

Kapitel 2.5 - Linguistische Analyse von Domänenbeschreibungen

SWT I – Sommersemester 2010

Walter F. Tichy, Andreas Höfer, Korbinian Molitorisz

IPD Tichy, Fakultät für Informatik



Linguistische Analyse von Domänenbeschreibungen

- Warum tut man sowas als Softwareingenieur?
- Antwort: Um den Kunden zufrieden zu stellen!
- Problem: Benutzer kann die Software nicht bedienen.
 - Er weiß nicht, **welche** Funktionen überhaupt zur Verfügung stehen.
 - Er weiß nicht, **was** die Funktion tut, die ihm gerade angeboten wird.
 - Er weiß nicht, **wo** er die Funktion findet, die er gerade braucht.

Zusammenhang der Begriffswelten

- Ergonomischer Lösungsansatz: Die Software soll im Innern mit den selben Begriffen arbeiten, wie ihr Benutzer „draußen“.
 - Es stehen ihm die selben Funktionen im Programm wie in der Realität zur Verfügung.
 - Die Funktionen tun im Programm das selbe wie in der Realität.
 - Die Funktionen finden sich im Programm da, wo sie sich in der Realität auch finden.
 - Beispiel: WYSIWYG Textverarbeitung (Word etc.)
- Grundidee der objektorientierten Programmierung!

OO-Analyse – Vorgehen

■ Vorgehensweise (abstrakt)

- Betrachte die Realität.
(i.e. OO-Analyse)
- Modelliere den **Ausschnitt**,
der von dem Programm
unterstützt werden soll. (z.B. UML-Modell)
- Mache das Modell **lauffähig** (i.e. erstelle z.B. ein Java-Programm)
- **Simuliere** den gewählten Ausschnitt der Welt (i.e. führe das Programm aus)

Dieser Schritt wurde Anfangs nicht bewusst (explizit) gemacht. Man hat statt dessen aus der Analyse direkt ein Programm erstellt. Heute ist das Modell ein zusätzliches Artefakt, mit dem der Softwareingenieur arbeitet: Er verfeinert und ergänzt das Modell so lange, bis er es für implementierungswürdig hält.

„Anwendungsdomäne“

⇒ Die Software arbeitet im Innern mit den selben **Dingen**, wie ihr Benutzer „draußen“.

Kompetenzprobleme

- Szenario: Wir wollen eine Banksoftware anbieten, mit der die Bank ihrerseits Produkte zur Kreditbesicherung berechnen kann
- Problem: Welche „Dinge“ gibt es denn in der Anwendungsdomäne?
 - Suche Softwareingenieur, der „beides“ gelernt hat (Hier: Informatik und Bankkaufmann)
 - Frage die „Domänenexperten“ (z.B. Kreditberater, Kundenbetreuer, Risiko-Manager, etc.)
⇒ Quelle der **Domänenbeschreibung**

- Anfänge: Russell J. Abbott; *Program Design by Informal English Descriptions*; ACM; 1983

“A technique is presented for developing programs from informal but precise English descriptions. The technique shows how to derive **data types from common nouns, variables from direct references, operators from verbs and attributes, and control structures from their English equivalents**. The primary contribution is the proposed relationships between common nouns and data types; the others follow directly. Ada is used as the target programming language because it has useful program design constructs.”
- Bekannter: Booch, KISS-Method (Kristen)

„Abbots Methode“ nach Brüggge

Textual Analysis (Abbot's Technique)

<i>Example</i>	<i>Part of speech</i>	<i>UML model component</i>
“Monopoly”	Proper noun	object
Toy	Improper noun	class
Buy, recommend	Doing verb	operation
is-a	being verb	inheritance
has an	having verb	aggregation
must be	modal verb	constraint
dangerous	adjective	attribute
enter	transitive verb	operation
depends on	intransitive verb	Constraint, class, association

© 2008 Bernd Bruegge

Software Engineering SS 2008

21

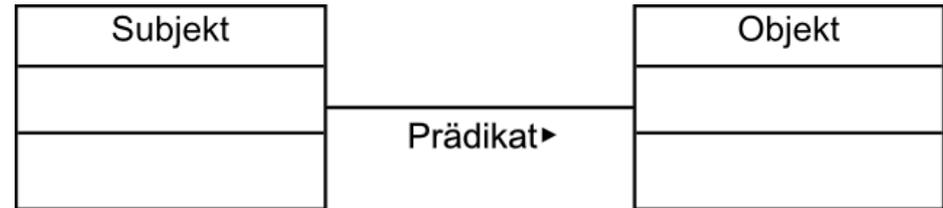
Quelle: http://www.bruegge.informatik.tu-muenchen.de/twiki/pub/Lehrstuhl/SoftwareEngineeringWiSe2008/L7_Analysis.pdf

Linguistische Analyse – Aber richtig!

■ Naiver Ansatz:

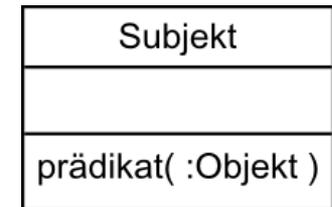
■ SPO \mapsto Beziehung

- Die Geschäftsstelle liegt in Karlsruhe. ✓
- Der Angestellte eröffnet Konten (für Kunden). ✗



■ SPO \mapsto Methode

- Der Angestellte eröffnet Konten (für Kunden). ✓
- Die Geschäftsstelle liegt in Karlsruhe. ✗



■ SpOOOO \mapsto ?

- Im ersten Schuljahr **ist** die allmähliche Umformung der Mundart oder Aussprache in die Schriftsprache, die Gewöhnung zum freien Sprechen, die Schärfung des Gehörs, die Vermehrung der Wort- und Sprachformen Gegenstand des Unterrichts.
(Beispiel aus Ludwig Reiners; *Stilkunst*, C.H.Beck; 2004)

- Weiteres Problem: Syntax-basierte Ansätze können nur mit Eingabetexten umgehen, die bereits **verbalisierte Entwürfe** sind!
 - Each adherent has one name.
 - Each adherent has one address.
 - Each adherent has at most one phone number.
 - Each book has one title.
 - Each book has at most one author.
 - Title is a string. Author is a name. Address is a string. (usw.)
(Beispiel aus Dignum et al.; *Constraint modelling using a conceptual prototyping language*; Elsevier; 1987)
- Von welchem Domänenexperten dürfen wir einen solchen Text erwarten?

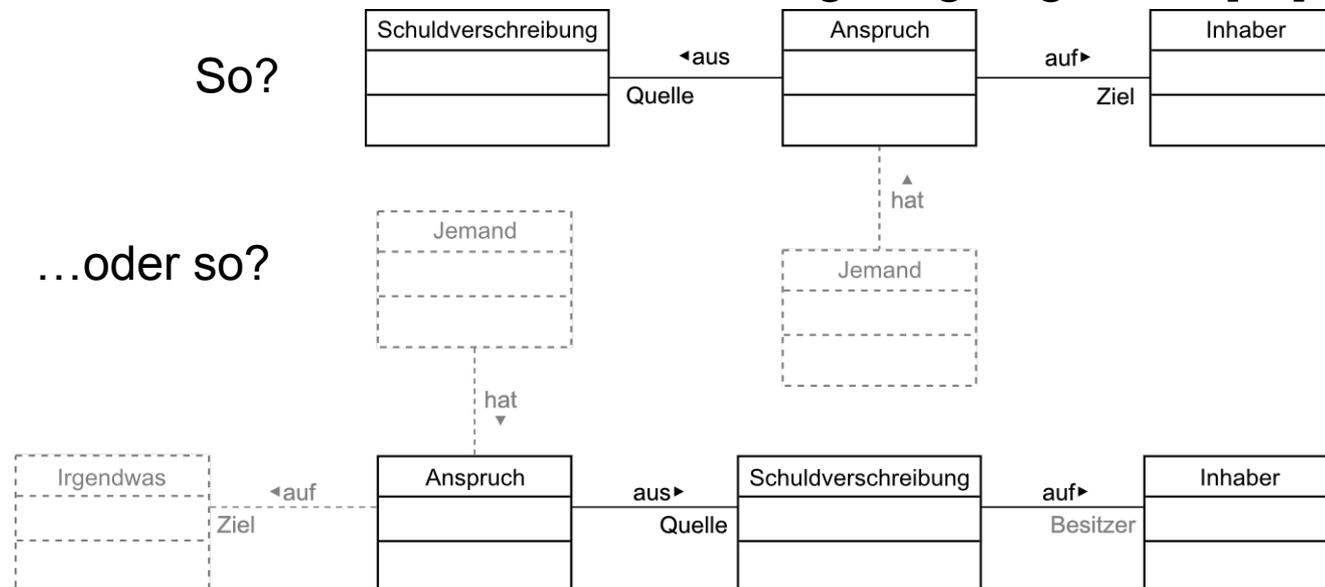
Gegenbeispiel: Beamten- und Ingenieursdeutsch

- BGB, Buch 2, Abschnitt 8, Titel 24
§ 801 Erlöschen; Verjährung

Subjekt

Der Anspruch aus einer Schuldverschreibung auf den Inhaber

erlischt mit dem Ablauf von 30 Jahren nach dem Eintritt der für die Leistung bestimmten Zeit, wenn nicht die Urkunde vor dem Ablauf der 30 Jahre dem Aussteller zur Einlösung vorgelegt wird. [...]



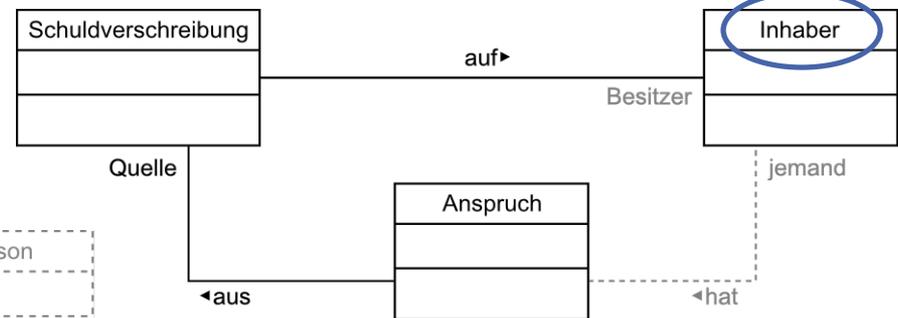
Gegenbeispiel: Beamten- und Ingenieursdeutsch

- BGB, Buch 2, Abschnitt 8, Titel 24

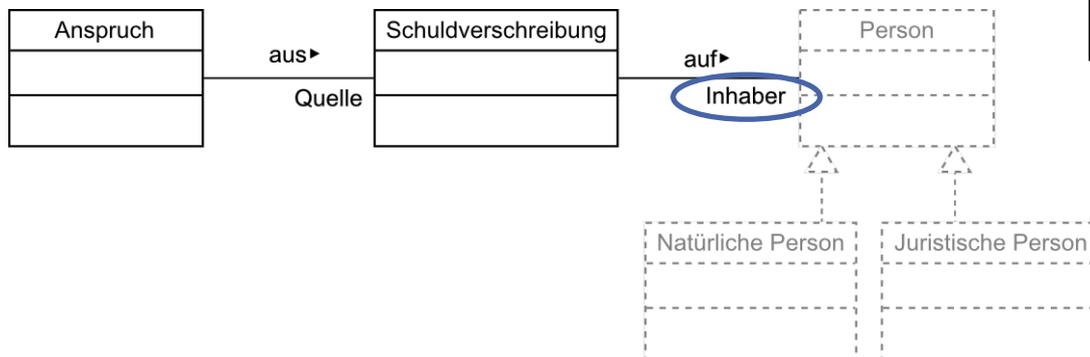
§ 801 Erlöschen; Verjährung

Der Anspruch aus einer Schuldverschreibung auf den Inhaber erlischt mit dem Ablauf von 30 Jahren nach dem Eintritt der für die Leistung bestimmten Zeit, wenn nicht die Urkunde vor dem Ablauf der 30 Jahre dem Aussteller zur Einlösung vorgelegt wird. [...]

Eigentlich ist es so:



...oder lieber so modellieren?



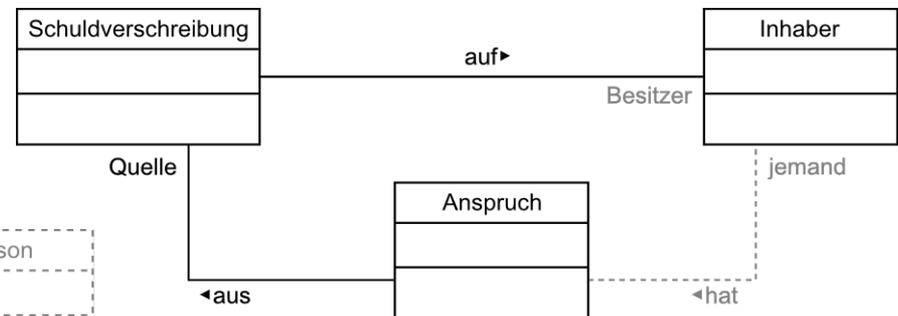
Gegenbeispiel: Beamten- und Ingenieursdeutsch

- BGB, Buch 2, Abschnitt 8, § 801 Erlöschen; Verjährung
Der Anspruch aus einer Schuld erlischt mit dem Ablauf von 30 Jahren ab dem Ablauf der Leistung bestimmten Zeit, wenn nicht die Urkunde vor dem Ablauf der 30 Jahre dem Aussteller zur Einlösung vorgelegt wird. [...]

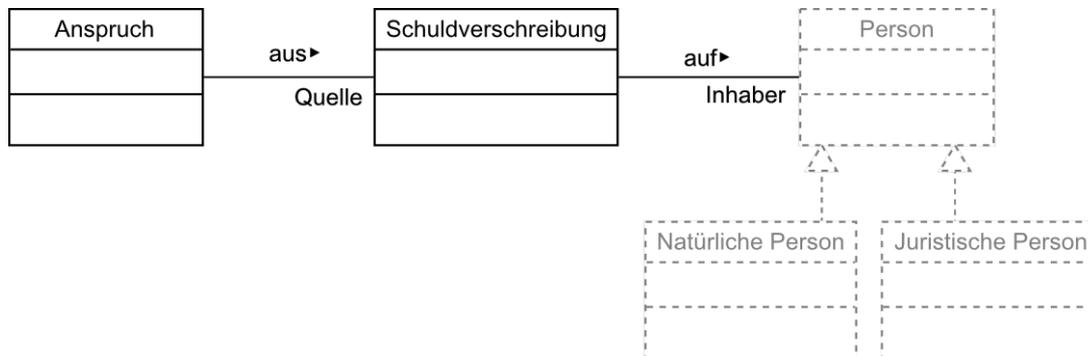
Offenbar gibt es zwei Freiheitsgrade:

- Die Entscheidung über die semantische Interpretation des Geschriebenen
- Die Entscheidung über die „beste“ (UML-) Modellierung des Inhalts

Eigentlich ist es so:



...oder lieber so modellieren?



Thematische Rollen

- Beobachtung: Die Syntax hat **keine Bedeutung!**
 - Fillmore (1969): „[...] traditional subject and object are not to be found among the syntactic functions to which semantic rules must be sensitive.“
 - The door will open.
 - The janitor will open the door.
 - The janitor will open the door with this key.
 - This key will open the door.
- } Wer öffnet die Tür?
- Lösung: Betrachte nicht die **syntaktischen**, sondern die **semantischen** bzw. **thematischen** Rollen.
 - Hier: Handelnder (*agens*, = janitor), Handlung (*actus*, = open), Behandler (*patiens*, = door) und Instrument (= key).
 - Die konkrete Syntax des Satzes ist dann unerheblich!

Linguistische Analyse mit thematischen Rollen

- Offene Wortklassen \approx Modellelemente
 - Nomen, Verben, Nominalphrasen, Verbphrasen, Teilsätze und Sätze spielen thematische Rollen und begründen Modellelemente, wenn sie einen semantischen Beitrag leisten.
 - Adjektive begründen Attribute oder Parameter.
- Geschlossene Wortklassen \approx Zusammenhang
 - Pronomen etc. sind Bezüge und werden aufgelöst.
 - Artikel, Präpositionen, Adverbien etc. codieren die Beziehungen (Wer spielt welche Rolle in Bezug auf wen?) und tauchen nicht im Ergebnismodell auf.

Spezifikation eines Ferraris

A close-up photograph of a Ferrari steering wheel, showing the central hub and the spokes. The wheel is dark, likely black or dark red, with a textured surface. The image is overlaid with several blue callout boxes containing white text, each pointing to a specific feature of the steering wheel.

450 PS

Schaltwippen am Lenkrad

Helles Leder

Aluminium Felgen

335/45 Reifen

Becherhalter

Was der Kunde wollte

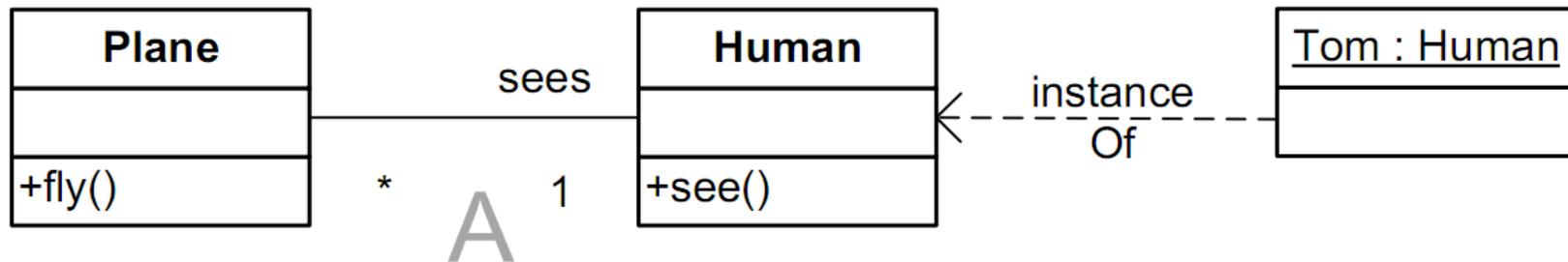


Das ist auch ein Lamborghini



Thematische Rollen – Syntax vs. Semantik

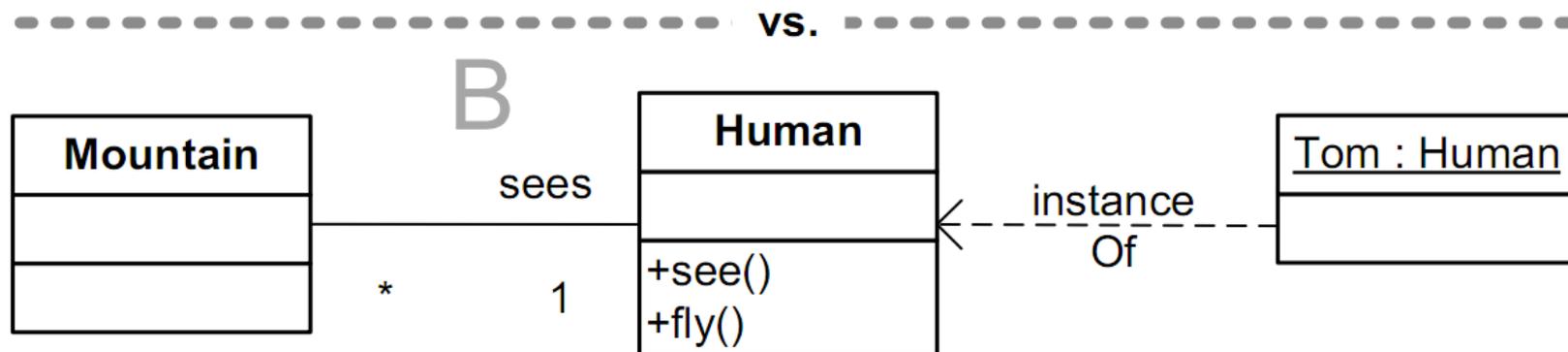
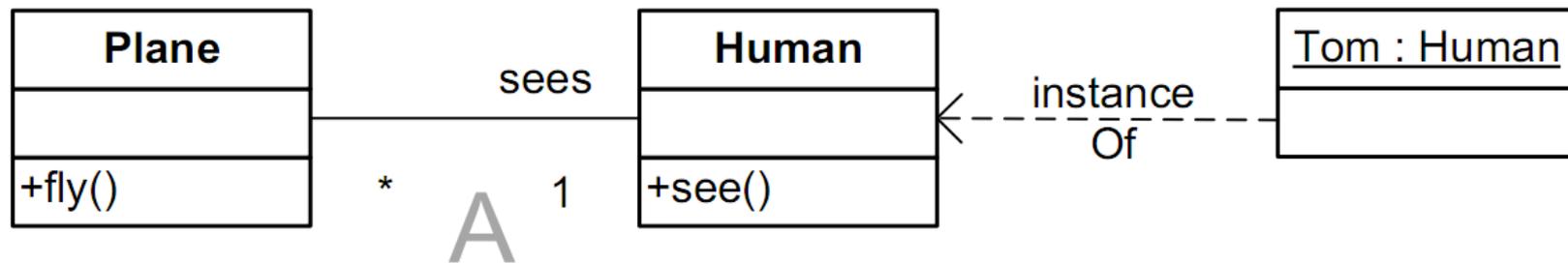
„Tom saw **the plane** flying.“



Thematische Rollen – Syntax hat keine Bedeutung

„Tom saw **the plane** flying.“

„Tom saw **the mountains** flying.“

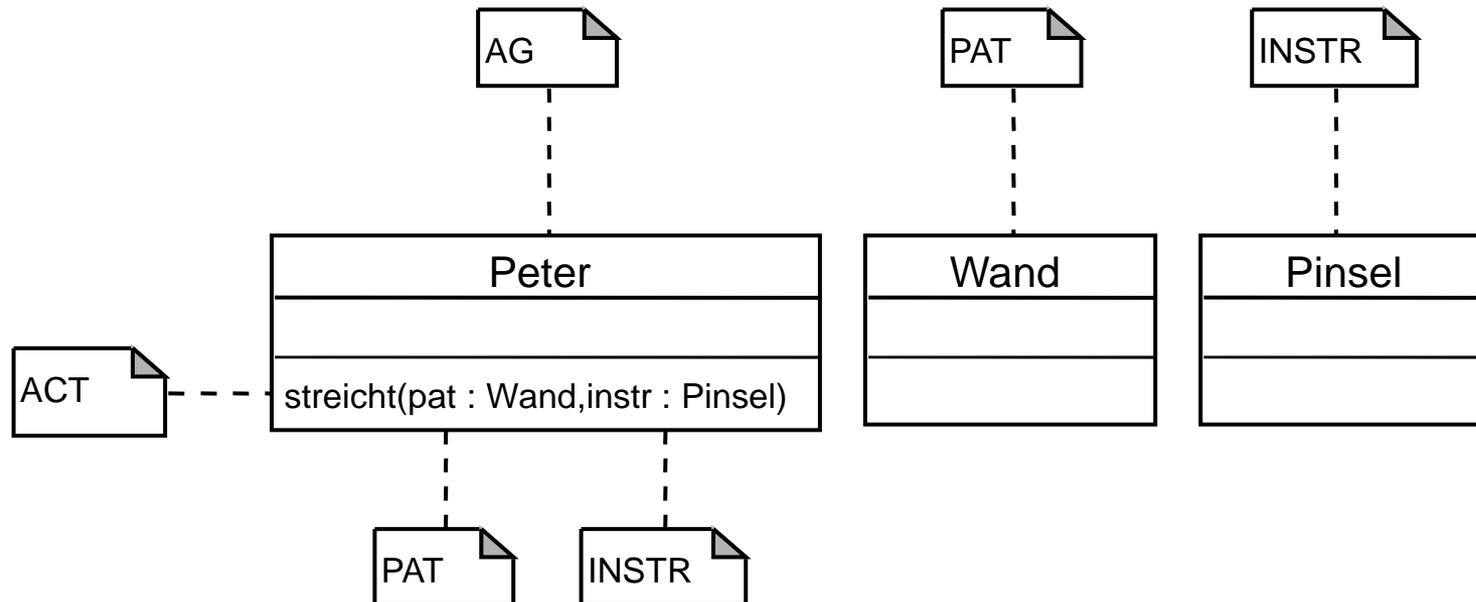


Thematische Rollen

- Einige für SW-Ingenieure interessante Rollen:
 - *agens* – der Handelnde.
patiens – der Behandelte.
Peter (AG) wirft einen Ball (PAT).
 - *actus* – die Handlung (für Tätigkeitsverben und deren Nominalisierungen)
Peter wirft (ACT) einen Ball. Peter tätigt einen Wurf (ACT) des Balls.
 - *instrumentum* – das Hilfsmittel
Peter streicht die Wand mit einem Pinself (INSTR).
 - *status* – der Zustand (für Zustandsverben und deren Nominalisierungen)
Peter wohnt (STAT) in Bonn.

Beispiele für Abbildungen: Methoden

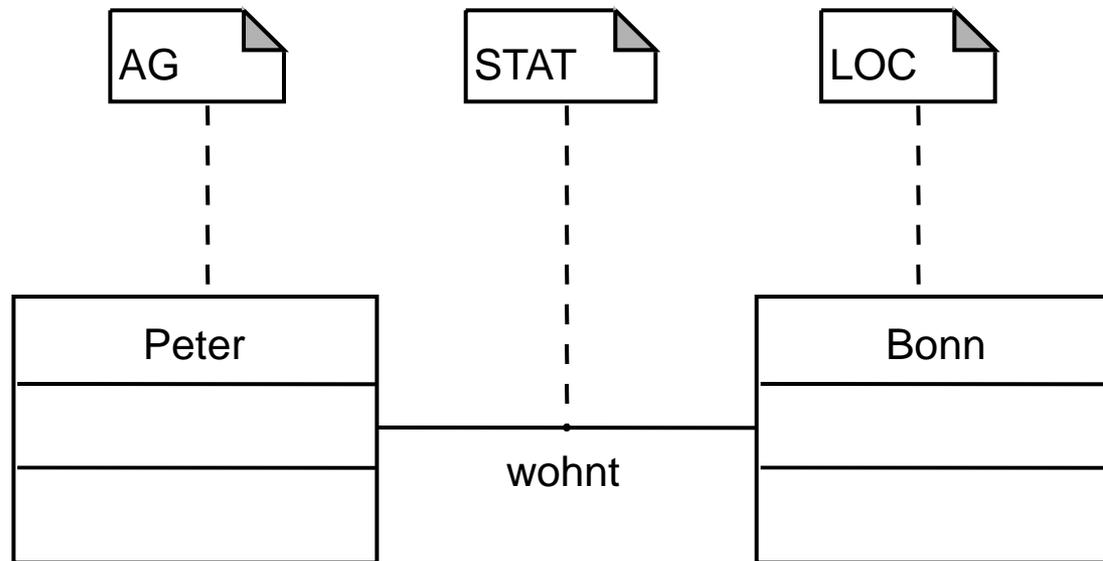
- Peter (AG) streicht (ACT) die Wand (PAT) mit einem Pinzel (INSTR).



Wir hätten auch schreiben können: „Die Wand (PAT) wird von Peter (AG) gestrichen (ACT), und zwar mittels eines Pinzels (INSTR).“

Beispiele für Abbildungen: Beziehungen

- Peter (AG) wohnt (STAT) in Bonn (LOC).



Beispiel für eine einfache Relation.

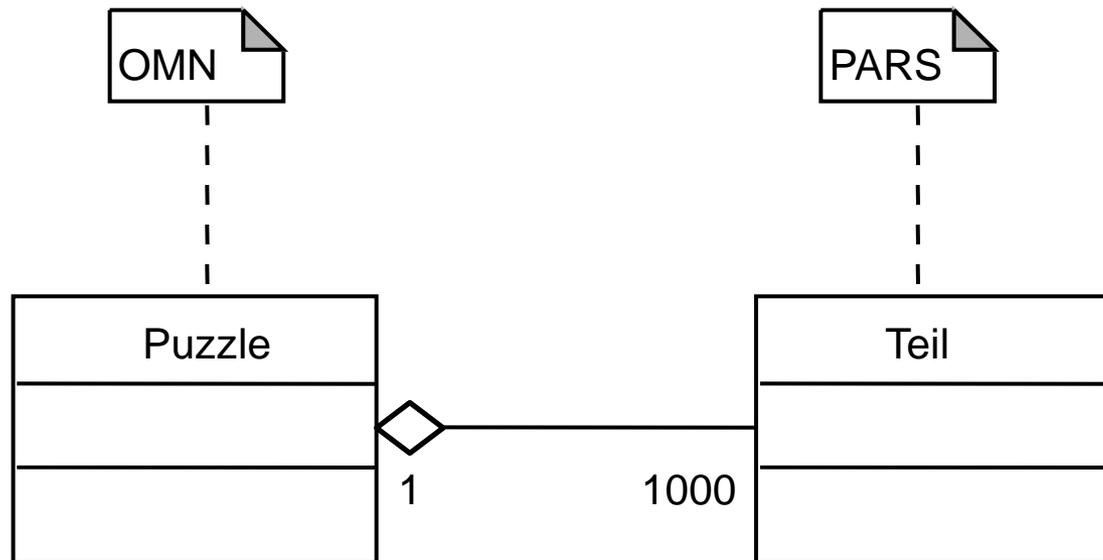
Thematische Rollen

- Einige für SW-Ingenieure interessante Rollen (Forts.):
 - *locus* – ein **Ort**
Peter ist in Frankreich (LOC) im Urlaub.
 - *omnium* – das **Ganze** (im Verhältnis zu seinen Teilen)
 - *pars* – das Teil (im **Verhältnis** zum Ganzen).
Peter puzzelt ein Puzzle (OMN) aus 1000 Teilen (PARS).

...u.v.m.

Beispiele für Abbildungen: Beziehungen

- Peter puzzelt ein Puzzle (OMN) aus 1000 Teilen (PARS).



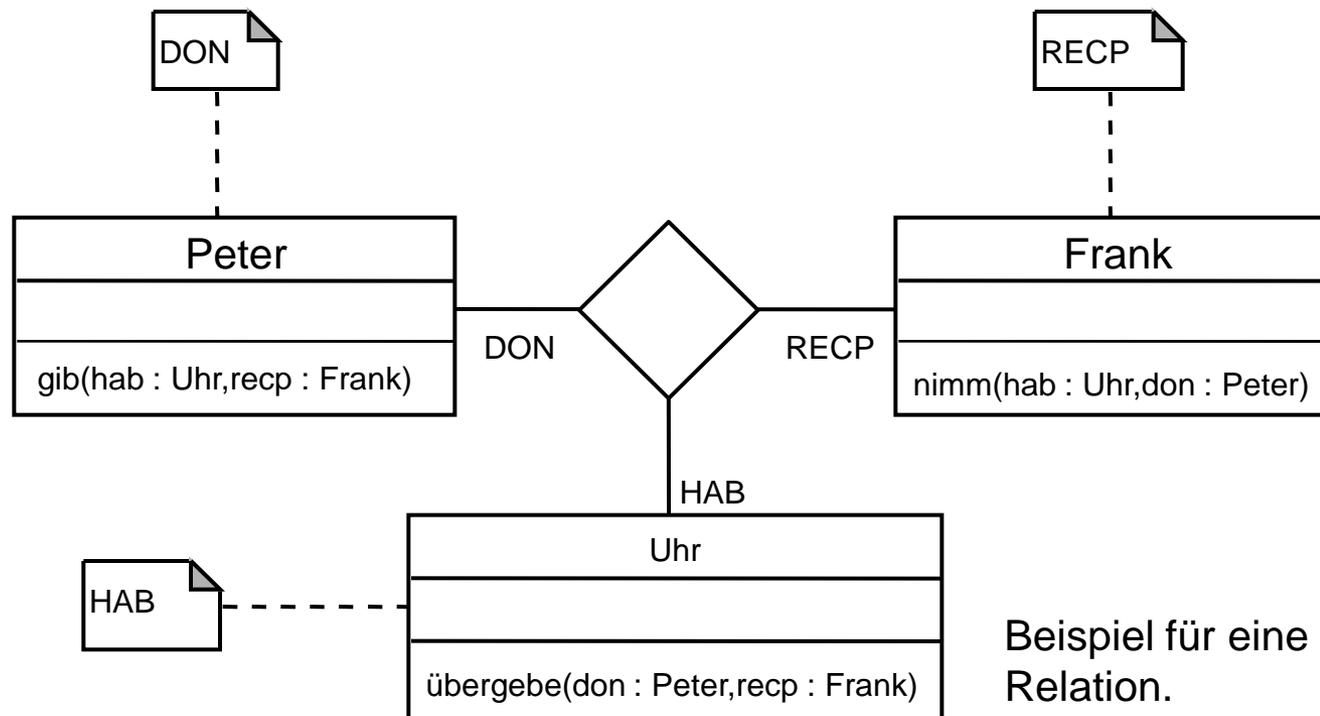
Beispiel für eine Kompositionsbeziehung.

Thematische Rollen

- Einige für SW-Ingenieure interessante Rollen (Forts.):
 - *donor* – der, der etwas **gibt**
 - *recipient* – der **Empfänger** einer Sache.
 - *habitum* – etwas, das **gegeben** oder „gehabt“ wird.
Peter (DON) schenk Frank (RECP) eine Uhr (HAB).

Beispiele für Abbildungen: Beziehungen

- Peter (DON) schenkt Frank (RECP) eine Uhr (HAB).



Beispiel für eine mehrstellige Relation.

■ Eine Spezifikation

```
A WHOIS_server listens on
TCP port 43 for requests from
WHOIS clients
    The WHOIS_client makes a
    text_request to the
    WHOIS_server
then the WHOIS_server replies
with text content
All requests are terminated with
ASCII CR then ASCII LF
```

Quelle: IETF WHOIS Protocol Specification, RFC 3912, Ch. 2, "Protocol Specification"

Arbeiten am Institut: Annotierter Text

■ Annotationsprache SAL_E

Referenz auf vorhergehende Instanz

Kommentar (1 Wort)

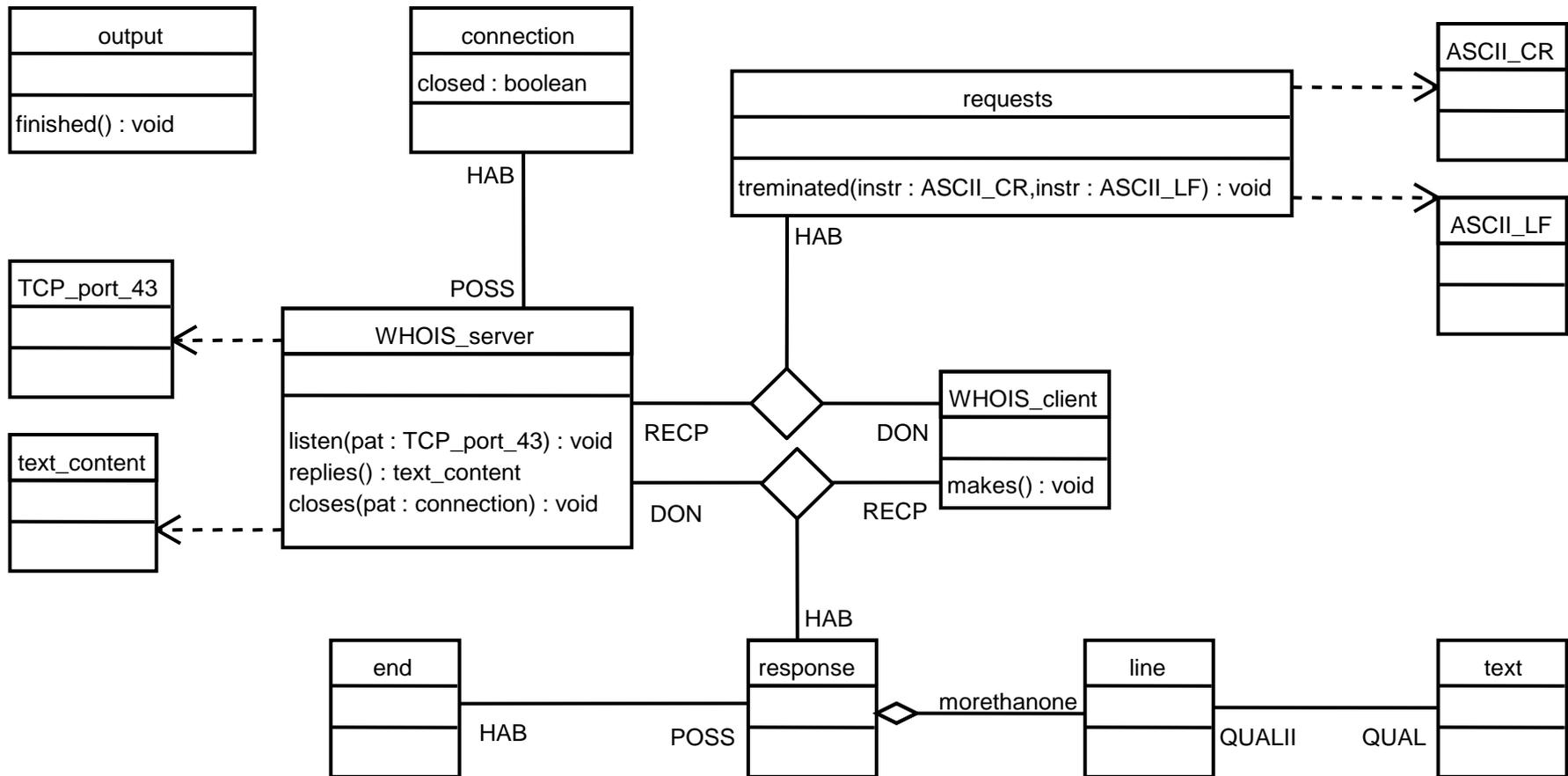
Relation

Rolle

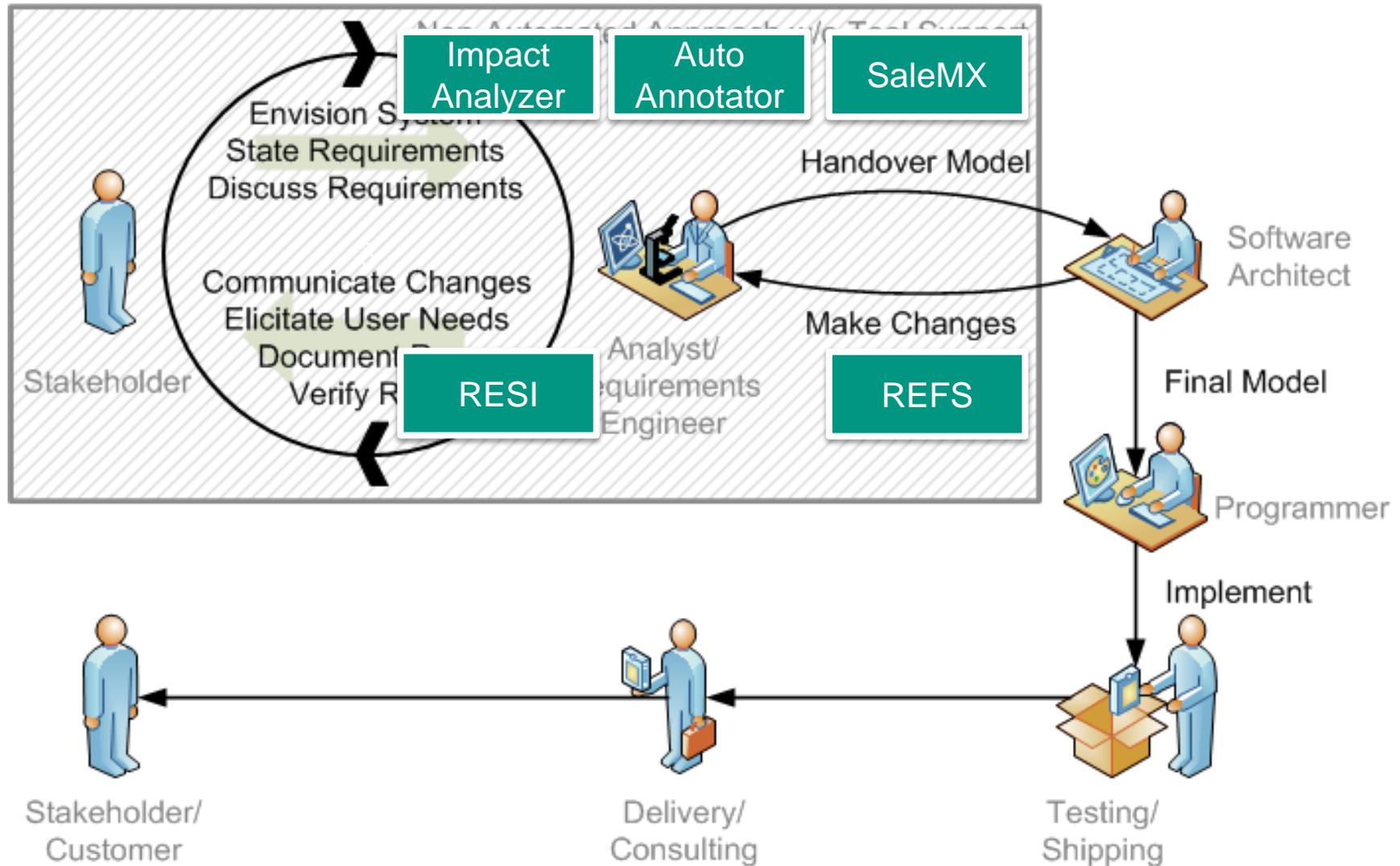
```
[ #A WHOIS_server|{AG,RECP} listens|STAT #on
TCP_port_43|PAT #for requests|HAB #from
WHOIS_clients|DON ]
[ #The WHOIS_client|{AG,DON} makes|ACT #a
text_request|{HAB} #to #the
@WHOIS_server|RECP ]|SUM
#then #the @WHOIS_server|AG replies|ACT
#with text_content|OPUS ]
[ #All @requests|PAT #are terminated|ACT #with
{ ASCII_CR AND #then ASCII_LF }|INST ]
```

Quelle: IETF WHOIS Protocol Specification, RFC 3912, Ch. 2, "Protocol Specification"

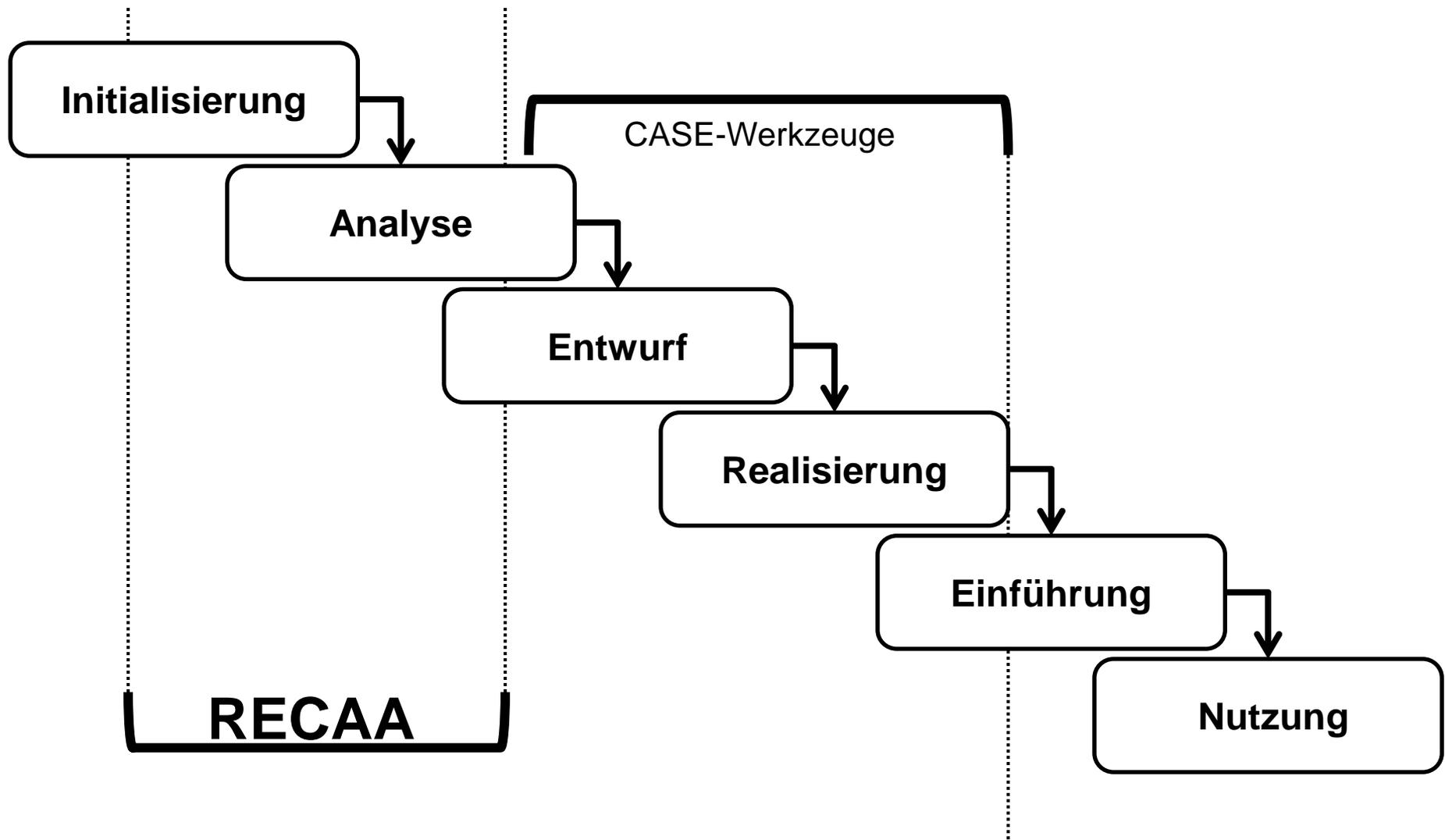
Arbeiten am Institut: Übersetzung in UML mit SAL_Emx



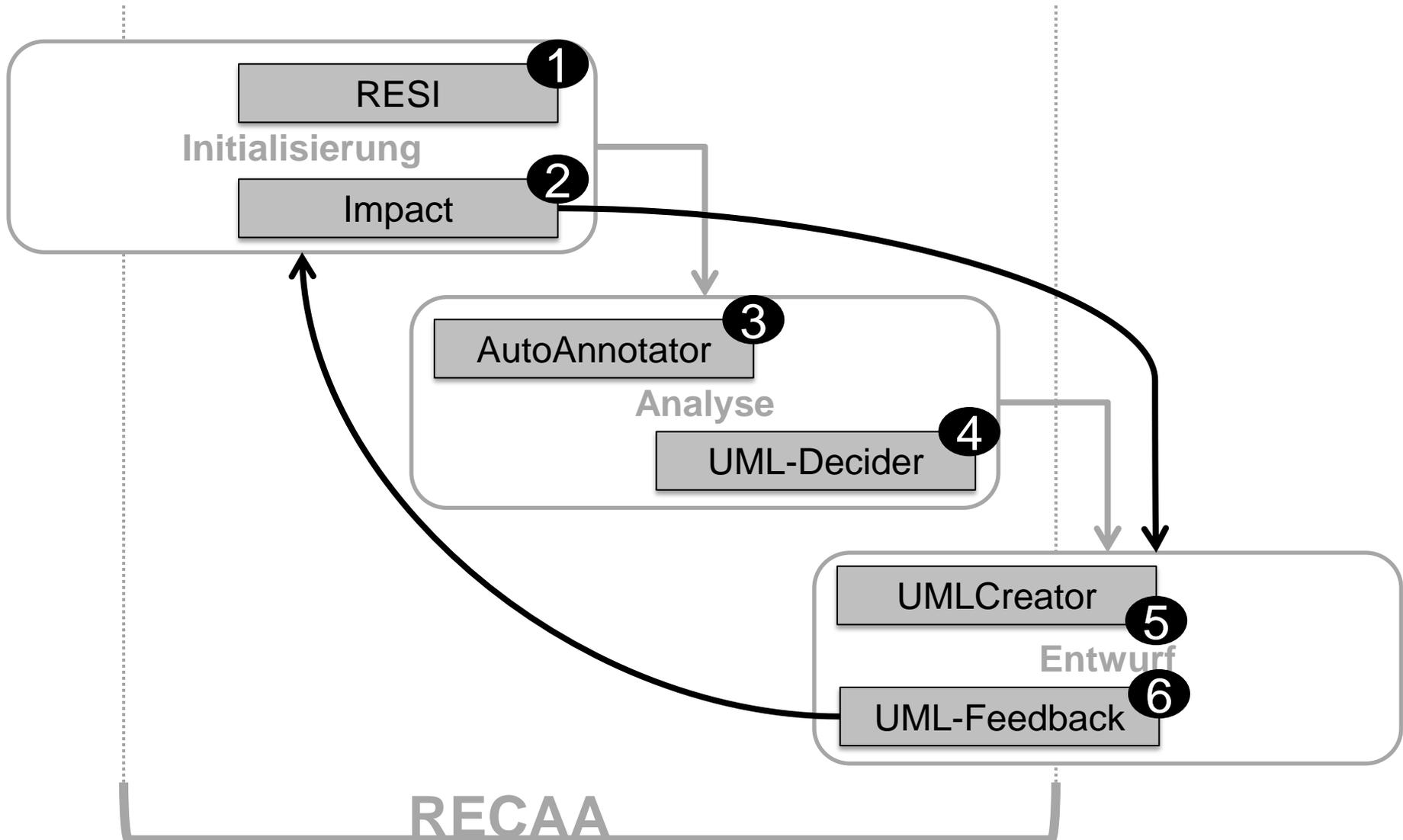
Natürliche Sprache als Hauptmittel der Anforderungsermittlung



Forschungsprojekt RECAA gliedert sich im Entwicklungszyklus ein



RECAA – eine Werkzeugensammlung für die Anforderungsermittlung und -Verwaltung



Falls es nicht funktioniert ...



...sollte es ein Kundenfehler sein

Literatur

- R. J. Abbott; *Program Design by Informal English Descriptions*; ACM; 1983
- T. Gelhausen, W. F. Tichy; *Thematic Role Based Generation of UML Models from Real World Requirements*; IEEE; 2007
- S. Körner, T. Gelhausen; *Improving Automatic Model Creation using Ontologies*; KSI; 2008
- S. Körner, T. Brumm; *RESI – A Natural Language Specification Improver*; ICSC; 2009
- S. Körner, Mathias Landhäußer; *Semantic Enriching of Natural Language Texts with Automatic Thematic Role Annotation*; NLDB 2010
- Siehe auch
 - RECAA + SAL_E**mx**: <https://svn.ipd.uni-karlsruhe.de/trac/mx>
 - AutoModel: <http://www.ipd.uka.de/Tichy/people.php?id=55>