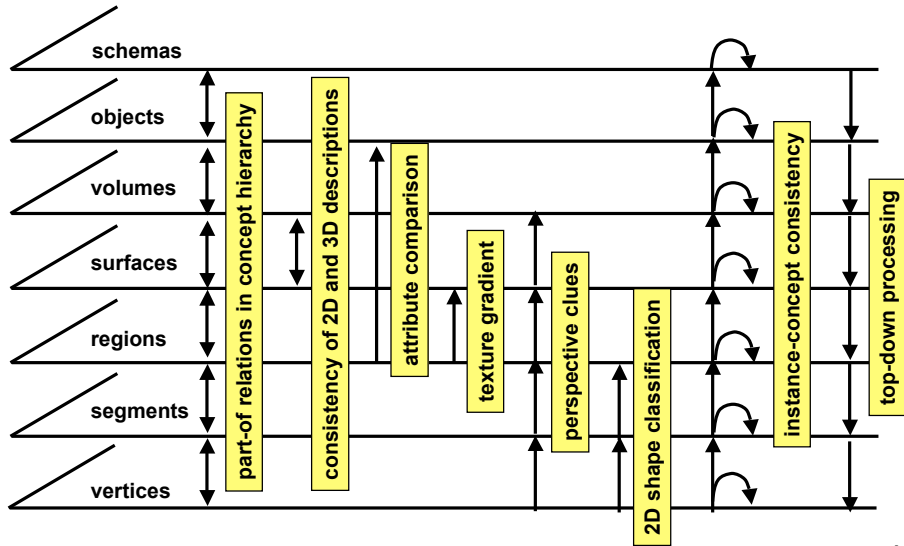
9

### Representation Hierarchy in VISIONS



10

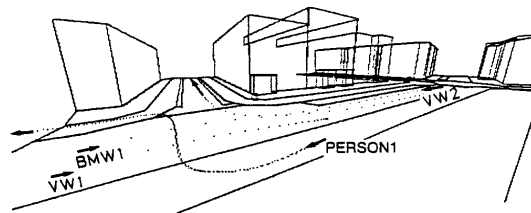
## Knowledge Sources for VISIONS



11

## NAOS - Natural Language Description of Object Motions in Traffic Scenes

Neumann, B.; Novak, H.-J., NAOS: Ein System zur natürlichsprachlichen Beschreibung zeitveränderlicher Szenen (NAOS: A System for Natural-language Description of Time-varying Scenes), Informatik Forschung und Entwicklung 1, 83 - 92, 1986



### English paraphrase of automatically generated description:

The scene contains four moving objects: three cars and a pedestrian.

A VW **drives** from the Alte-Post to the front of the FBI. It **stops**.

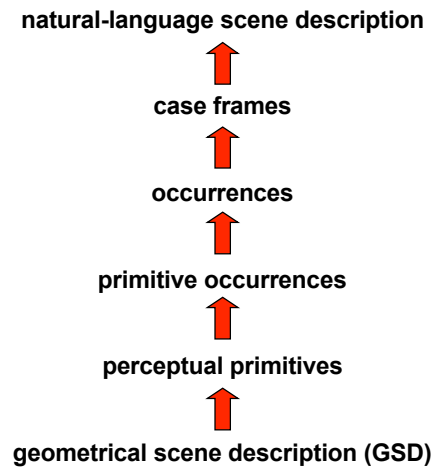
Another VW **drives** towards Dammtor. It **turns off** Schlueterstrasse. It **drives** on Bieberstrasse towards Grindelhof.

A BMW **drives** towards Hallerplatz. While doing so, it **overtakes** the VW which has stopped, before Bieberstrasse. The BMW **stops** in front of the traffic lights.

The pedestrian **walks** towards Dammtor. While doing so, he **crosses** Schlueterstrasse in front of the FBI.

12

## From Scene Data to a Natural-language Scene Description (NAOS)



13

## Geometrical Scene Description (GSD) in NAOS

Quantitative description of all objects in a time-varying scene:

- name of all objects (class or identity)
- position of all objects at all times (location and orientation)
- illumination (if required for high-level description)

Example of a synthesized GSD in NAOS:

```

(LAGE VW2 (779. 170. 0.) (-1.0 0.0 0.0) 0)
(LAGE VW2 (753. 170. 0.) (-1.0 0.0 0.0) 1)
(LAGE VW2 (727. 170. 0.) (-1.0 0.0 0.0) 2)
(LAGE VW2 (701. 170. 0.) (-1.0 0.0 0.0) 3)
(LAGE VW2 (675. 170. 0.) (-1.0 0.0 0.0) 4)
(LAGE VW2 (649. 170. 0.) (-1.0 0.0 0.0) 5)
(LAGE VW2 (623. 170. 0.) (-0.999 0.037 0.0) 6)
(LAGE VW2 (596. 171. 0.) (-1.0 0.0 0.0) 7)
(LAGE VW2 (570. 171. 0.) (-1.0 0.0 0.0) 8)
(LAGE VW2 (544. 171. 0.) (-1.0 0.0 0.0) 9)
.
.
.
  
```

⏟
⏟
↑

location
orientation
time

14

## Occurrence Model for "OVERTAKE" (NAOS)

(OVERTAKE OBJ1 OBJ2 T1 T2)  $\Leftrightarrow$   
 (MOVE OBJ1 T1 T2)  
 (MOVE OBJ2 T1 T2)  
 (BEHIND OBJ1 OBJ2 T1 T3)  
 (BESIDE OBJ1 OBJ2 T3 T4)  
 (BEFORE OBJ1 OBJ2 T4 T2)  
 (APPROACH OBJ1 OBJ2 T1 T3)  
 (DIS-APPROACH OBJ1 OBJ2 T4 T2)

- ➔ temporal constraint satisfaction for occurrence recognition
- ➔ principled definition of primitive occurrences

15

## Temporal Relations in NAOS

- Observations provide begin and end time-points of occurrences
- Models express qualitative constraints on time-points

Unary temporal constraints:  $t_{\min} \leq t \leq t_{\max}$

Binary temporal constraints:  $t_1 \geq t_2 + c_{12}$

Convex interval relations may be expressed by inequalities:

$$\begin{aligned}
 I_1 \text{ during } I_2 \Rightarrow & I_2.tb \leq I_1.tb \\
 & I_1.te \leq I_2.te
 \end{aligned}$$

*tb = begin time-point*  
*te = end time-point*

NAOS temporal constraint propagation was later identified as a convex time point algebra [Vila 94].

16

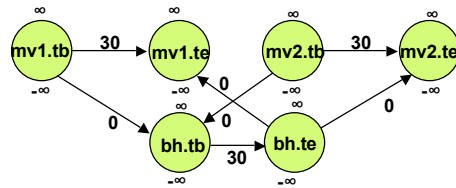


## Constraint Propagation for Occurrence Verification in NAOS (1)

**Example:**

Verify occurrence "two moving objects, one behind the other"

**1. Initialize constraint net of occurrence model**



- mv1 and mv2 must last at least 30 time units
- bh must last at least 30 time units
- bh must occur during mv1 and during mv2

**2. Compute primitive events for scene**

ID:	move1
instance:	move
parts:	mv-ob = obj1
	mv-tr = trj1
times:	mv-tb = 13
	mv-te = 98

ID:	behind1
instance:	behind
parts:	bh-ob1 = obj1
	bh-ob2 = obj2
times:	bh-tb = 20
	bh-te = 53

(and many more)

17

## Constraint Propagation for Occurrence Verification in NAOS (2)

**3. Instantiate parts in occurrence model**

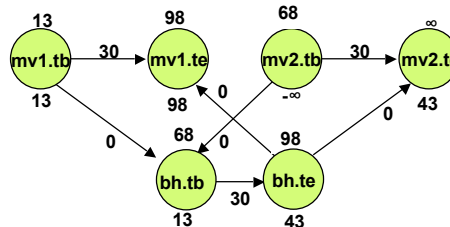
Propagate minima and maxima of time-points through constraint net:

- minima in edge direction  $t_{2min} = \max \{t_{2min}, t_{1min} + c_{12}\}$
- maxima against edge direction  $t_{1max} = \min \{t_{1max}, t_{2max} - c_{12}\}$

**Example:**

- move1 in scene instantiates mv1 of model
- time point values are propagated

ID:	move1
instance:	move
parts:	mv-ob = obj1
	mv-tr = trj1
times:	mv-tb = 13
	mv-te = 98



18

## Constraint Propagation for Occurrence Verification in NAOS (3)

### 4. Consistency and completeness test

A (partially) instantiated model is inconsistent, if for any node T one has:  $T_{min} > T_{max}$

=> search for alternative instantiations or terminate with failure

An occurrence has been recognized if the occurrence model is instantiated with sufficient completeness and the instantiation is consistent.

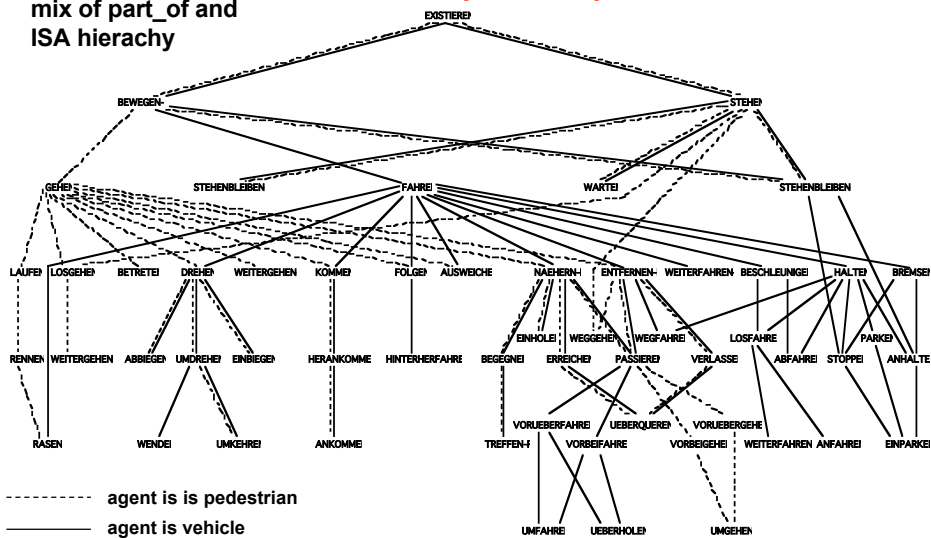
#### Note:

- Incremental occurrence recognition follows an evolving scene
- A-posteriori occurrence recognition is carried out after observing a scene (choice of order!)
- Partially instantiated models may be used for scene prediction

19

## Hierarchy for Object Motions in Street Traffic (NAOS)

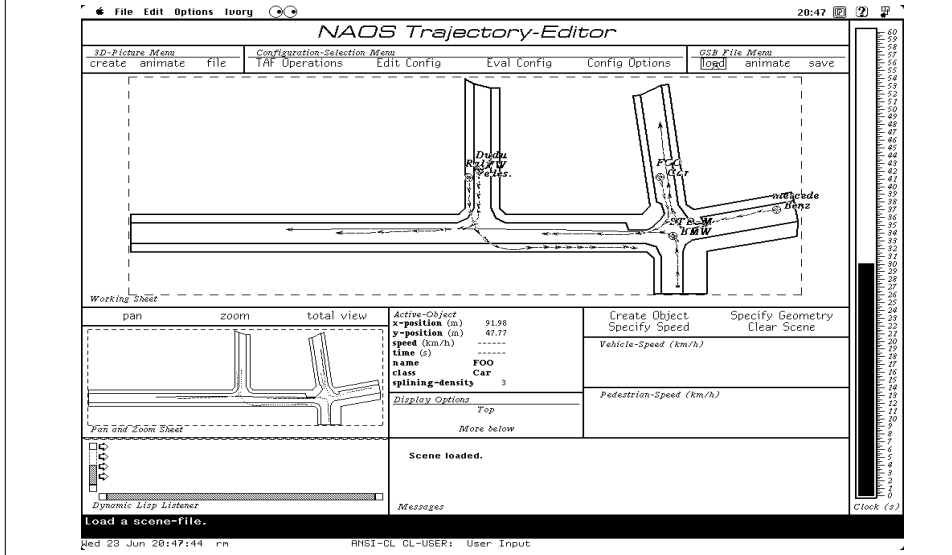
mix of part\_of and ISA hierachy



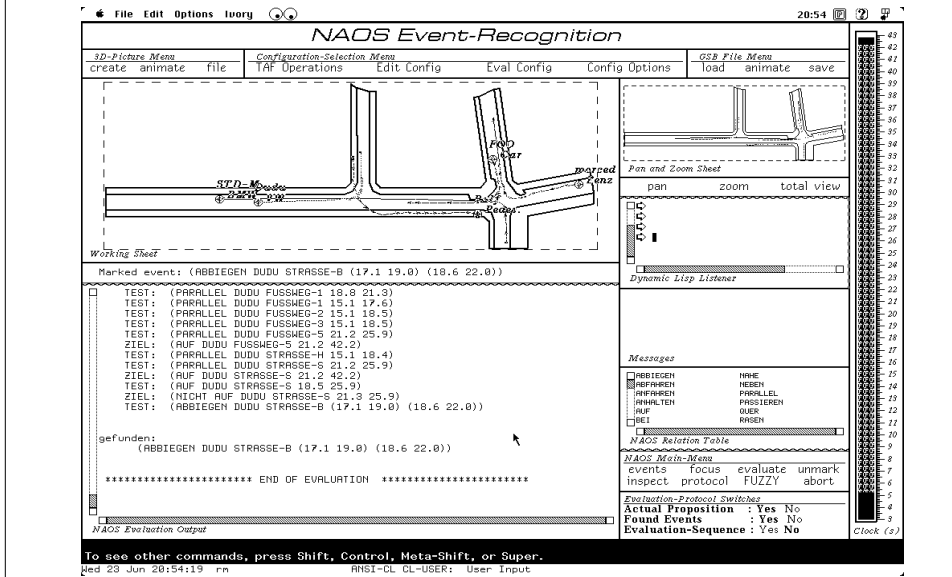
20

# NAOS System: Editor

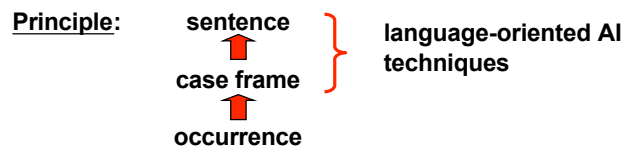
The NAOS system used artificially generated trajectory input data



# NAOS System: Recognizing a Turning-off Occurrence



## Generating a Natural-language Description in NAOS



### Problems:

- Which occurrences should be selected for verbalization?
- Which deep cases should be filled?
- Which additional time or location information is required?
- In which order should the information be presented?

### Solution:

Speech planning based on hearer simulation

informing a hearer  $\Leftrightarrow$  enabling a hearer to imagine the scene

23

## Standard Plan for Generating Natural-language Scene Descriptions in NAOS

- rules which assure that the hearer will be able to imagine the scene
- summary + descriptions of all object trajectories, each in chronological order
- no explicit hearer simulation

### Description of an object trajectory

1. Each time interval is described by the most special occurrence
2. The first occurrence begins at the beginning of the scene
3. The next occurrence follows in temporal order
4. Location information is given by prepositional expressions as required
5. Temporal information is given by prepositional expressions or references to other occurrences as required

24

## Lessons Learnt from Early Work on Scene Interpretation

- Scene interpretation requires representation and recognition of object motions and change.
- Representations may involve taxonomies and partonomies.
- Representations may be in quantitative and/or qualitative terms.
- Representations may involve temporal and spatial constraints on objects.
- Recognition may be incremental or post-mortem.
- A natural-language description is one possible form of a high-level scene interpretation.
- The success of scene interpretation experiments and applications depends heavily on the quality of low-level image analysis.