

Sommersemester 2006 • Institut für Germanistik I

Vorlesung Computerphilologie

Themenfeld Semantische Repräsentationen

„Wie kann man Inhalte in einem Text mit dem Computer systematisch finden, repräsentieren und verarbeiten?“

Aufgaben der semantischen Analyse

Zugrunde liegt das Kompositionalitätsprinzip G. Freges

nach Menzel

Semantische Analyse besteht aus den drei Schritten:

- Konstruktion (d. h. Aufbau einer Repräsentation),
 - Resolution (d. h. Ersetzungen oder Identifikationen) und
 - Auswertung (d. h. Antwort, Füllung aus einer Datenbank, Verifikation),
- Semantikkonstruktion
 - Ermitteln des semantischen Potentials einer Äußerung
 - Grundlage nur sprachliche Information:
 - Bedeutung der Wörter
 - Deren syntaktische Anordnung
 - keine Kontextinformation, keine zusätzlichen Wissensquellen

Das Ergebnis kann mehrdeutig sein.

Aufgaben der semantischen Analyse - 2 -

nach Menzel

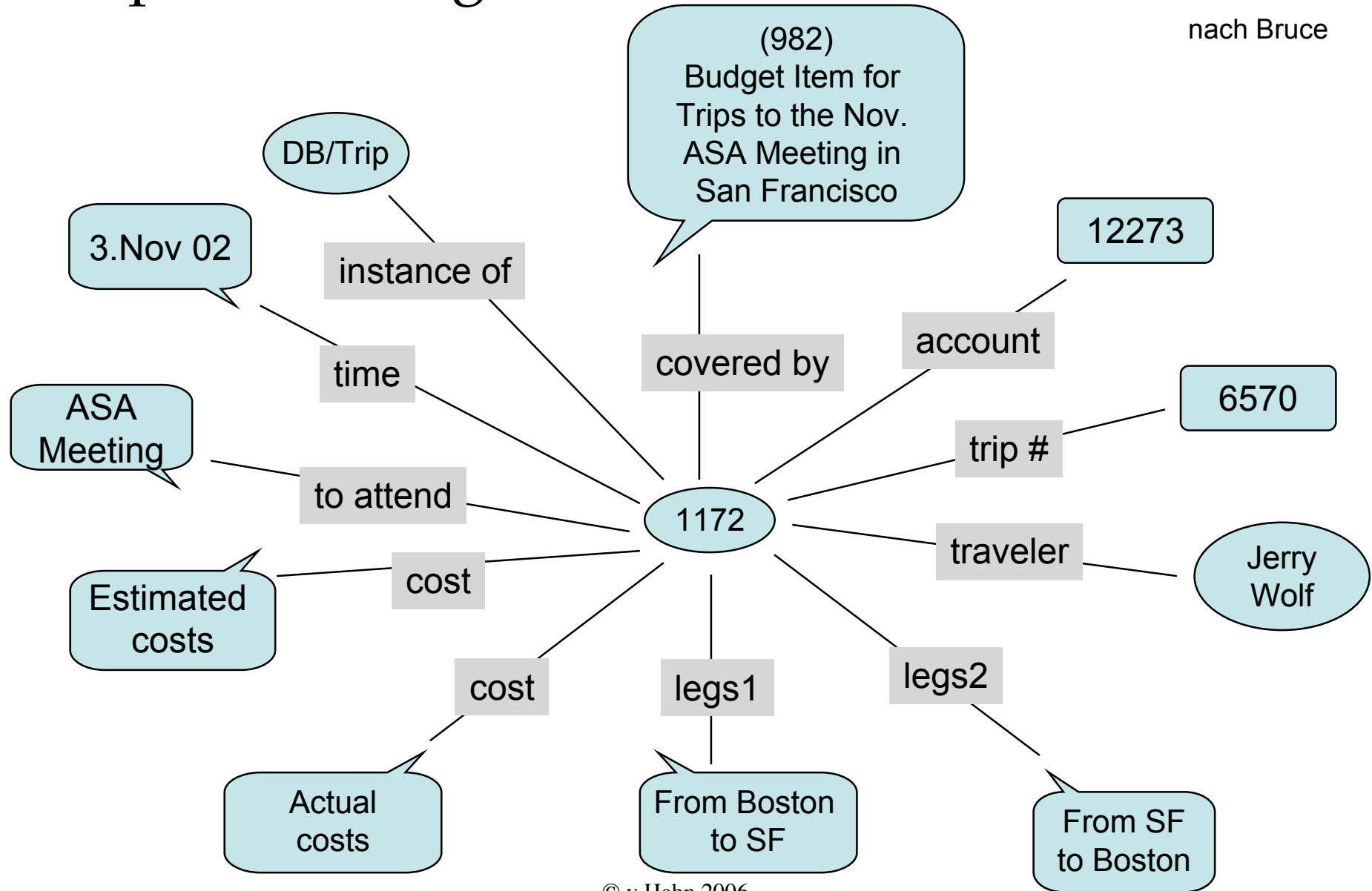
- Semantische Resolution
 - Ermitteln des aktuellen semantischen Werts im Äußerungskontext (meist “Kotext” genannt)
 - Desambiguierung
 - Identifikation der Referenzobjekte
 - Grundlage: Kontextwissen
- Semantische Auswertung
 - Ermitteln relevanter Äußerungsinhalte
 - interpretationsspezifische Schlußfolgerungen
 - Anwendung von Schlußfolgerungen auf dem semantischen Wert
 - Grundlage: Weltwissen und Alltagswissen

Designentscheidungen für die Semantische Verarbeitung

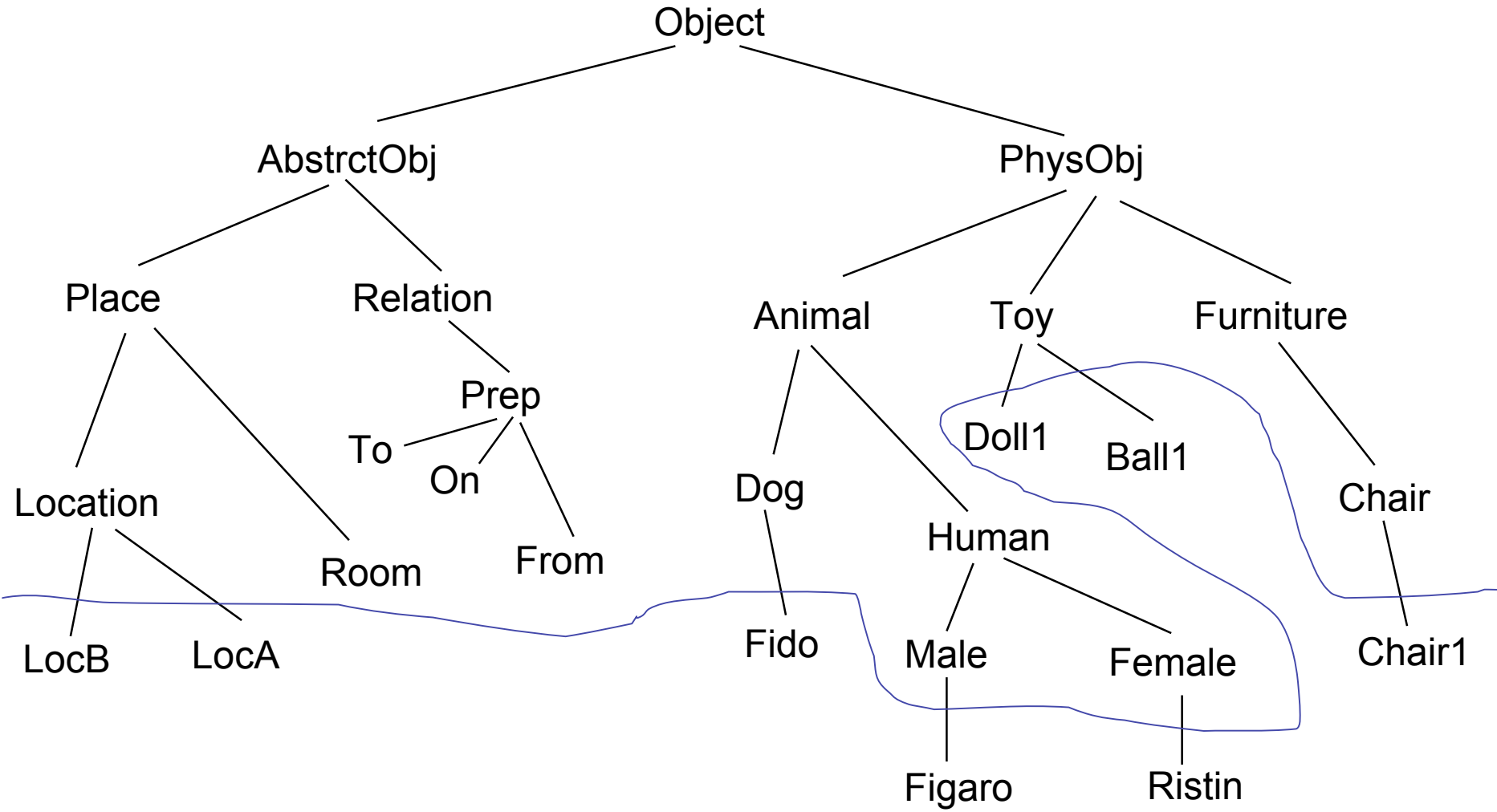
- Semantik im Parser? \Rightarrow Semantischer Parser / Kasusgrammatik
- Trennung von Konstruktion, Resolution und Auswertung? \Rightarrow Mehrphasenverarbeitung vs ganzheitliche Extraktion
- Frameorientierte Verarbeitung oder kompositionelle Verfahren
- Domänenwissen in der Semantik? \Rightarrow Referenzsemantik
- Konzeptuelles Wissen, Referenz und Faktenwissen trennen?
- Rezeptionsspezifische Interpretation? \Rightarrow Rezipientenmodell
- Zeitliche oder andere Abhängigkeiten? \Rightarrow Zeitlogik, modale Logiken

Script einer Flugreise

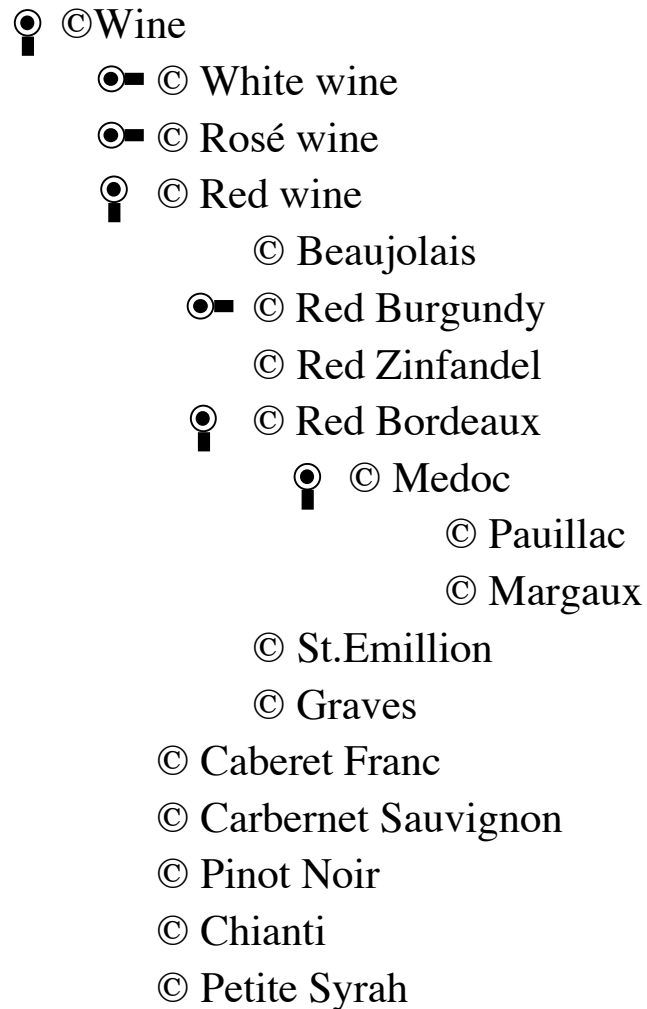
nach Bruce



Ontologien



Wie baut man eine Ontologie auf?



Darstellung einer
Ontologie in
Protégé-2000

Ontologien bestehen aus
Beziehungen zwischen
Konzepten oder Namen,
aber nicht Wörtern

Ontologie-Tools

- Protégé-20000 <http://protege.stanford.edu/>
- Ontolingua <http://www-ksl-svc.stanford.edu:5915/doc/frame-editor/guided-tour/index.html>
- Chimaera <http://www.ksl.stanford.edu/software/chimaera/>

Eine gute allererste Übersicht bietet:

- N.F.Noy und D.L.McGuinness: *Ontology Development 101: A Guide to Creating Your First Ontology*:
http://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html

Conceptual Dependency -1-

Entwickelt von

R.C. Schank 1973 ff

Objekt-Typen:

PP "picture producer", handelnde oder behandelte (belebte) Objekte

ACT Auswahl aus den 11 primitiven Aktionen

LOC Ort, Raumkoordinaten

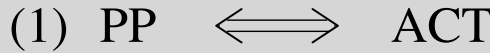
T "time", absolute oder relative Angabe auf einer Zeitachse

AA "action ailer", Modifikationen von ACTs

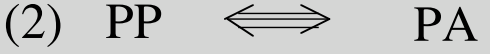
PA <Status><Wert>-Paare als charakteristische Merkmale von PPs

CD -2-

Relationen
zwischen
Objekten:



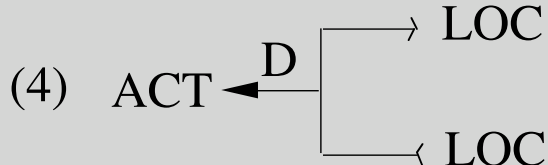
Certain PPs can ACT.



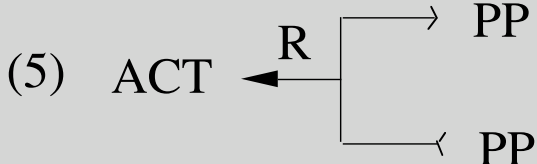
PPs (and some conceptualizations) can be described by an attribute.



ACTs have objects.



ACTS have directions.

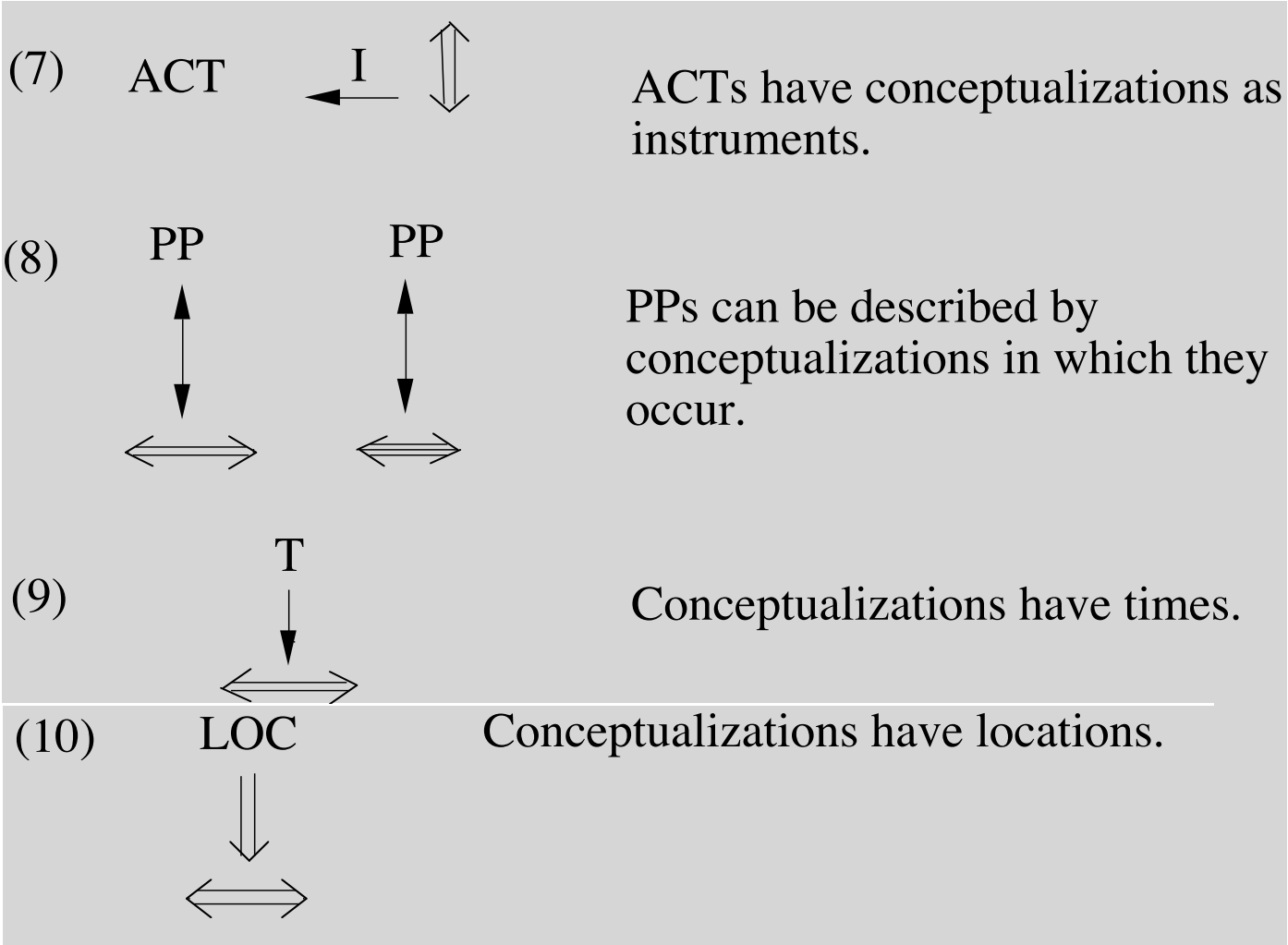


ACTs have recipients.

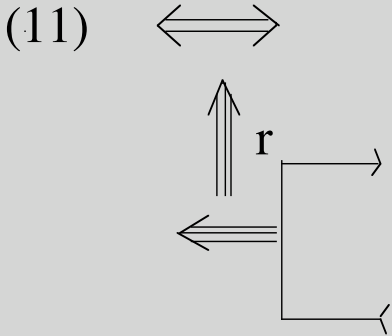


MTRANS requires conceptualizations or combinations of conceptualizations as objects, and MBUILD has its own special object type.

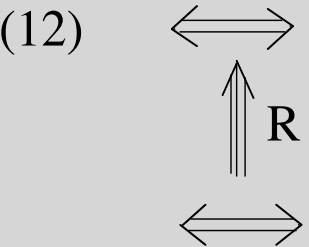
Conceptual Dependency - 3 -



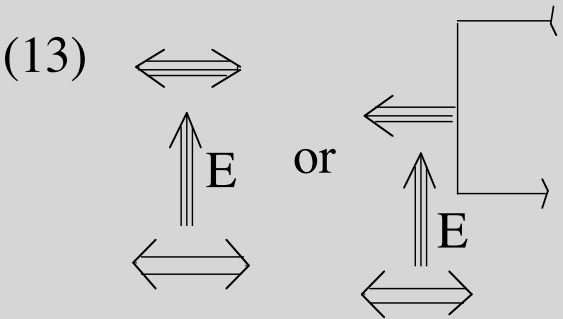
Conceptual Dependency - 4 -



Conceptualizations can result in state changes for PPs.



Conceptualizations involving mental ACTs can serve as reasons for conceptualizations.



States or state changes can enable conceptualizations to occur.

Conceptual Dependency - 5 -

(14) PP \longleftrightarrow PP One PP is equivalent to an instance of another PP.

(15) ACT
 ↑
 |
 AA

ACTs can be varied along certain dimensions (e.g., speed for motion ACTs).

Conceptual Dependency - 6 -

Primitive Actions:

| | | | | |
|--------|--------------------|--------|-------|-------|
| PROPEL | MOVE | INGEST | EXPEL | GRASP |
| PTRANS | (Phys. Transfer) | | | |
| ATRANS | (Abstr. Transfer) | | | |
| ATTEND | (Wahrnehm.) | | | |
| SPEAK | | | | |
| MTRANS | (Inform. Transfer) | | | |
| MBUILD | (Gedanken fassen) | | | |

tenses and modes:

| | |
|----|---------------------|
| p | past |
| f | future |
| / | negation |
| ts | start of transition |
| tf | end of transition |

| | |
|-----|---------------|
| c | conditional |
| k | continuous |
| ? | interrogative |
| ∞ | timeless |
| nil | present |

| | | |
|----------------|-----|----------------------|
| MEMORY: | CP | Conceptual Processor |
| | IM | Intermediate Memory |
| | LTM | Long Term Memory |

Conceptual Dependency - 7 -

| States: | Skales | | from - to |
|----------------|--------|-----|-------------------------------|
| Health | -10 | +10 | dead - perfect health |
| Fear | -10 | 0 | terrified - calm |
| Anger | -10 | 0 | furious - calm |
| Mental State | -10 | +10 | catatonic - ecstatic |
| Physical State | -10 | +10 | dead - OK |
| Consciousness | 0 | +10 | unconscious - awake |
| Hunger | -10 | +10 | starving - stuffed |
| Disgust | -10 | 0 | nauseated - bothered (-2) |
| Surprise | 0 | +10 | surprised (5) - astounded (9) |
| Intelligence | -10 | +10 | moron - genius |

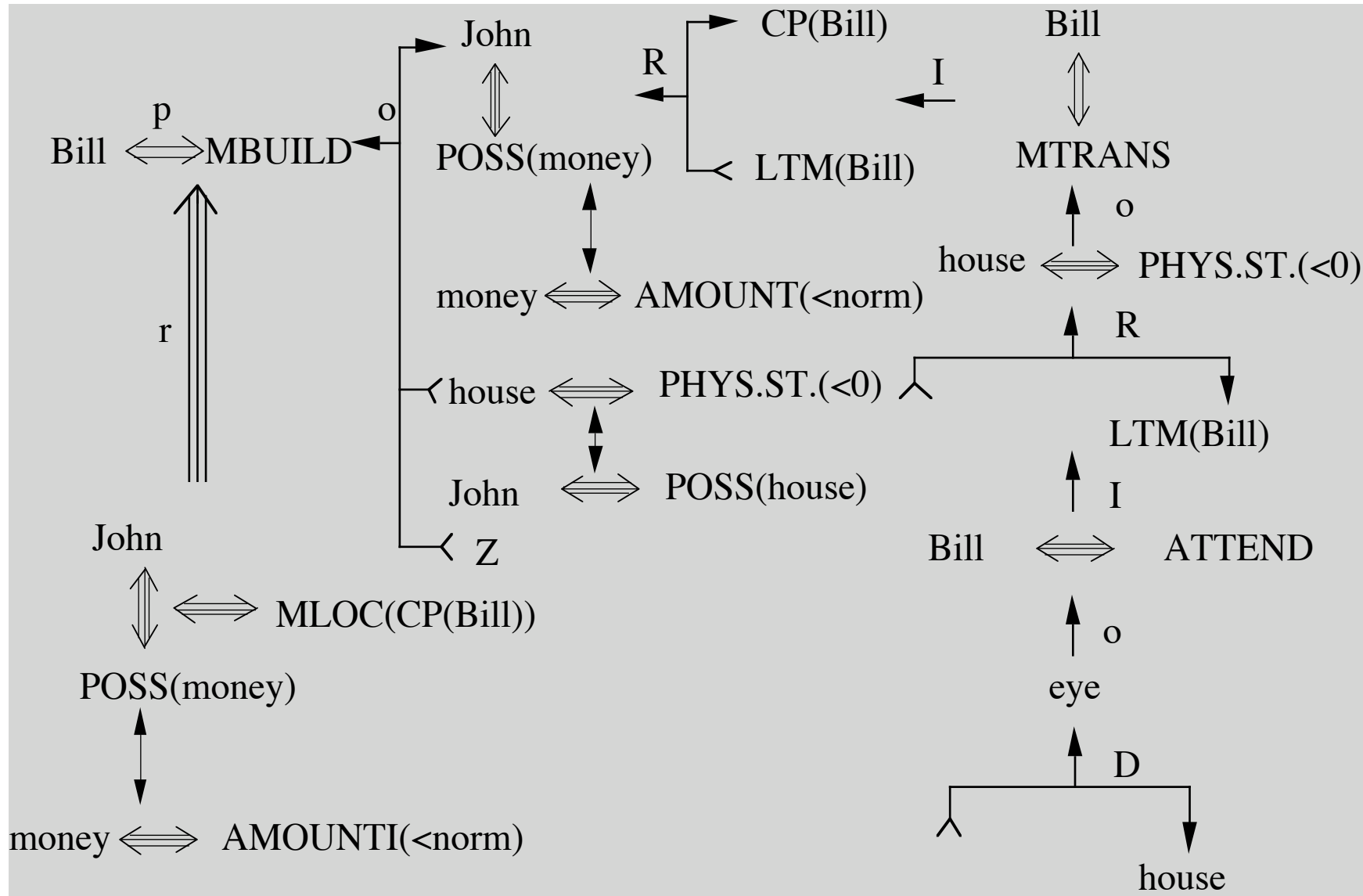
Nicht skalar: Length, Color, Light Intensity, Mass, Speed, Control

Sonstige: Part, Poss, Ownership, Contain, Proximity, Location, Phys.Contents

"Bill became aware of John's poverty when he saw the state of his house"

CD - Beispiel

Schank Conceptual mod. S. 64



Script - 1 -

Schank/Abelson Scripts, Plans, Goals S. 43

Script: RESTAURANT
 Track: Coffee Shop
 Props: Tables
 Menu
 F-Food
 Check
 Money

Roles: S-Customer
 W-Waiter
 C-Cook
 M-Cashier
 O-Owner

Entry conditions: S is hungry.
 S has money.

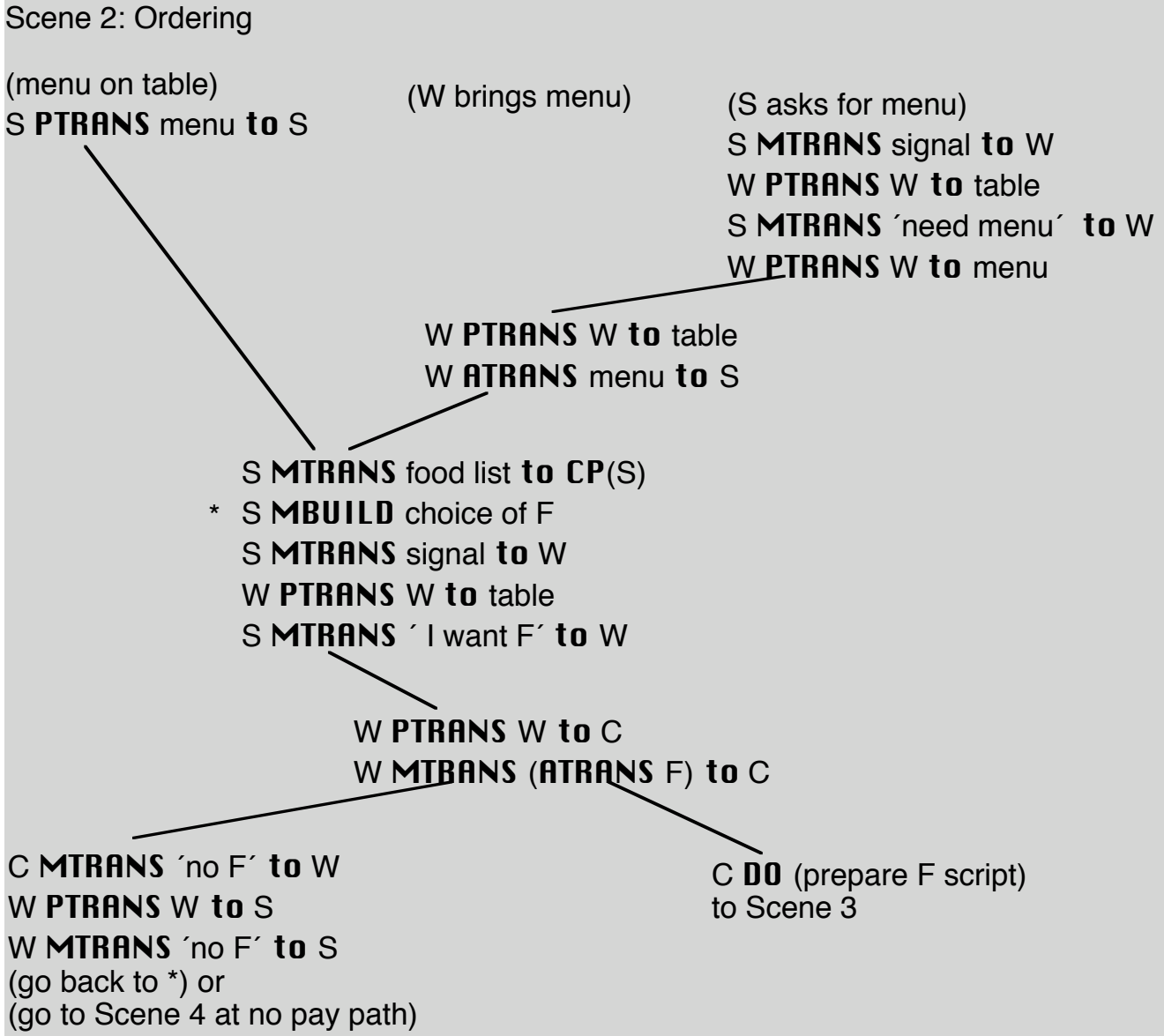
Results: S has less money.
 O has more money.
 S is not hungry.
 S is pleased (optional).

Scene 1: Entering

S **PTRANS** S into restaurant
 S **ATTEND** eyes **to** tables
 S **MBUILD** where **to** sit
 S **PTRANS** S **to** table
 S **MOVES** S **to** sitting position

Script - 2 -

Schank/Abelson:
Scripts, Plans, Goals S. 43



KL-ONE

Semantisches Netzwerk (Brachman)

Concept: Formales Objekt zur Darstellung von intensionalen Objekten in der Domäne, ihren Eigenschaften und Beziehungen.

Es gibt zwei Typen von Concepts, generische, d.h. Klassen von Individuen und individuelle

Role: Beziehungen zwischen den Concepts

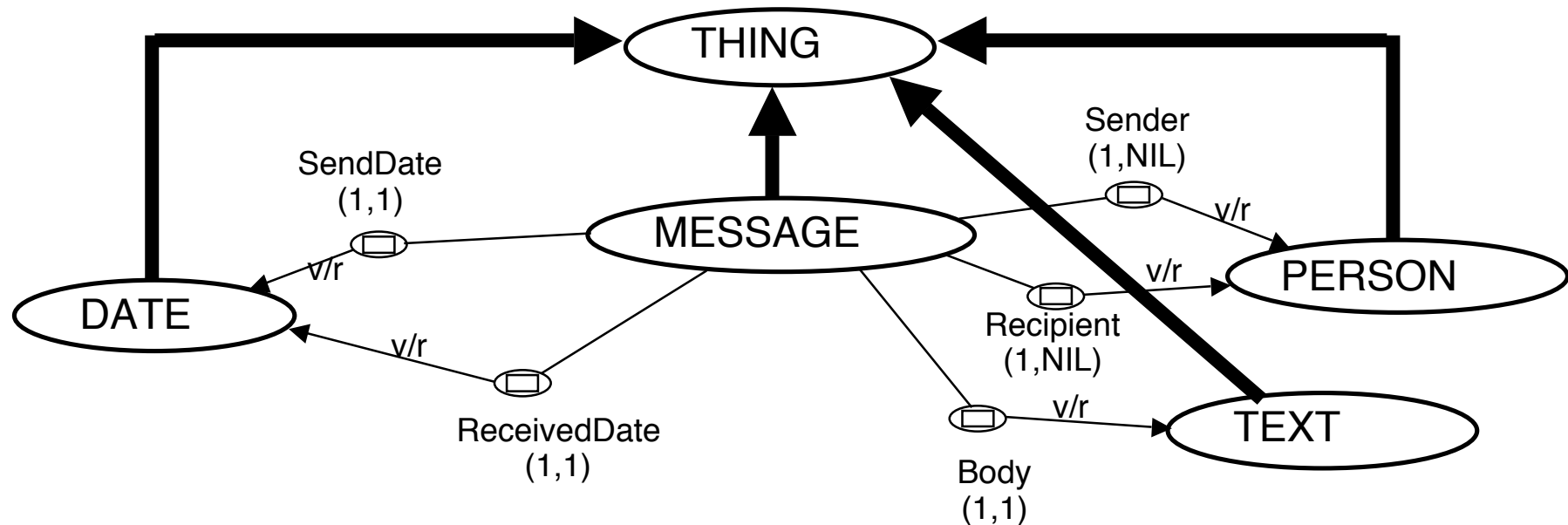
Role-Value-Maps (RVM)

Gleichheit oder Teilmengenbeziehung zwischen Rollenfüllern

Structural Description (SD)

Beliebige Beziehungen zwischen Rollenfüllern

KL-ONE: Beispiel



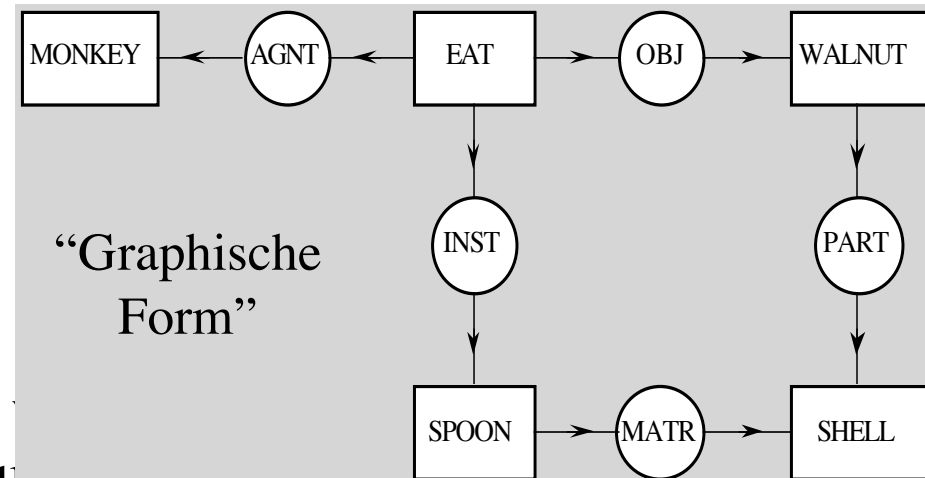
"A MESSAGE is, among other things, a THING with at least one Sender, all of which are PERSONs, at least one Recipient, all of which are PERSONs, a Body, which is a TEXT, a SendDate, which is a DATE, and a ReceivedDate, which is a DATE."

Conceptual Graphs

Conceptual Graphs (Sowa 84): Ein Wissensrepräsentationsformalismus für

- natürlichsprachliche Semantik
- Wissensbasen (z.B.Ontologien)
- Datenbanken
- Konzeptuelle Modellierung
- Theorembeweisen

- Beispiel: "A monkey eating a walnut with a spoon made out of the walnut's shell"

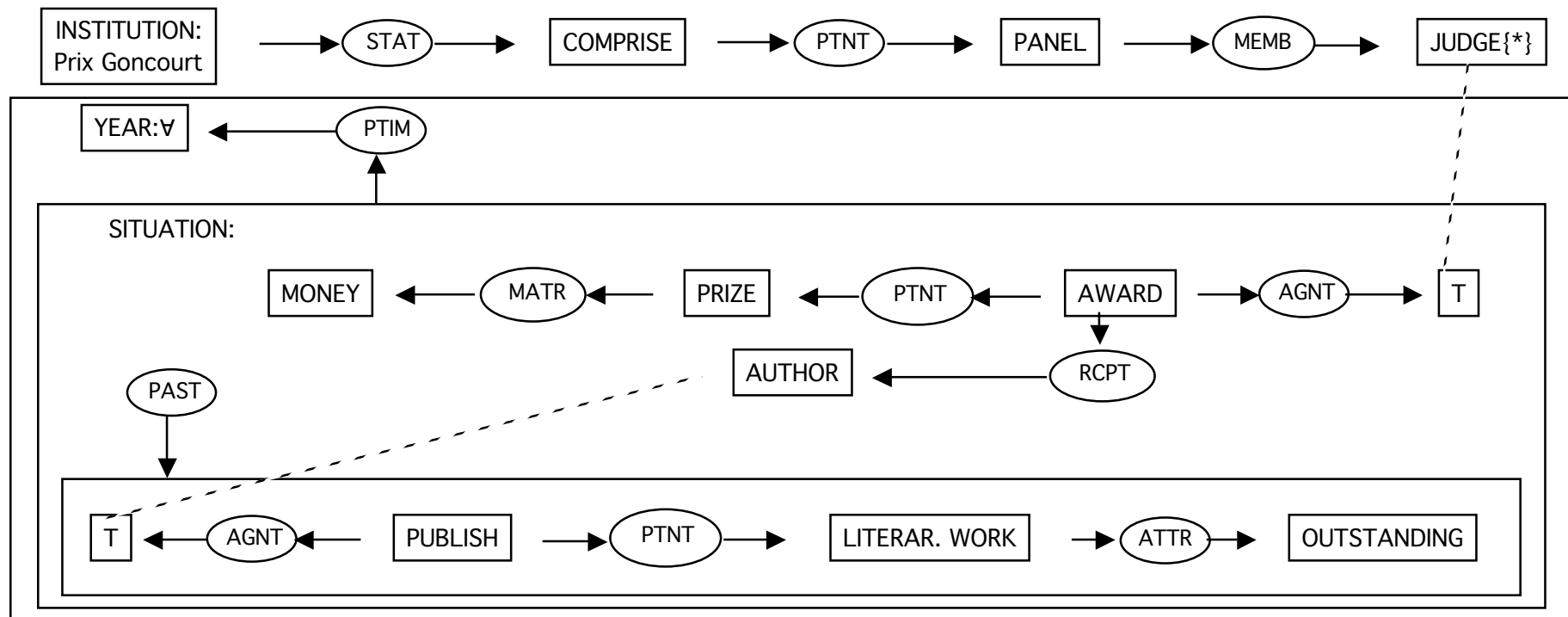


```
[ EAT ] -
  ( AGNT ) -> [ MONKEY ]
  ( OBJ ) -> [ WALNUT : *x ]
  ( INST ) -> [ SPOON ] -> ( MATR ) -> [ SHELL ] <- ( PART ) <- [ WALNUT : *x ] .
```

„Lineare Form“

- Konzepte und Relationen ("semantisches Netz")
- logisches Kalkül mit Sorten (Verband)

Sowa's CGs Beispiel



Graph für "Prix Goncourt"

Conceptual Graphs und natürliche Sprache

Ziel der CGs:

Den propositionalen Gehalt natürlicher Sprache so einfach und direkt wie möglich auszudrücken.

Gewöhnliche Nomen, Verben, Adjektive:

lady => [LADY], *dance* => [DANCE], *happy* => [HAPPY]

Eigennamen: *Yojo* => [CAT: Yojo]

Kontextuell definierte Referenten: *the cat* => [CAT: #], *this* => [T: #this]

Plurale: *ladies* => [LADY: {*}]

Anzahl: *nine ladies* => [LADY: {*}@9]

Phänomene natürlicher Sprache

Tempora: *Tom went*

(PAST)->[SITUATION: [PERSON: Tom]<-(AGNT)<-[GO]].

Koreferenzvariablen: *Mary is a teacher*

[PERSON: Mary*x] [TEACHER: #*x].

Kasusrahmen: *I sent a book to Fred by mail*

(PAST)->[SITUATION: [SEND]-

(AGNT)->[PERSON: #I]

(PTNT)->[BOOK]

(RCPT)->[PERSON: Fred]

(INST)->[MAIL]].

Konjunktionen: Relationen zwischen Kontexten

Restriktive Relativsätze: Lambda-Ausdrücke

Probleme der semantischen Analyse - 1 -

Skopus-Festlegung

(1) Quantoren: "Jeder Abteilungsleiter liest einen Bericht" kann heißen:

$\forall m . \text{TYPE} (m, \text{ABTEILUNGSLEITER}) \dots$

$(\exists d . \text{TYPE} (d, \text{BERICHT}) \ \& \ \text{LIEST} (m,d))$ (Lesart: viele Berichte)

oder:

$\exists d . \text{TYPE} (d, \text{BERICHT}) \ \& \ (\forall m . \text{TYPE} (m, \text{ABTEILUNGSLEITER}) \dots \text{LIEST} (m,d))$
(Lesart: ein einziger Bericht)

(2) Negation "Jeder Abteilungsleiter liest nicht einen Bericht" kann heißen:

$\forall m . \text{TYPE} (m, \text{ABTEILUNGSLEITER}) \dots \neg \text{LIEST} (m, \text{BERICHT})$

(Lesart: keiner liest)

oder

$\neg (\forall m . \text{TYPE} (m, \text{ABTEILUNGSLEITER}) \dots \text{LIEST} (m, \text{BERICHT}))$

(Lesart: nicht jeder liest)

Probleme der semant. Analyse - 2 -

(3) Satzoperatoren und Satzadverbien

"Ich glaube, daß ..." "hoffentlich ...", "vermutlich ..."

(4) Modifizierende Ausdrücke:

(1) Hecken: "ziemlich",

(2) Nonintersectives: "angeblich", "durchschnittlich", "Scherz-"

(5) Genitiv-Interpretation:

"Die Begründung der Projektgruppe ist mangelhaft" d.h. jemand begründet die Projektgruppe) vs (die Projektgruppe begründet)

(6) Koordination:

"XR1 und XR2 werden im Probetrieb gefahren" (jeweils?)

"Philips und Siemens erhalten EG-Mittel" (einzeln oder zusammen?)

"Philips und Siemens kooperieren" (zusammen, nicht jeder mit einem anderen)

Probleme der semant. Analyse - 3 -

Vagheit

Semantische Unbestimmtheit (Einteilung nach Pinkal)

1. Mehrdeutigkeit

1.1 Homonymie

"*Schloß*"

1.2 Polysemie

(abgetrennte Bedeutungen)

1.3 syntaktische Ambiguität

"*Möchtest Du Kaffee oder Tee?*"

1.4 Referentielle Vieldeutigkeit

"*dort*"

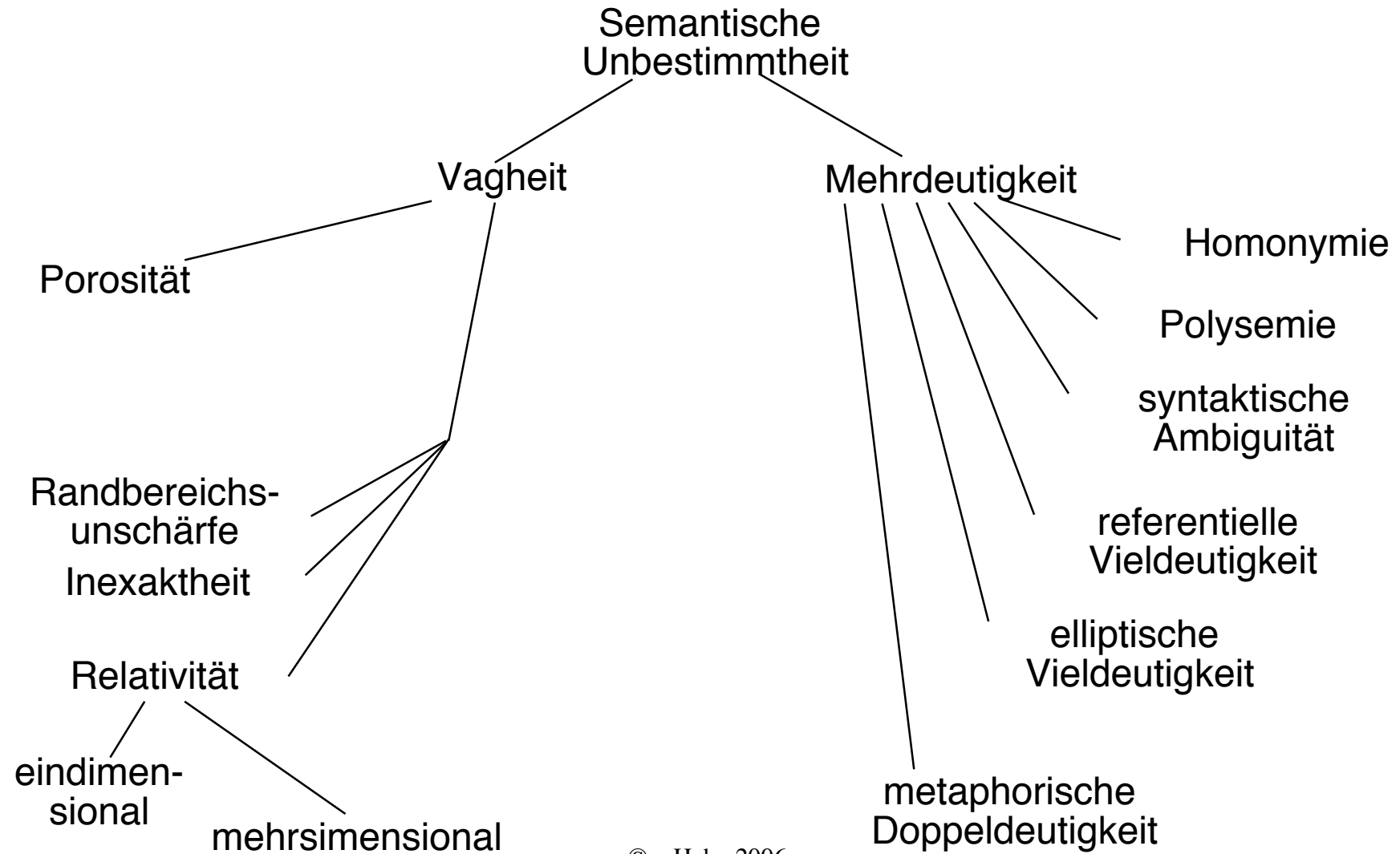
1.5 Elliptische Vieldeutigkeit

"Fangen wir an!"

1.6 Metaphorik

"*Kopf*"

Vagheit (nach Pinkal)



Probleme der semant. Analyse - 4 -

2. Vagheit im engeren Sinne

2.1 Porosität "*Schwergutfrachter*" (nicht für alle denkbaren Schiffe im voraus definiert)

2.2 Relativität

2.2.1 eindimensional "*groß*", "*langsam*" etc.

2.2.2 mehrdimensional "*schwierig*"

2.3 Inexaktheit "*Otto ist 1,80 m groß*" (oder nahe an diesem Wert)

2.4 Randbereichsunschärfe "*rund*", "*hier*", "*am Hals*" (unklare Grenze des Wahrheitsbereichs)

3. Ontologische Vagheit "begabt"

Weitere Repräsentationsprobleme

Kontinua

Ausgedehnte Objekte

"Natürliche" Mengenangaben

Defaults

Common Sense

Naive Physics

Beliefs

Wasser

Bürgersteig

einige, viele

Klaviere: meist schwarz

Reparatur --> macht Lärm

Vakuum ist "in" einem Behälter

Alle Studenten sind faul

Auswertung (Evaluation)

- Sammelbegriff für eine Vielzahl von Prozessen auf einer semantischen Repräsentation, die teilweise sequentiell, teilweise ineinander verschachtelt ablaufen
- Art, Anzahl und Anordnung von System zu System unterschiedlich
- keine einheitliche Theorie, stark auf Heuristiken beruhend
- Ausgangs- und Zielsprache von Auswertungsprozessen:
 - i.a. die semantische Repräsentation (sprachunabhängig)
- Referenzanalyse (Anaphora, Kataphora)
- Quantorenauswertung
- Prüfung von Präsuppositionen
- Inferenzen unter Ausnutzung des Weltmodells

Auswertung (Evaluation)

- Abbildung der semantischen Repräsentation auf “Interpretations“-System (Hermeneutischer Entwurf)
- Interpretation des Ergebnisses
- Oft sehr komplexe Kontrollstruktur
- Oft mehrfaches Backtracking
- Mehrfach ineinander geschachtelte Prozesse:
 - Präsuppositionsüberprüfung
 - Quantorenauswertung
 - Mehrfachlösungen
 - Schreib- / Scannfehlerkorrektur

Referenzauflösung

Einfachster Fall: Namen

“Die Gräfin Eboli tritt herein”

⇒ eindeutige Objektkennzeichnung

Syntaktischer Fall: Komplexe NP

“Eine grüne Wanduhr mit römischen Ziffern“ ⇒ Objektidentifikation

Semantischer Fall: logische Beziehungen zwischen den Konstituenten
(Anaphorik)

“Der erste von den drei Genannten ...”

Pragmatischer Fall: Sprecher-Origo-Bezug (Deixis)

“Der Herr, der vorhin dort stand ...”