

Semantische¹ Kodierung

Übersicht:

Spitäler müssen für die DRGs Diagnosen und Prozeduren kodieren.

Wieweit kann der Kodierprozess durch Computer unterstützt werden?

Wir untersuchen, was Kodierung und Semantik bedeutet.

Wir entdecken Charakteristika der semantischen Repräsentation.

Wir vergleichen 4 Methoden der Kodierunterstützung.

¹Semantik meint die Bedeutung von Zeichen, z.B. von Wörtern



Themenliste

1. Was bedeutet Kodierung?
 - Informationsreduktion
 - Zwei Arten von Kodierung
2. Semantik und Information
 - Bits und Bedeutung
 - Das semiotische Dreieck
3. Vom Text zum Kode: Probleme der automatisierten Kodierung
4. Begriffsarchitekturen:
Hierarchisch, multidimensional, multifokal
5. Vier Methoden der Kodierunterstützung:
Vom Stringvergleich zur Semantik
6. Semantische Interoperabilität
7. Live-Demo einer semantischen Kodierung



Was bedeutet Kodierung? - Informationsreduktion

Fragen, die Sie sich vor den nächsten Folien stellen können:

- Wie entsteht der Kode aus dem Fall?
- Welche Information ist im Kode enthalten?
- Welche Information ist nicht enthalten?
- Wie wird zwischen beidem ausgewählt?
- Kann der Kode den Fall vollständig abbilden?
- Gibt es ein perfektes Kodiersystem?
- Was hat das Kodierwerk mit dem Zweck der Kodierung zu tun?



Vom Patient zum Kode

Realität



Dossier/Freitext

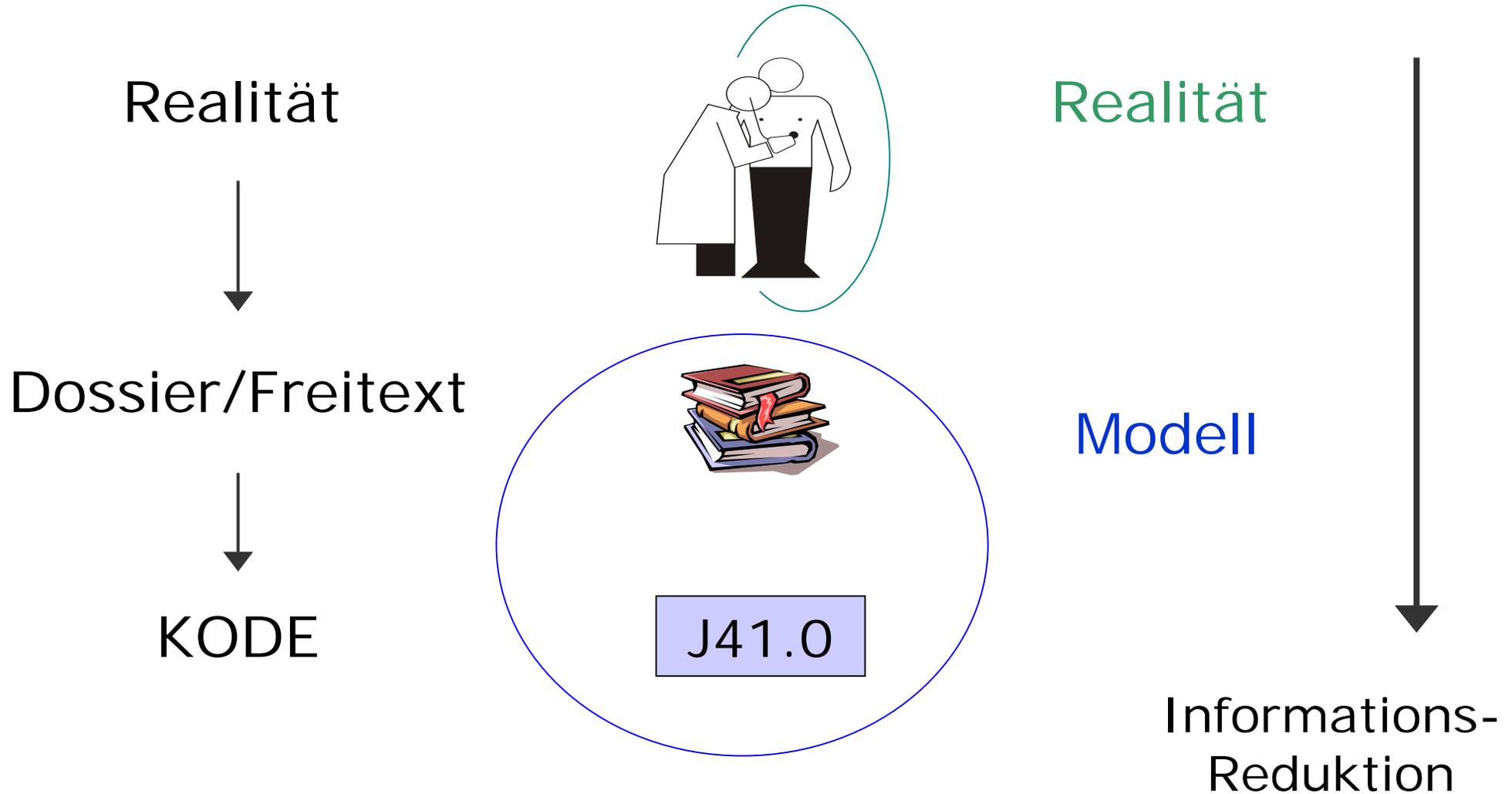


KODE

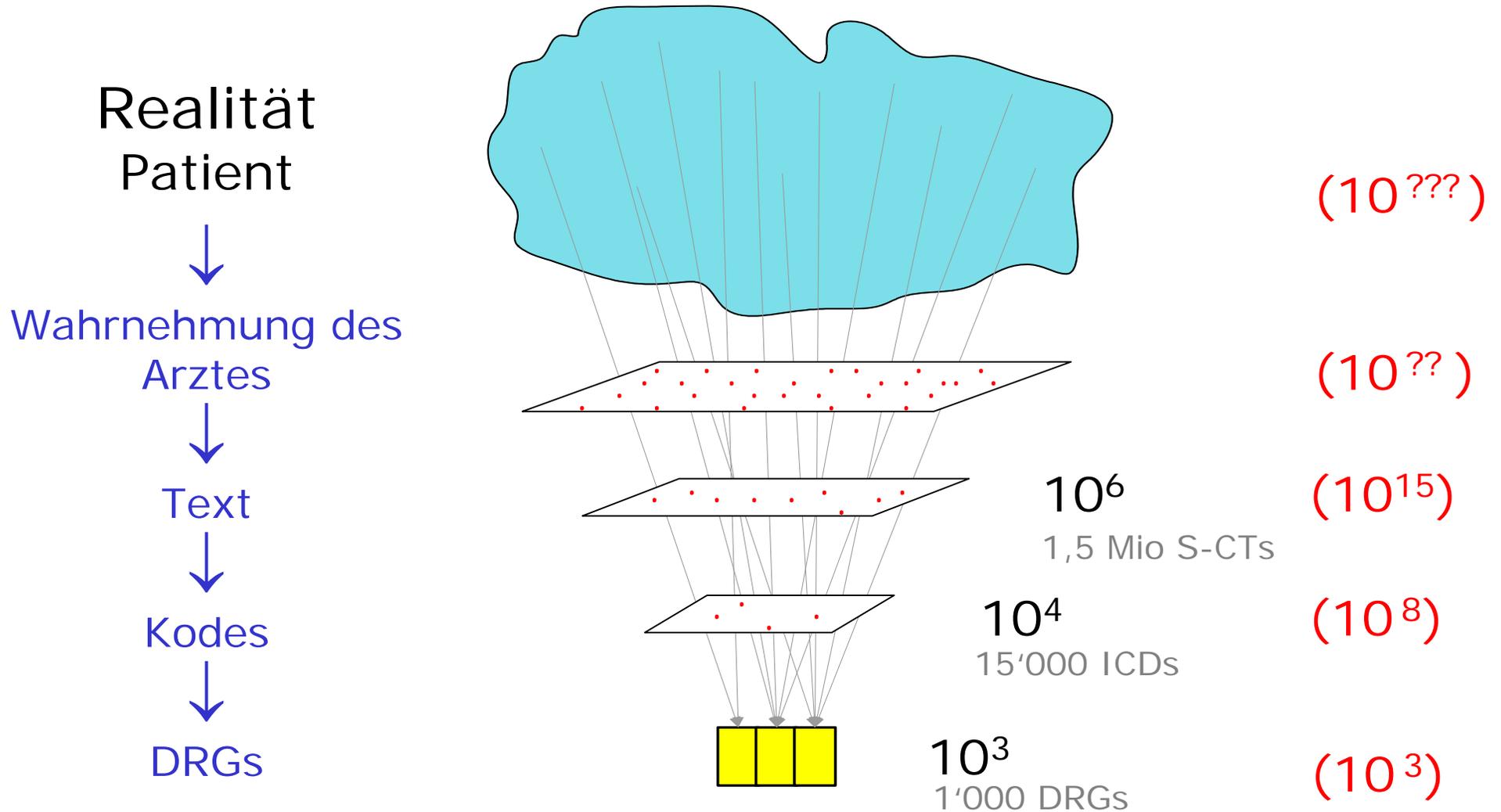
J41.0



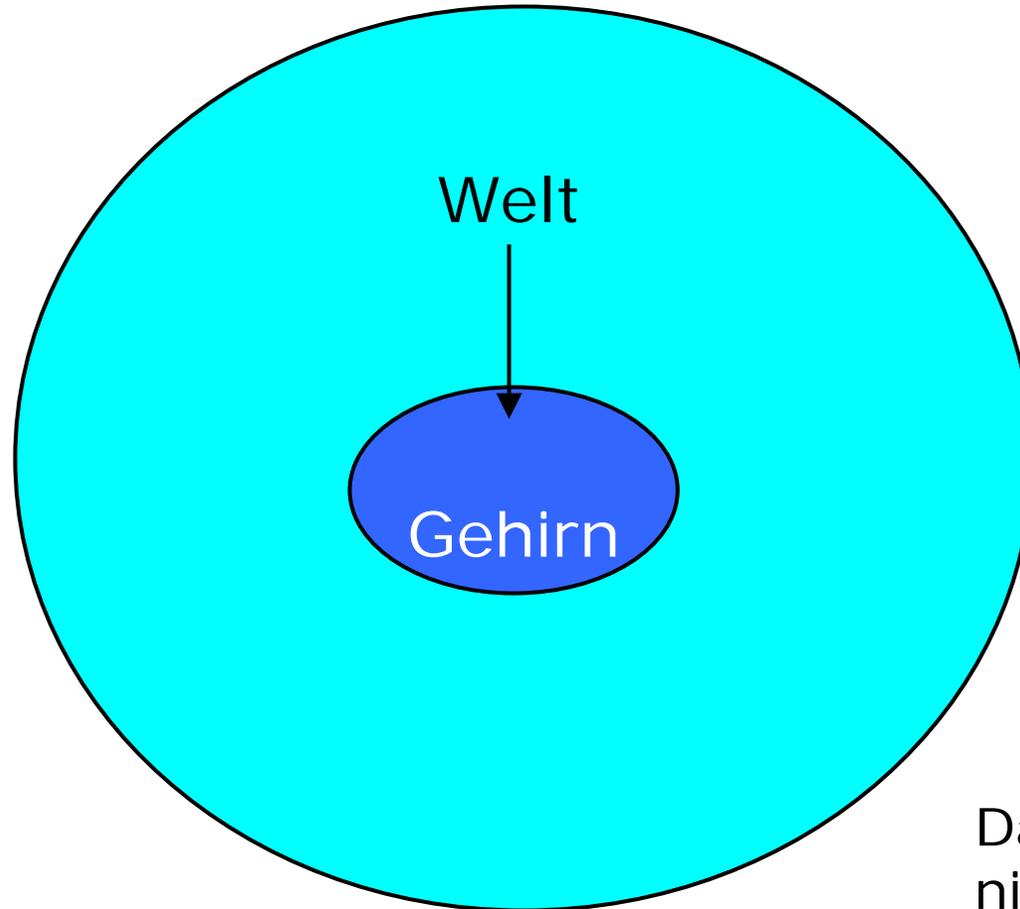
Vom Patient zum Kode



Quantifizierung der Informationsreduktion



Informationsfluss (Grundmodell)



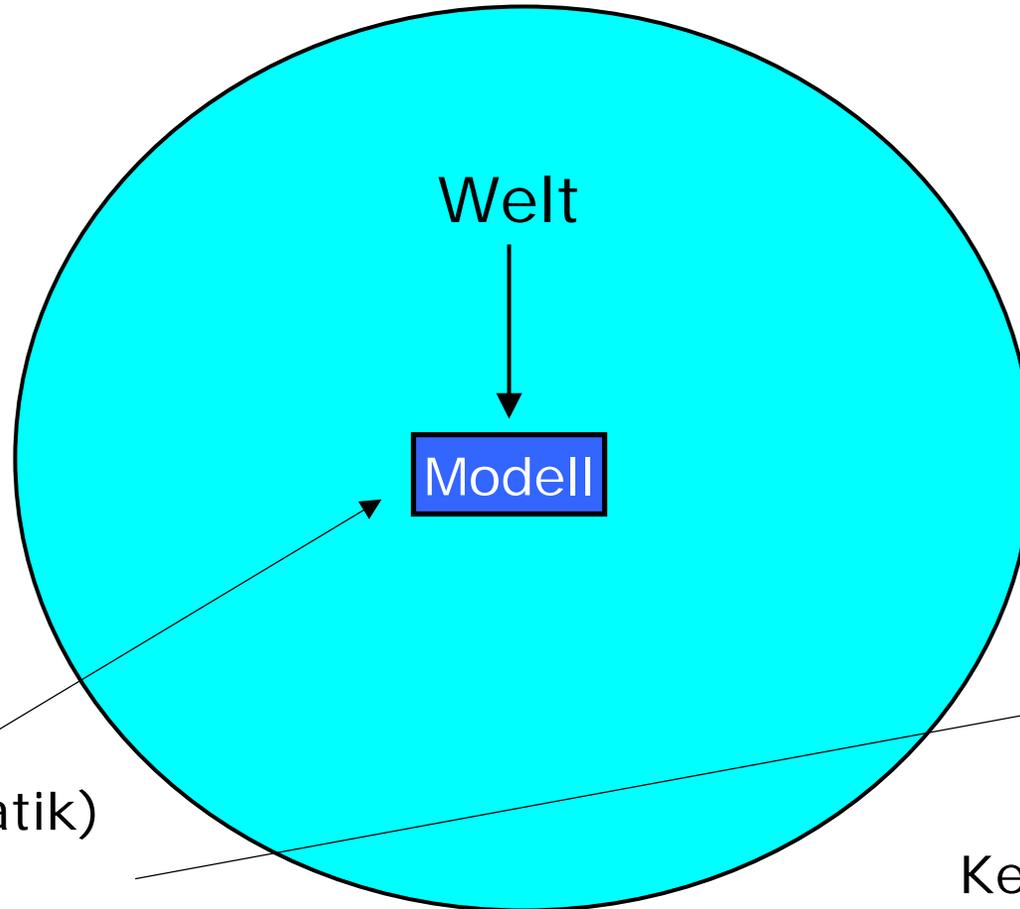
Ökonomie-
prinzip



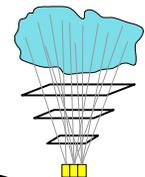
Das Gehirn kann
nicht die gesamte
Welt abbilden



Informationsfluss (Grundmodell)



Informations-
Reduktion ist
unumgänglich

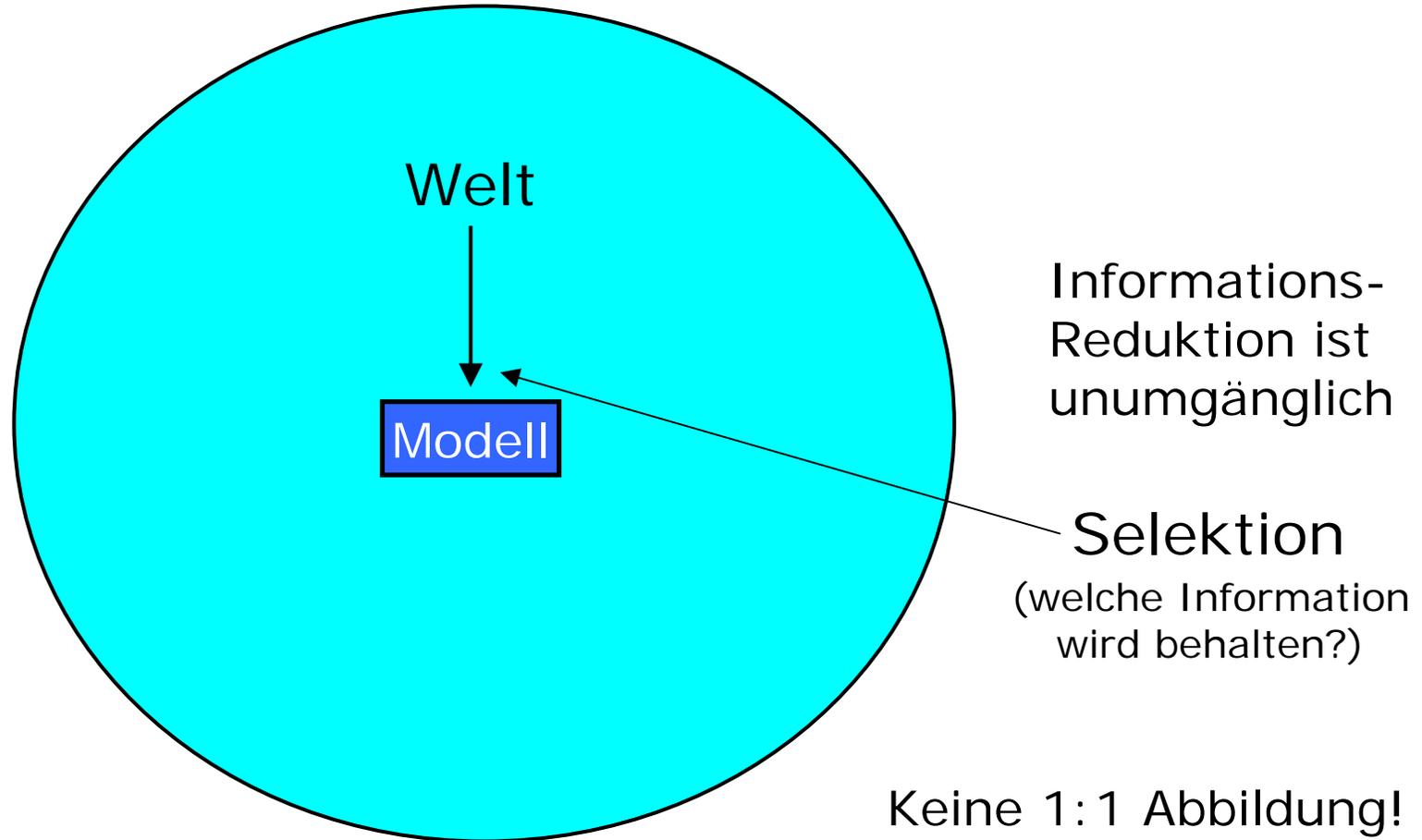


Keine 1:1 Abbildung!

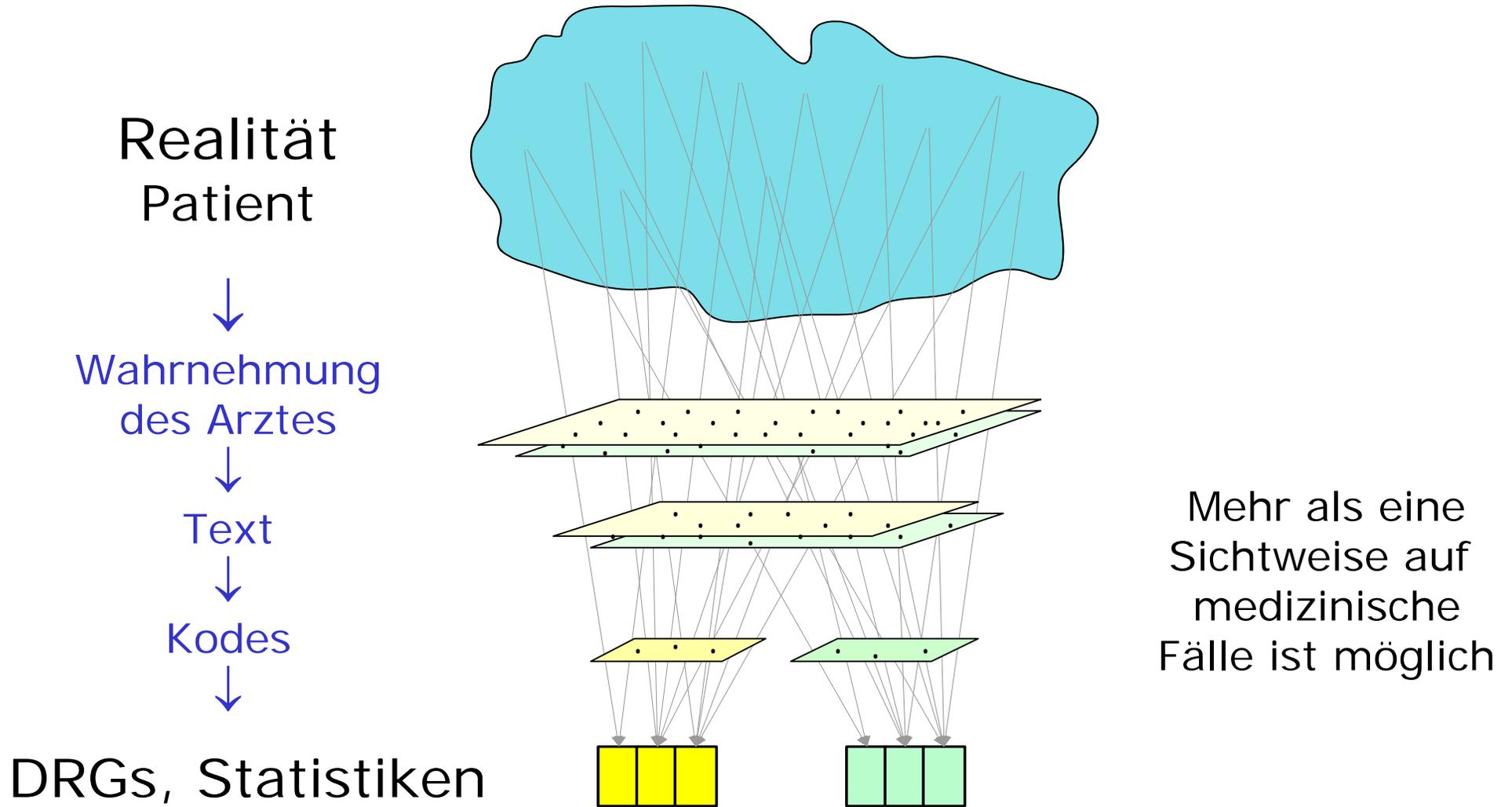
- Klassen (Informatik)
- Mengen (Math.)
- Begriffe (Semantik)



Informationsfluss (Grundmodell)



Die Bedeutung der Selektion



Realität und Modell

Bei der Erstellung eines Modells können wir nicht von einer 1:1-Abbildung der Realität ausgehen



Realität und Modell

Information: Selektion
entscheidet



Realität und Modell

Es gibt > 1 Modell

Es gibt > 1 Terminologie

Es gibt > 1 Klassifikation



Themenliste

1. Was bedeutet Kodierung?
 - Informationsreduktion
 - **Zwei Arten von Kodierung**
2. Semantik und Information
 - Bits und Bedeutung
 - Das semiotische Dreieck
3. Vom Text zum Kode: Probleme der automatisierten Kodierung
4. Begriffsarchitekturen:
Hierarchisch, multidimensional, multifokal
5. Vier Methoden der Kodierunterstützung:
Vom Stringvergleich zur Semantik
6. Semantische Interoperabilität
7. Live-Demo einer semantischen Kodierung

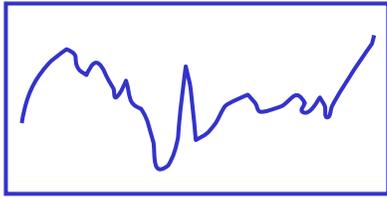


Was bedeutet Kodierung? - 2 Arten von Kodierung

Fragen, die Sie sich vor den nächsten Folien stellen können:

- Welche Kodierungen gibt es neben der medizinischen Kodierung?
- Was ist der Unterschied zwischen einem Morsekode und der ICD-10?
- Was bedeutet das "C" in ICD-10?
- Sind Kodierungen reversibel?
- Was hat die Informationsreduktion des Vorkapitels damit zu tun?



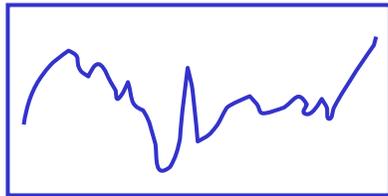


Tonaufnahme



01100110100110

Musik-CD
(dig. **Kode**)



Wiedergabe

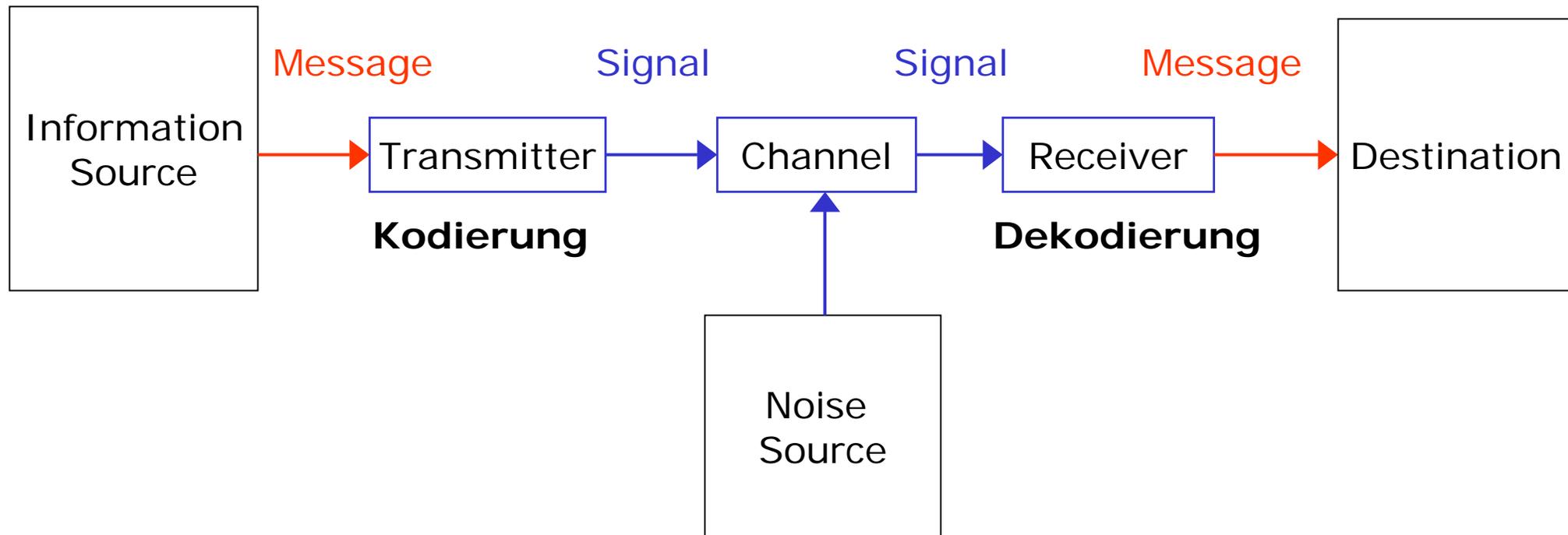
Kode = Umformung

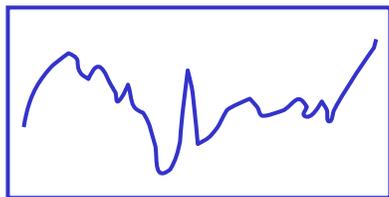
Informationsverlust minimieren!



„General Communication System“

nach C.E. Shannon (1949)





Tonaufnahme



01100110100110

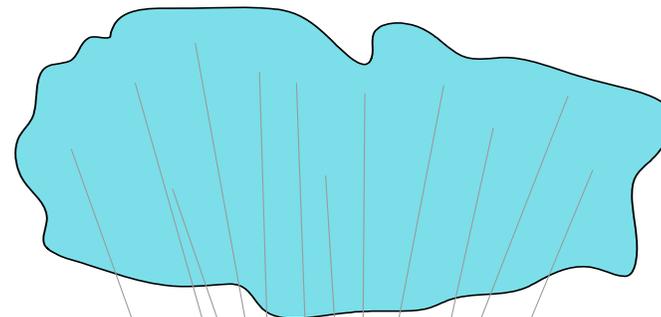
Musik-CD
(dig. **Kode**)



Wiedergabe

Kode = Umformung

Informationsverlust minimieren!



Realer
Patient



Kode des Falles
Klassifikationen
Begriffe

Kode = „Schublade“

Informationsverlust ist Ziel!



2 Arten von Kodierung

Umformung

Telefon, Morsealphabet,
Musik-CD, Kryptologie

→ Informationsverlust vermeiden!

Schubladen-
bildung

Klassierung (ICD), DRGs
Wahrnehmung, Interpretation

→ Informationsverlust unvermeidlich!



Themenliste

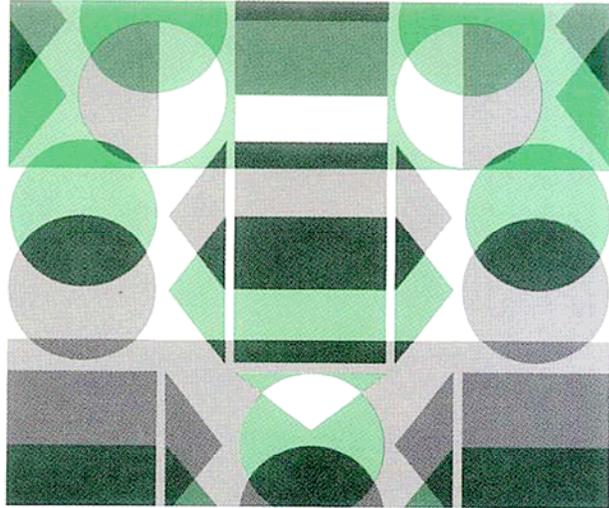
1. Was bedeutet Kodierung?
 - Informationsreduktion
 - Zwei Arten von Kodierung
2. Semantik und Information
 - Bits und Bedeutung
 - Das semiotische Dreieck
3. Vom Text zum Kode: Probleme der automatisierten Kodierung
4. Begriffsarchitekturen:
Hierarchisch, multidimensional, multifokal
5. Vier Methoden der Kodierunterstützung:
Vom Stringvergleich zur Semantik
6. Semantische Interoperabilität
7. Live-Demo einer semantischen Kodierung



Fragen, die Sie sich vor den nächsten Folien stellen können:

- Bits sind die Grundbausteine der Information? Was heisst das?
- Welche Information (Bedeutung) trägt ein Bit?
- Kann Bedeutung überhaupt mit Bits ausgedrückt werden?
- Mit anderen Worten: Was haben Bits mit Semantik zu tun?





THE MATHEMATICAL THEORY OF COMMUNICATION

by Claude E. Shannon and Warren Weaver

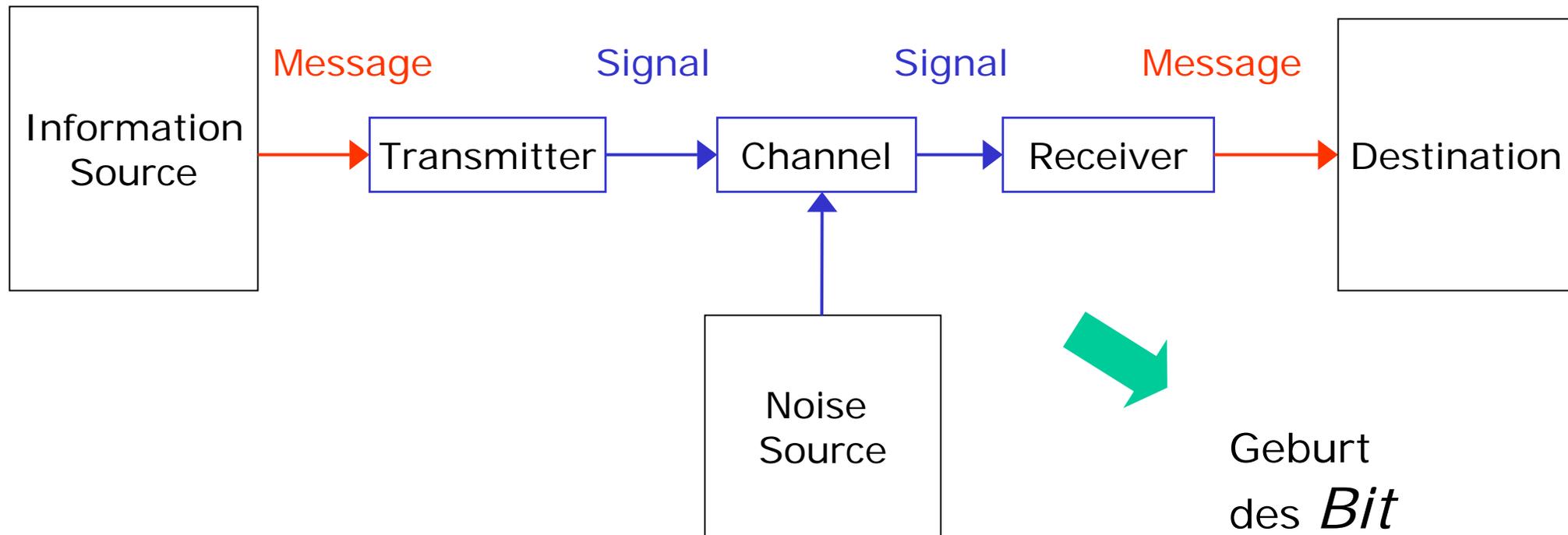
C.E. Shannon, 1949:

Startschuss für die
Informationstheorie



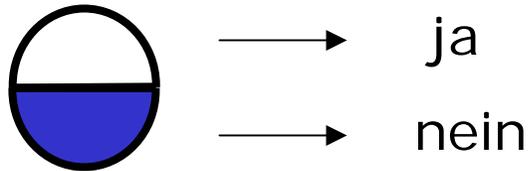
„General Communication System“

nach C.E. Shannon (1949)



Bit und Bedeutung

Was bedeutet ein Bit?

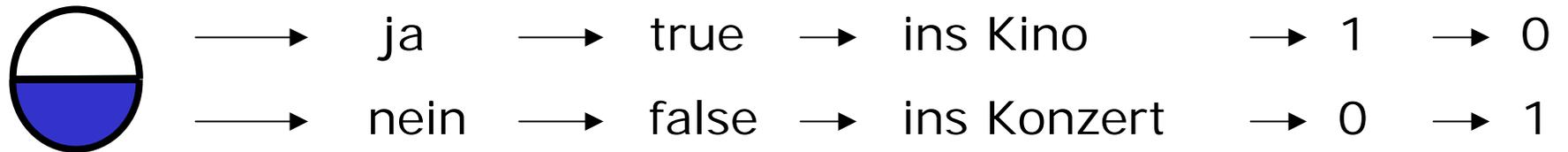


Was bedeutet ja?
Was bedeutet nein?



Bit und Bedeutung

Was bedeutet ein Bit?



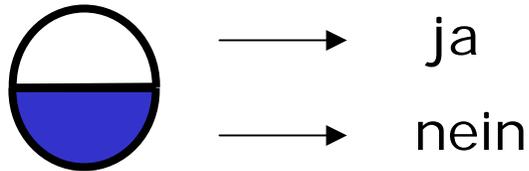
Ja/nein bedeuten
nichts ohne die
zugehörige Frage



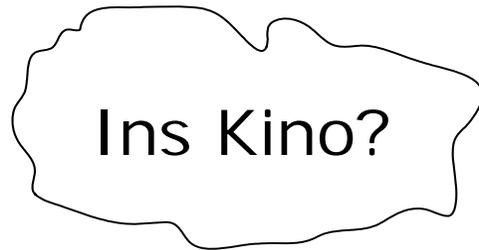
Was bedeutet ja?
Was bedeutet nein?

Bit und Bedeutung

Was bedeutet ein Bit?



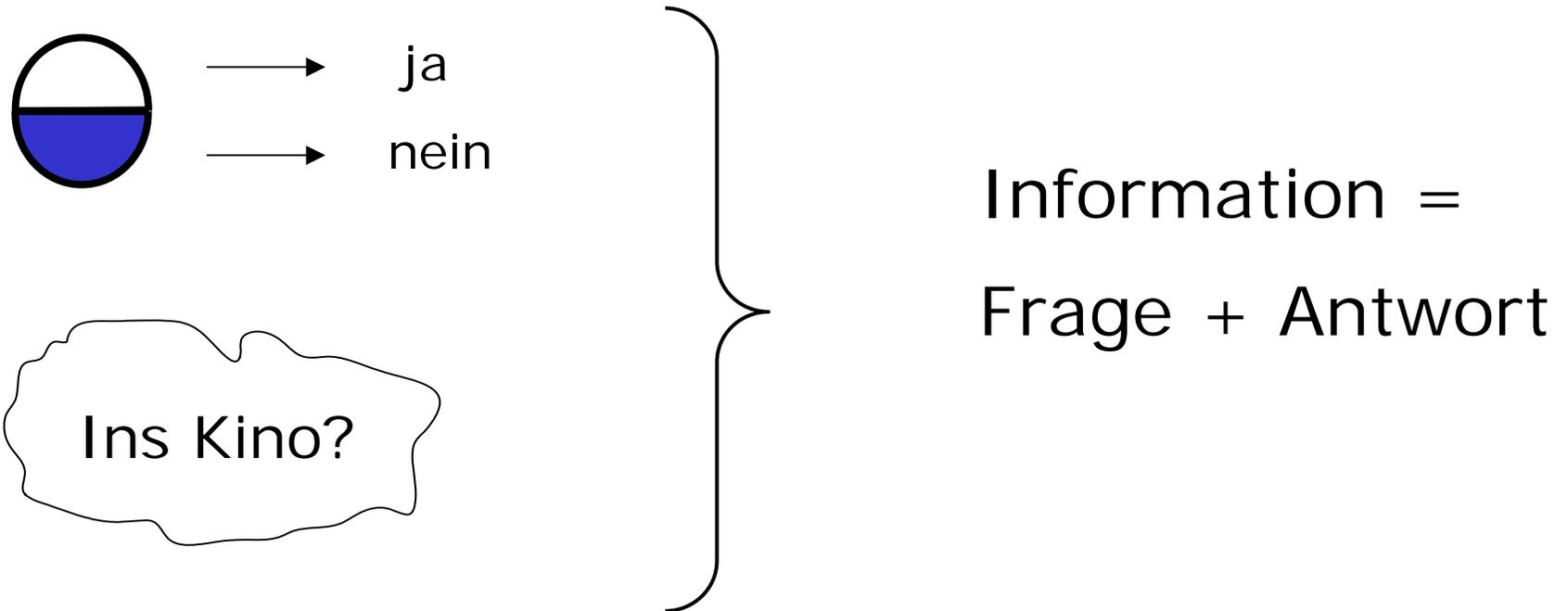
Bit = Antwort



Die Frage bestimmt die **Semantik** des Bits

Bit und Bedeutung

Was bedeutet ein Bit?



Bit und Bedeutung

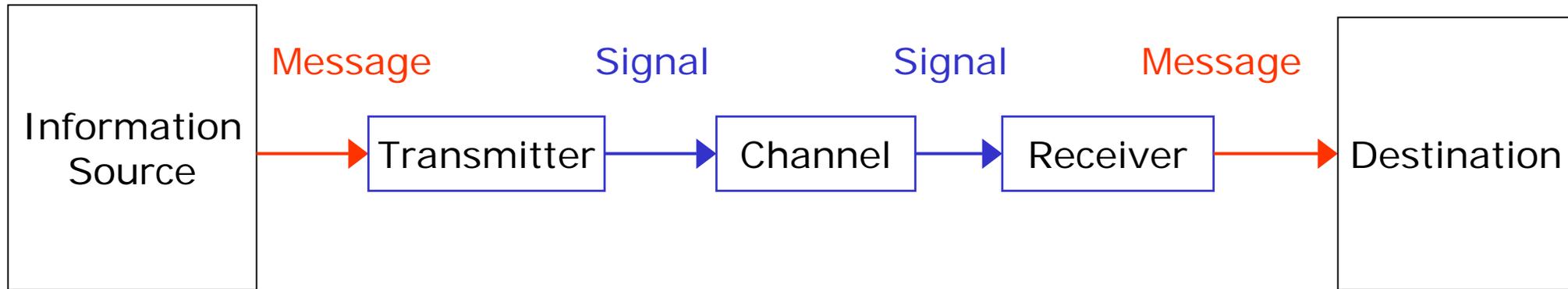
Was bedeutet ein Bit?

	Bit	Bedeutung
Sichtbarkeit	sichtbar	unsichtbar
Ausdehnung	Punkt	Feld (Kontext)
Fassbarkeit	objektiv	subjektiv
Rolle, Funktion	Signal, Zeichen	Semantik



Bit und Bedeutung

Was bedeutet ein Bit?



„Frequently the messages have *meanings* ...
These *semantic* aspects of communication
are *irrelevant* to the engineering problem.“

(C.E. Shannon: Mathematical Theory of Communication)



Themenliste

1. Was bedeutet Kodierung?
 - Informationsreduktion
 - Zwei Arten von Kodierung
2. Semantik und Information
 - Bits und Bedeutung
 - [Das semiotische Dreieck](#)
3. Vom Text zum Kode: Probleme der automatisierten Kodierung
4. Begriffsarchitekturen:
Hierarchisch, multidimensional, multifokal
5. Vier Methoden der Kodierunterstützung:
Vom Stringvergleich zur Semantik
6. Semantische Interoperabilität
7. Live-Demo einer semantischen Kodierung

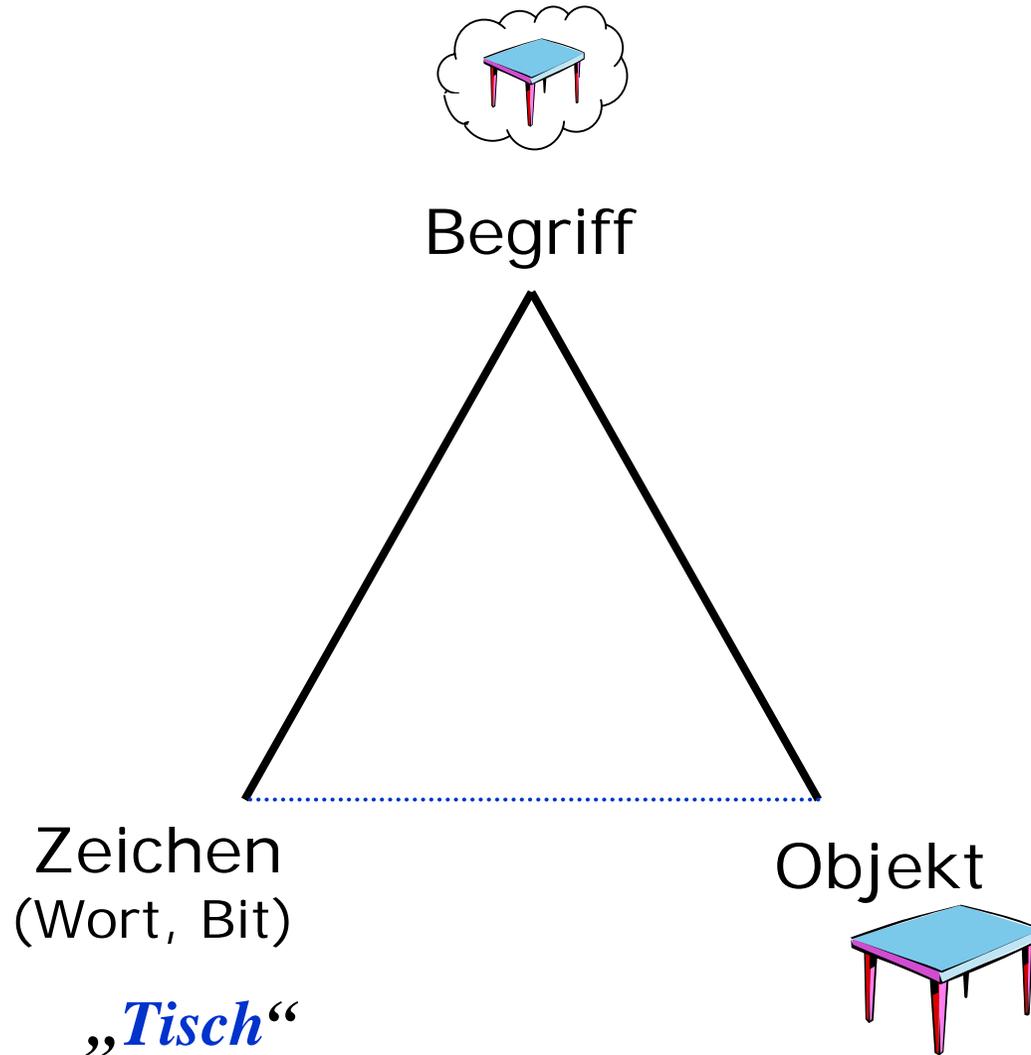


Fragen, die Sie sich vor den nächsten Folien stellen können:

- Wie ist ein Wort mit einem Gegenstand verbunden?
- Hat jedes Wort genau eine Bedeutung?
- Hat jede Bedeutung genau ein Wort?
- Lässt sich die Bedeutung eines Wortes lokalisieren?
- Warum überlappen sich Bedeutungen?
- In einem Satz folgt ein Wort auf das andere. Sind Begriffe (d.h. Bedeutungen von Wörtern) auch so angeordnet?
- Falls nein, wie sonst?



Semiotisches Dreieck



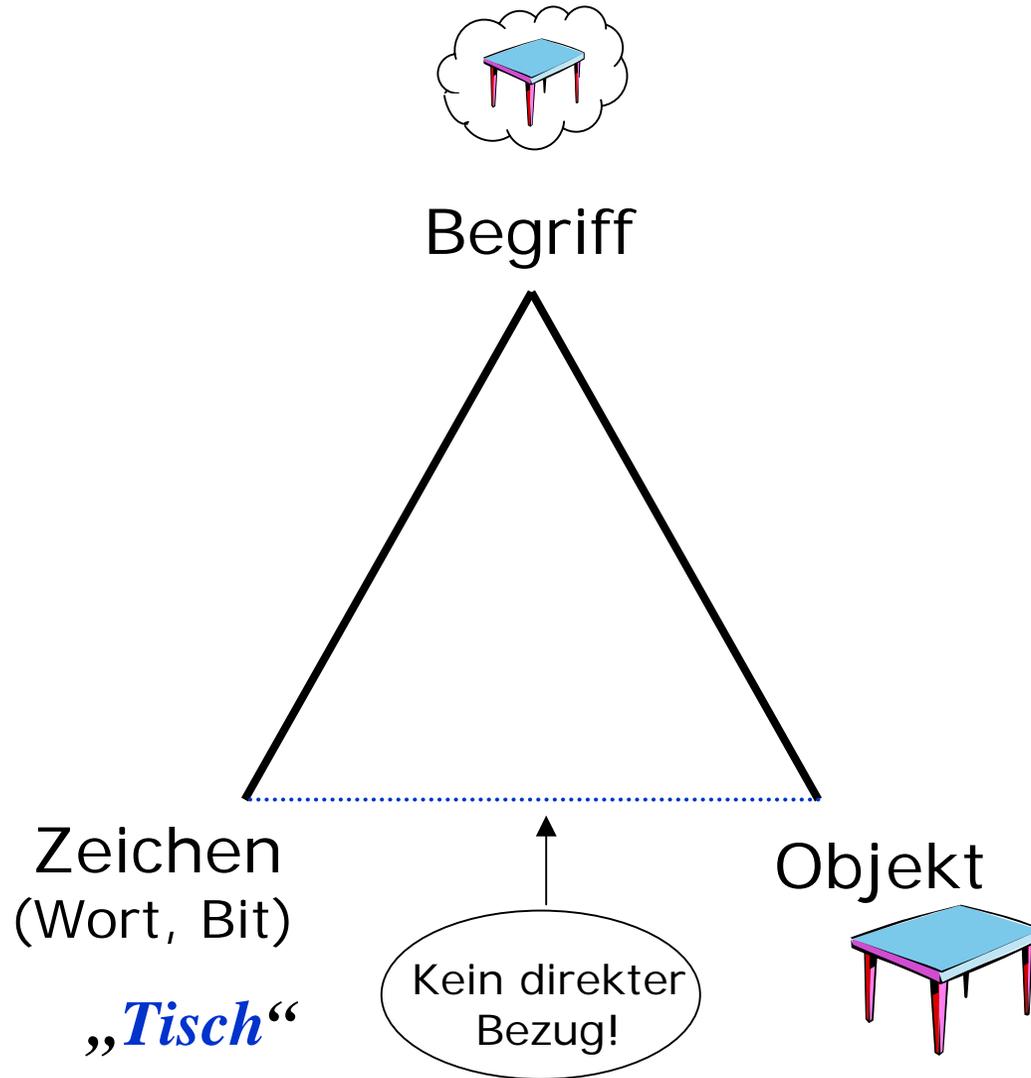
Autoren:

J. F. Sowa
Ogden / Richards
Charles S. Peirce
F. de Saussure
Aristoteles

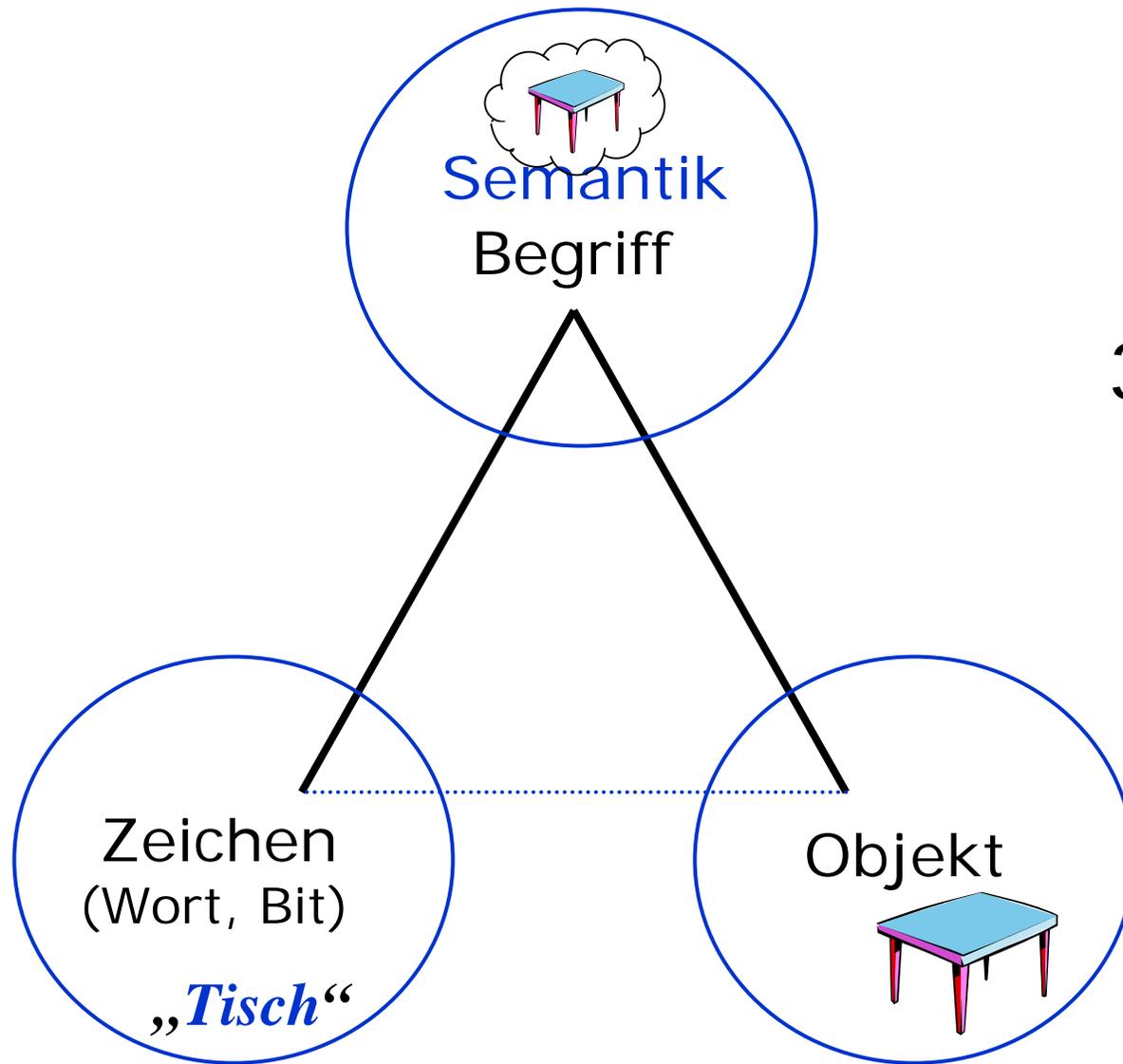
...



Semiotisches Dreieck



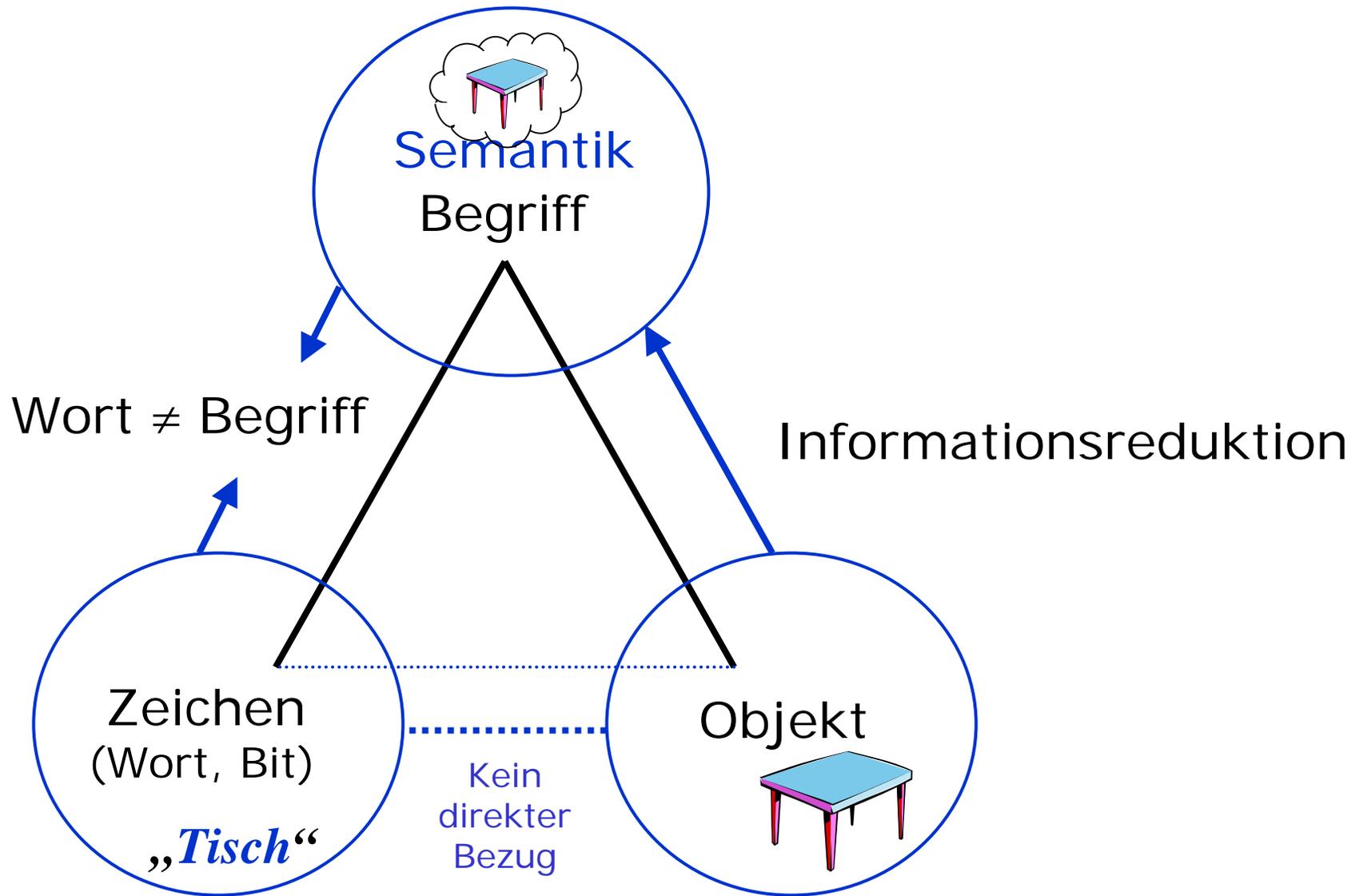
Semiotisches Dreieck



3 Welten



Semiotisches Dreieck



Wort \neq Begriff

	Wort	Begriff
Sichtbarkeit	sichtbar	unsichtbar
Abgrenzbarkeit	disjunkt	überlappend
Anordnung	linear	multidimensional

Überlappung und Mehrdimensionalität der Begriffe:

→ Begriffsmoleküle



Wort \neq Begriff

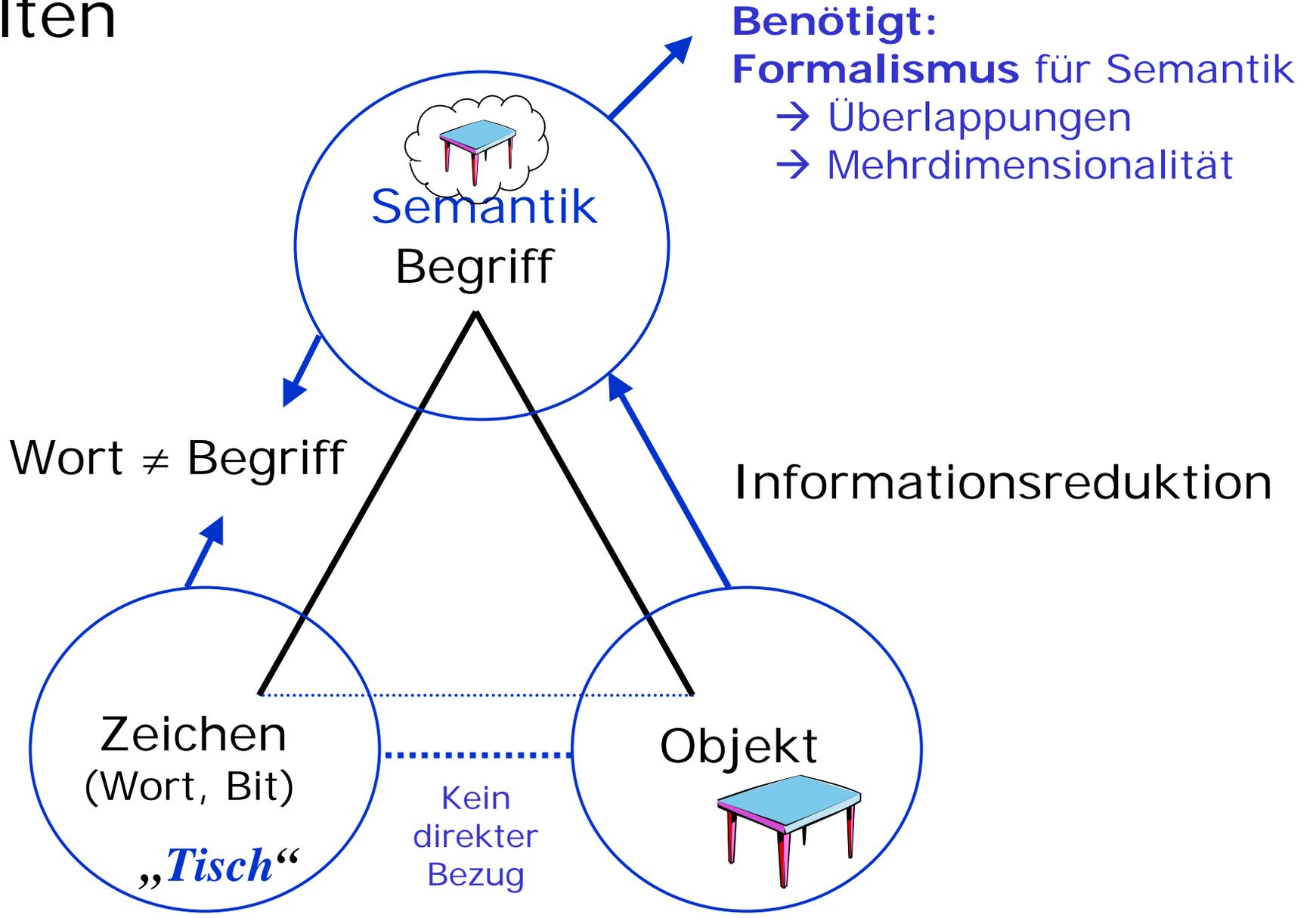
Semantische Technologien
dürfen sich nicht auf Wörter
abstützen

Überlappung und Mehrdimensionalität der Begriffe:

→ Begriffsmoleküle



3 Welten



Themenliste

1. Was bedeutet Kodierung?
 - Informationsreduktion
 - Zwei Arten von Kodierung
2. Semantik und Information
 - Bits und Bedeutung
 - Das semiotische Dreieck
3. Vom Text zum Kode: Probleme der automatisierten Kodierung
4. Begriffsarchitekturen:
 - Hierarchisch, multidimensional, multifokal
5. Vier Methoden der Kodierunterstützung:
 - Vom Stringvergleich zur Semantik
6. Semantische Interoperabilität
7. Live-Demo einer semantischen Kodierung



Automatisch vom Text zum Kode: Probleme

Fragen, die Sie sich vor den nächsten Folien stellen können:

- Wie lang wäre eine Tabelle mit allen Diagnosen?
- Wieviele Formulierungen gibt es für eine Diagnose?
- Wie interpretiert ein Computer Wörter mit mehreren Bedeutungen?
- Wie findet der Computer implizite Bedeutungen von Wörtern?
- Steht ein Kode für eine oder für mehrere (verschiedene) Diagnosen?
- Hat umgekehrt eine Diagnose eventuell mehrere Kodes?
- Lassen sich alle Diagnosen in Baumform (=Hierarchie) anordnen?
- Wie werden kombinierte Diagnosen kodiert? Einmal? Mehrfach?



1. Problem: Kombinatorische Explosion



Risswunde ...
Messerstich ...
Bisswunde ...
Tiefer Riss ...
Klaffende Verletzung ...
Platzwunde ...
Schnittwunde ...
Stichwunde ...
Offene Verletzung ...

... Knie
... parapatellar
... Fossa poplitea
... neben der Kniescheibe

S81.0 Offene Wunde des Knies



1. Problem: Kombinatorische Explosion

× 50 Formulierungen für offene Wunde
× 10 Formulierungen für Knie
× 10 x 20 x 20 Ausschmückungen
= 2'000'000 Formulierungen für S81.0

× 13'000 Codes

= 26 Mrd. Formulierungen → Aufzählung möglich?
↘ Wartung möglich?

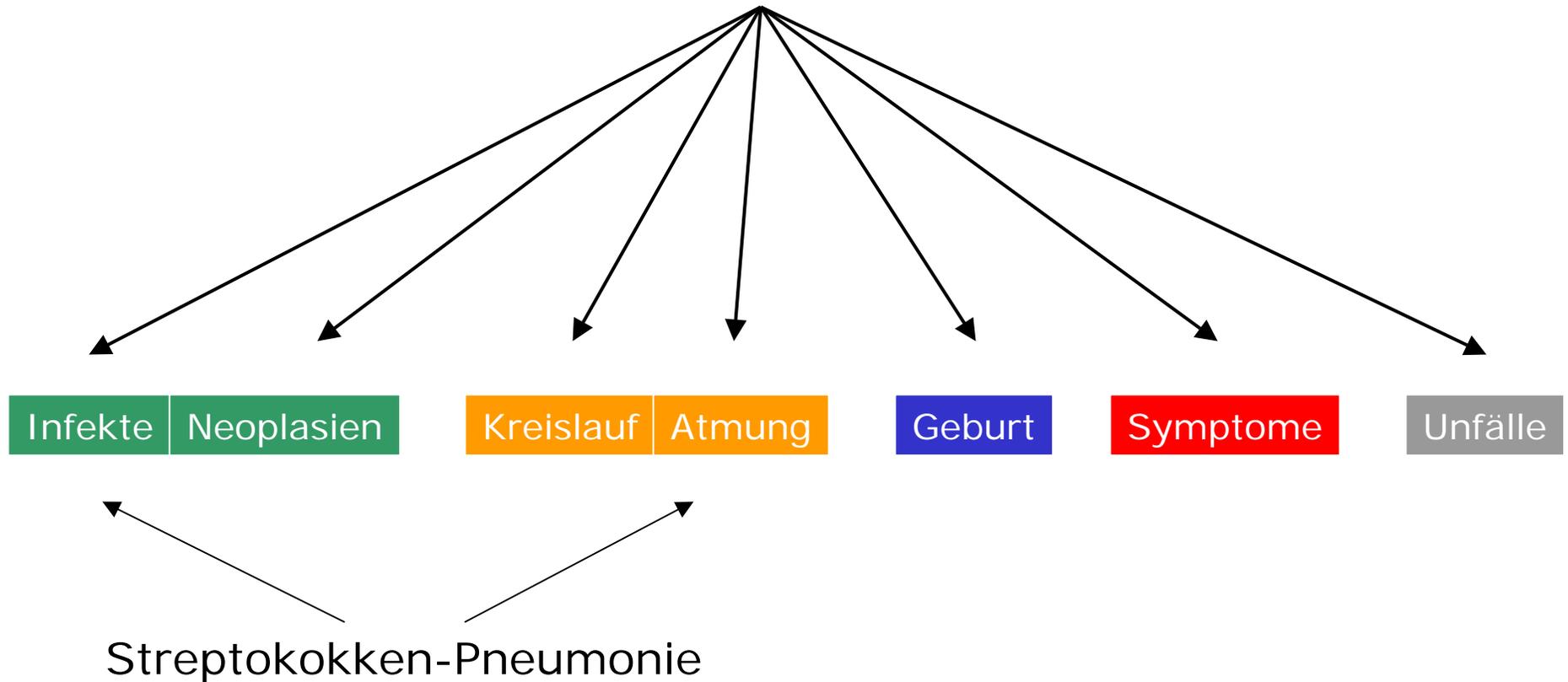


2. Problem: Komplexität



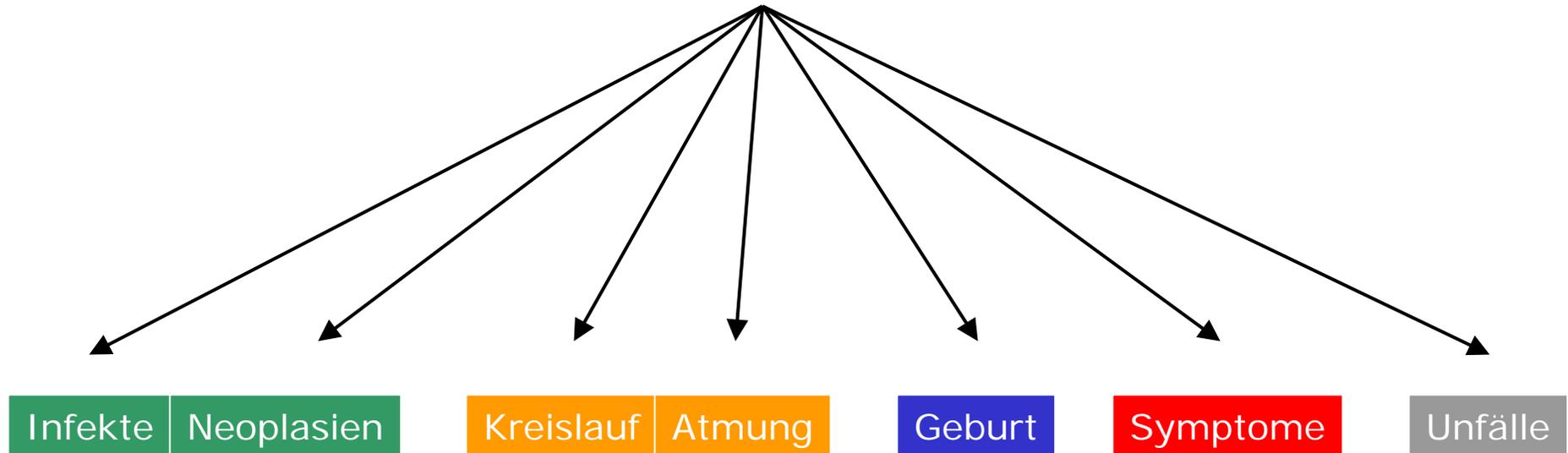
2. Problem: Komplexität

Kapitel der ICD-10



2. Problem: Komplexität

Kapitel der ICD-10

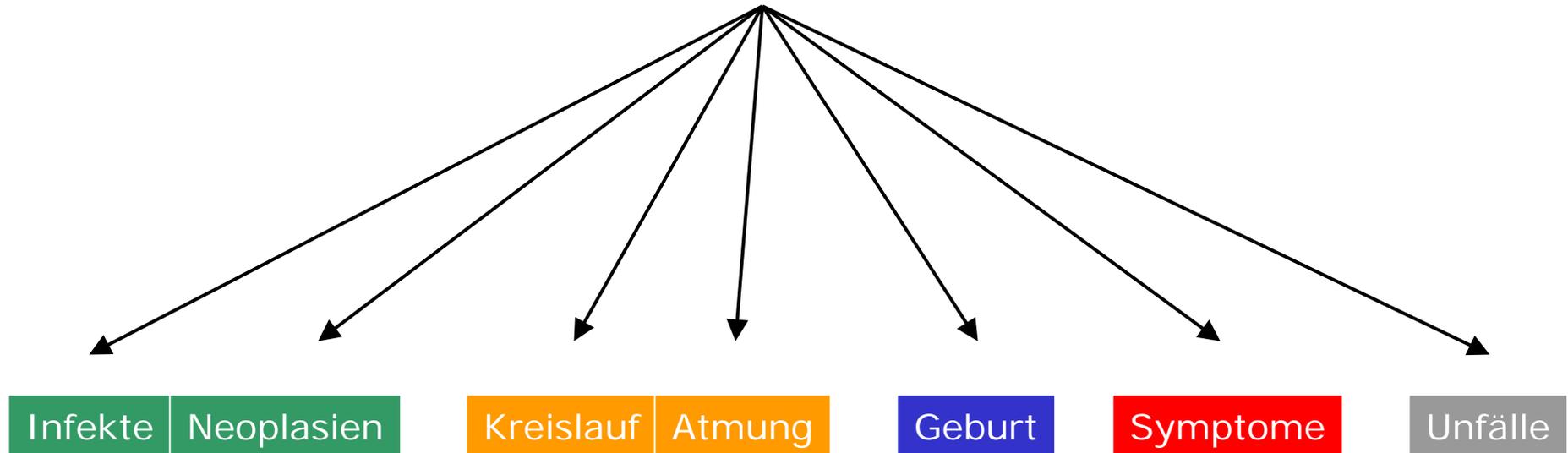


Streptokokken-Pneumonie

In der Realität **überschneiden** sich die hierarchischen Kategorien

2. Problem: Komplexität

Kapitel der ICD-10



Streptokokken-Pneumonie

→ komplexe Inklusiva und Exklusiva
(für Maschine und Mensch)



2. Problem: Komplexität

- Überschneidungen
- Mehrdeutigkeiten (Kopf, Blase)
- Negation (Diabetes, nicht insulinpflichtig)
- Nicht-Information (Diabetes, n.n.b.)
- Implikationen (Radius → Knochen, Unterarm)
- Auslassungen (Fraktur von Humerus und Radius)
- Mehrfachdiagnosen mit gegenseitigem Einfluss:
 Kreuz-Stern, komplexe Inklusiva und Exklusiva
- Summe-Summanden-Problem!



Themenliste

1. Was bedeutet Kodierung?
 - Informationsreduktion
 - Zwei Arten von Kodierung
2. Semantik und Information
 - Bits und Bedeutung
 - Das semiotische Dreieck
3. Vom Text zum Kode: Probleme der automatisierten Kodierung
4. **Begriffsarchitekturen:**
 - Hierarchisch, multidimensional, multifokal
5. Vier Methoden der Kodierunterstützung:
 - Vom Stringvergleich zur Semantik
6. Semantische Interoperabilität
7. Live-Demo einer semantischen Kodierung



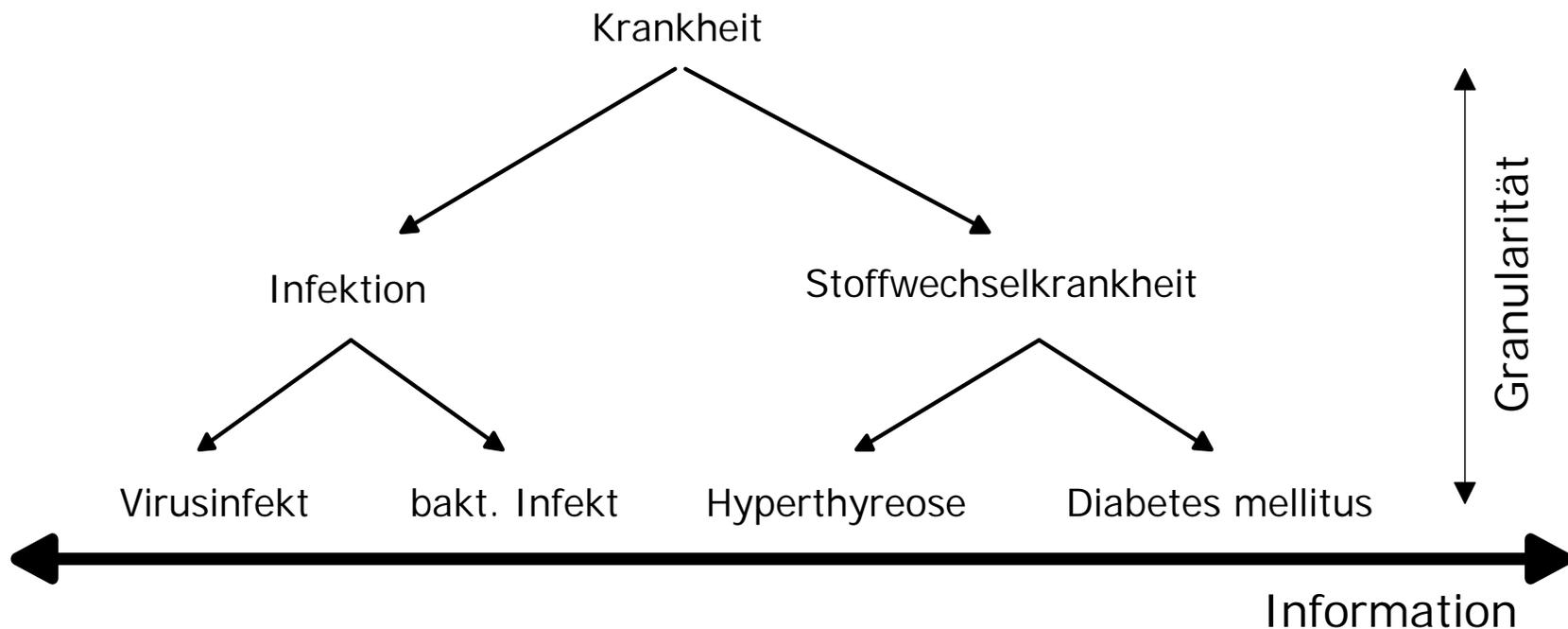
Begriffsarchitekturen (Systematiken)

Fragen, die Sie sich vor den nächsten Folien stellen können:

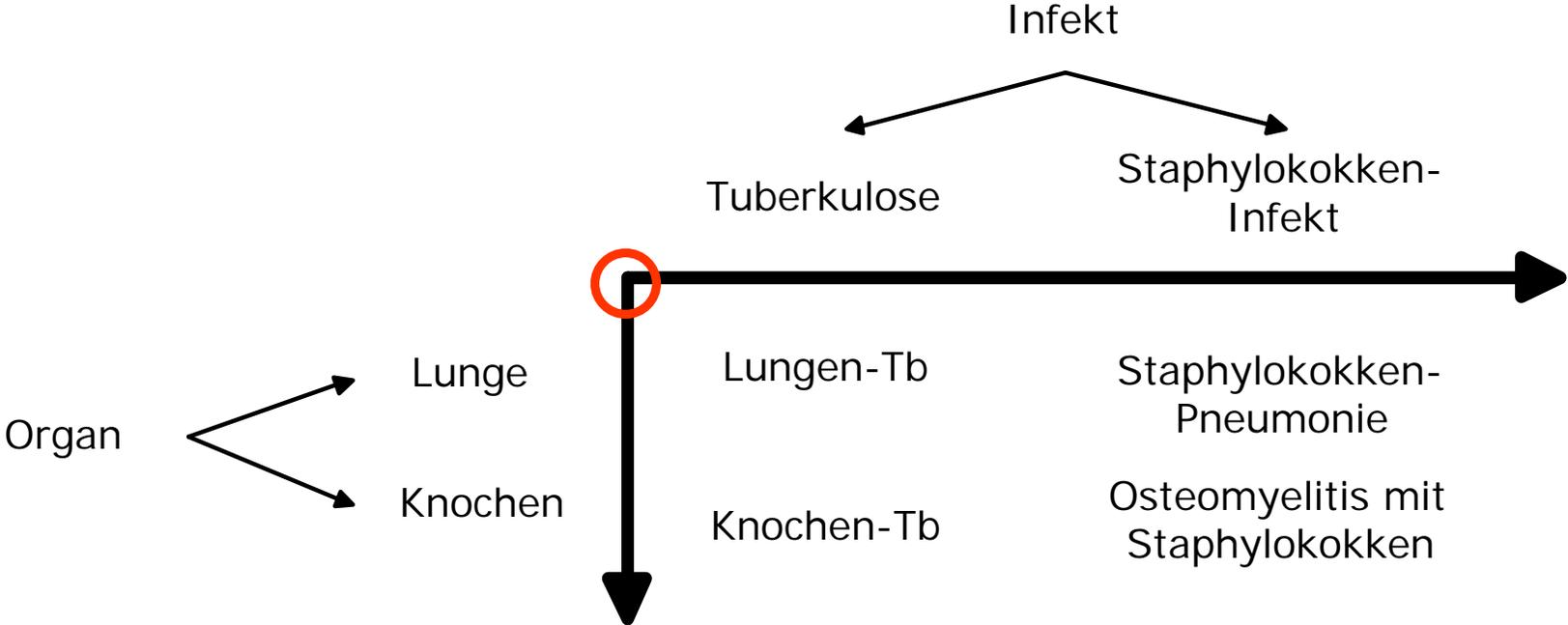
- Welche Beziehungen (Relationen) sind zwischen zwei Begriffen möglich?
- Lassen sich alle Diagnosen in einem Baum anordnen?
- Wenn sich eine Gruppe von Diagnosen in zwei oder mehr Merkmalen unterscheidet (z.B. Infektionen bezüglich Erreger und Organ), was bedeutet das für den Hierarchiebaum?
- Wieviele solche Merkmale (Organe, Erreger, usw.) gibt es?
- Macht jedes Merkmal für alle Diagnosen Sinn?



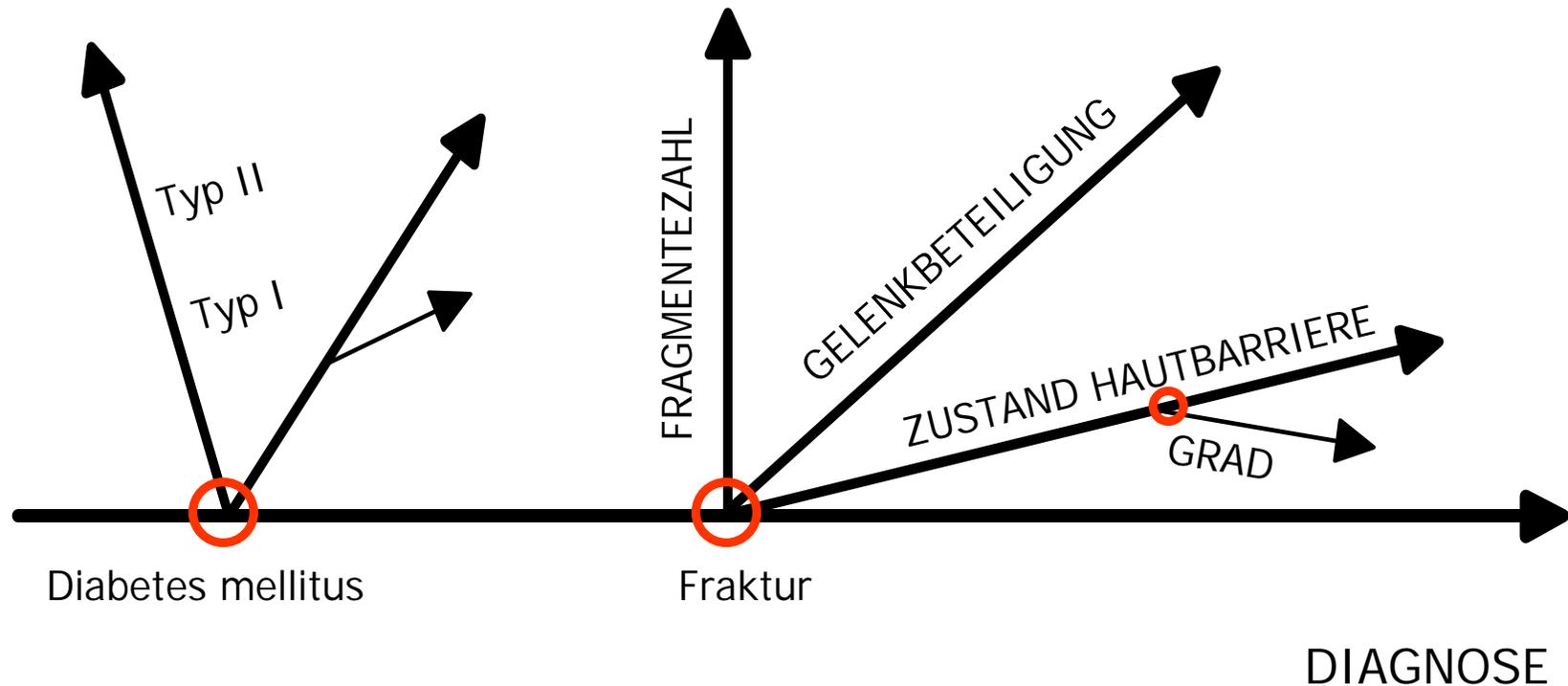
Die Hierarchie: 1 Achse = 1 Dimension



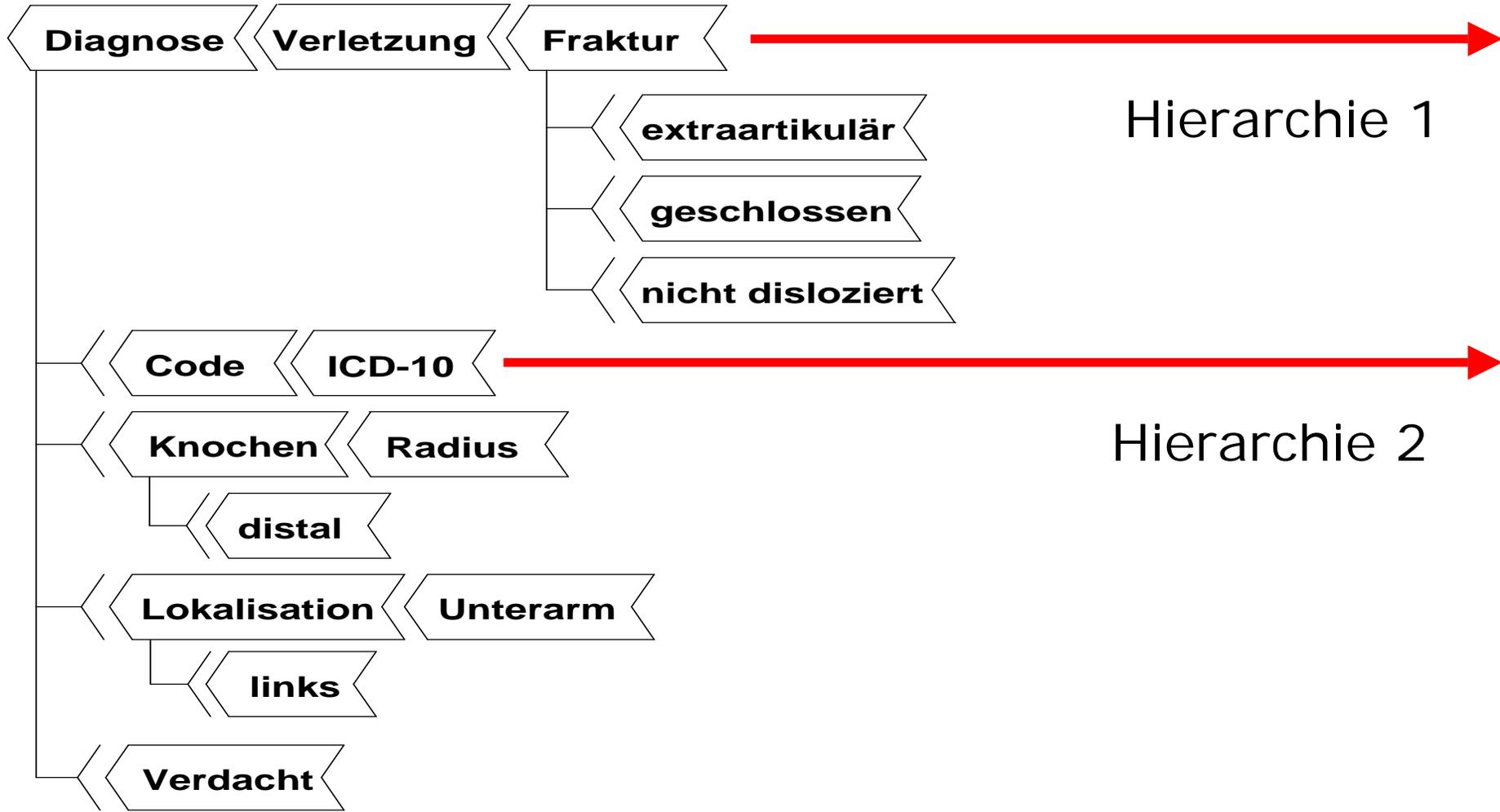
Die semantische Matrix: multidimensional



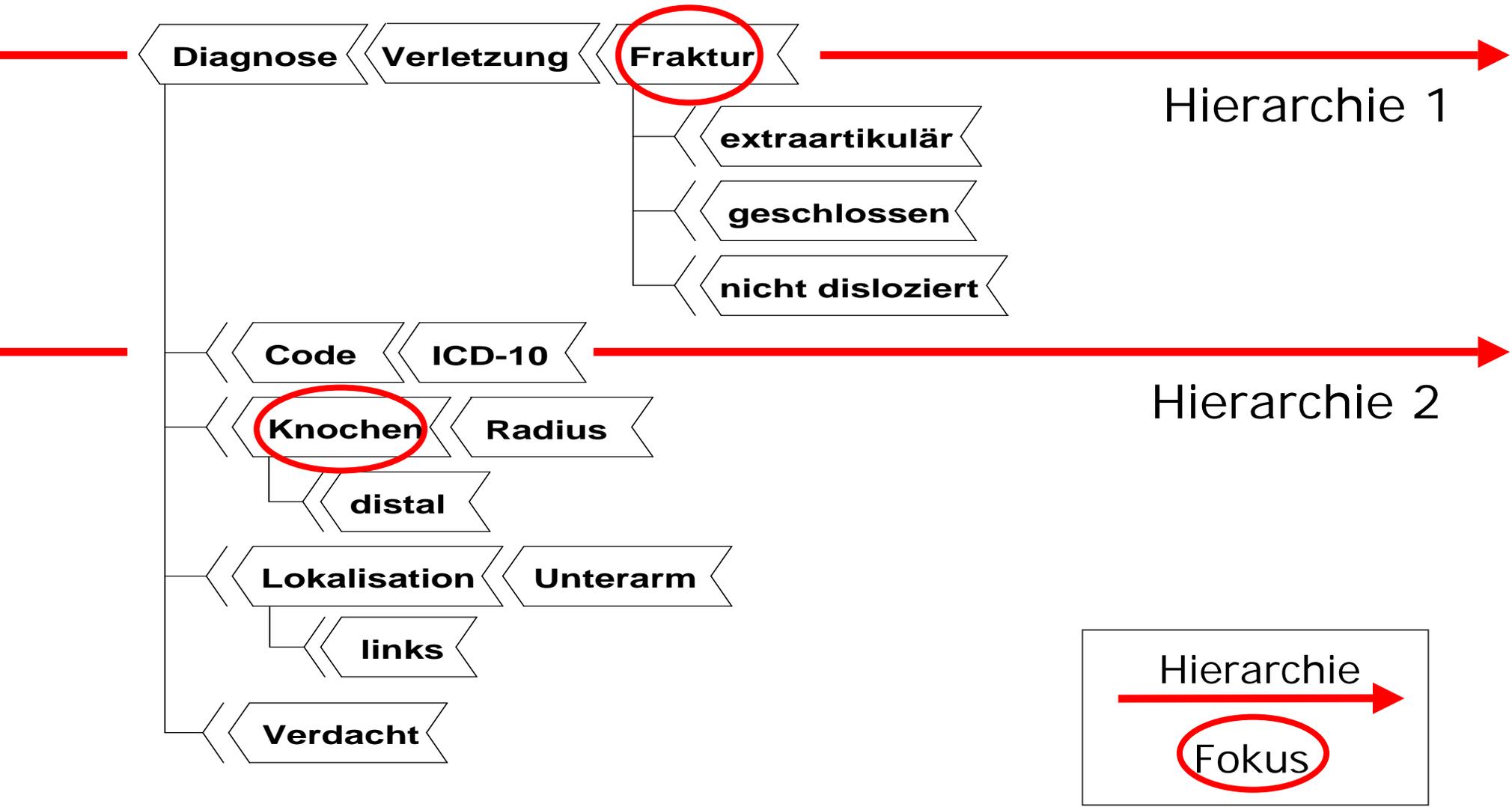
Gekapselte Achsen: multi-fokales System



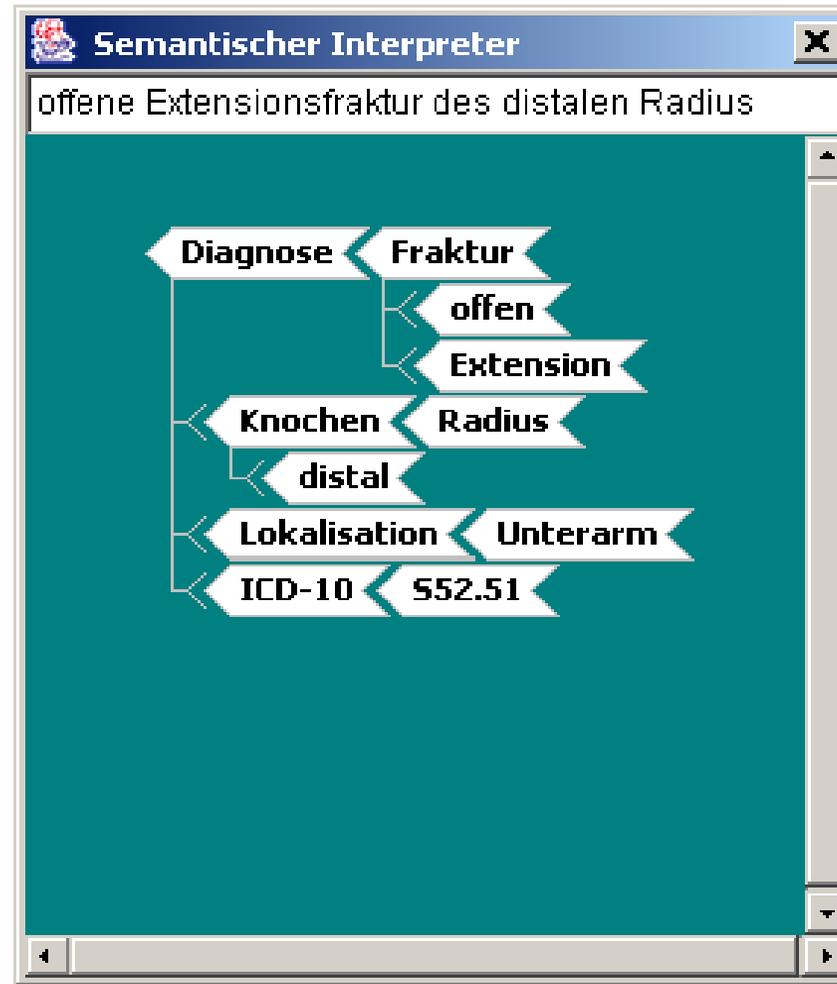
Begriffsmolekül



Begriffsmoleküle zeigen den semantischen Raum!



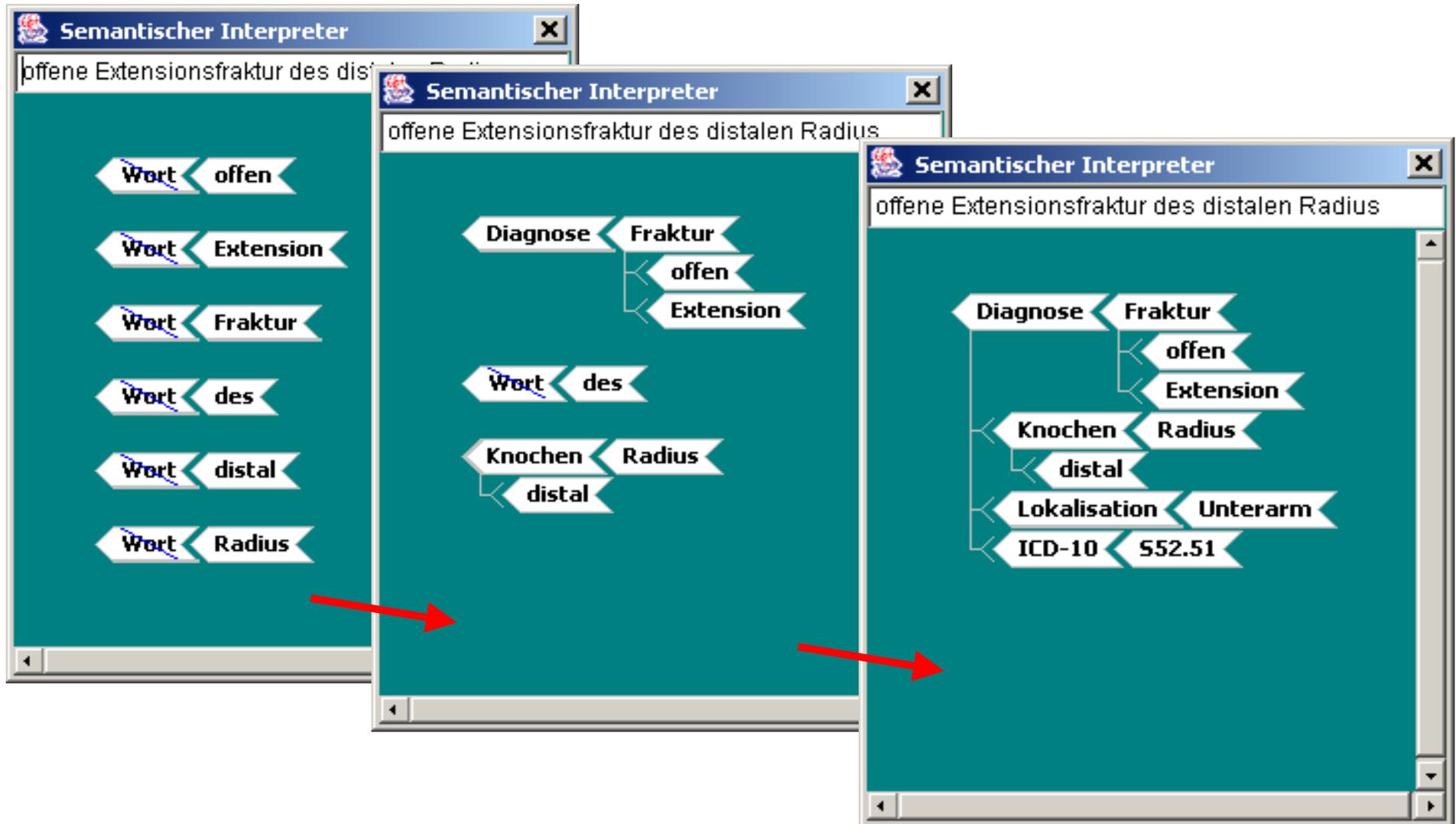
Begriffsrepräsentation in Begriffsmolekülen



← Eingabe

← Interne Repräsentation

Begriffsmoleküle: Aus Worten wird Bedeutung



Begriffsmolekül: strukturiert + komplett



Themenliste

1. Was bedeutet Kodierung?
 - Informationsreduktion
 - Zwei Arten von Kodierung
2. Semantik und Information
 - Bits und Bedeutung
 - Das semiotische Dreieck
3. Vom Text zum Kode: Probleme der automatisierten Kodierung
4. Begriffsarchitekturen:
Hierarchisch, multidimensional, multifokal
5. **Vier Methoden der Kodierunterstützung:**
Vom Stringvergleich zur Semantik
6. Semantische Interoperabilität
7. Live-Demo einer semantischen Kodierung



4 Methoden der Kodierunterstützung

Fragen, die Sie sich vor den nächsten Folien stellen können:

- Was erwarten Sie von einem Kodiertool?
- Soll ein Kodiertool auf eine Texteingabe einen Kode liefern oder mehrere?
- Soll ein Tool gezielte Fragen stellen?
- Soll es Kreuz- und Sternkodes als Kombination liefern?
- Soll es Synonyme und Unterbegriffe erkennen?
- Soll es Mehrdeutigkeiten (HWI, Kopf, Blase, ..) auflösen können?



4 Methoden der Kodierunterstützung

1. Stringvergleich in Thesauren
2. Koordinatenvergleich
3. Kaskadierte Abfrage
4. Echte semantische Modellierung



1. Stringvergleich in Thesauren

Osteoidosteom distaler Humerus → Kode?

- Osteoidosteom D16.9
- Jaffé-Osteoidosteom D16.9
- distale Humerusfraktur S42.4
- juvenile Osteochondrose des Humerus M92.0
- ...
- ...
- ... etc. ...



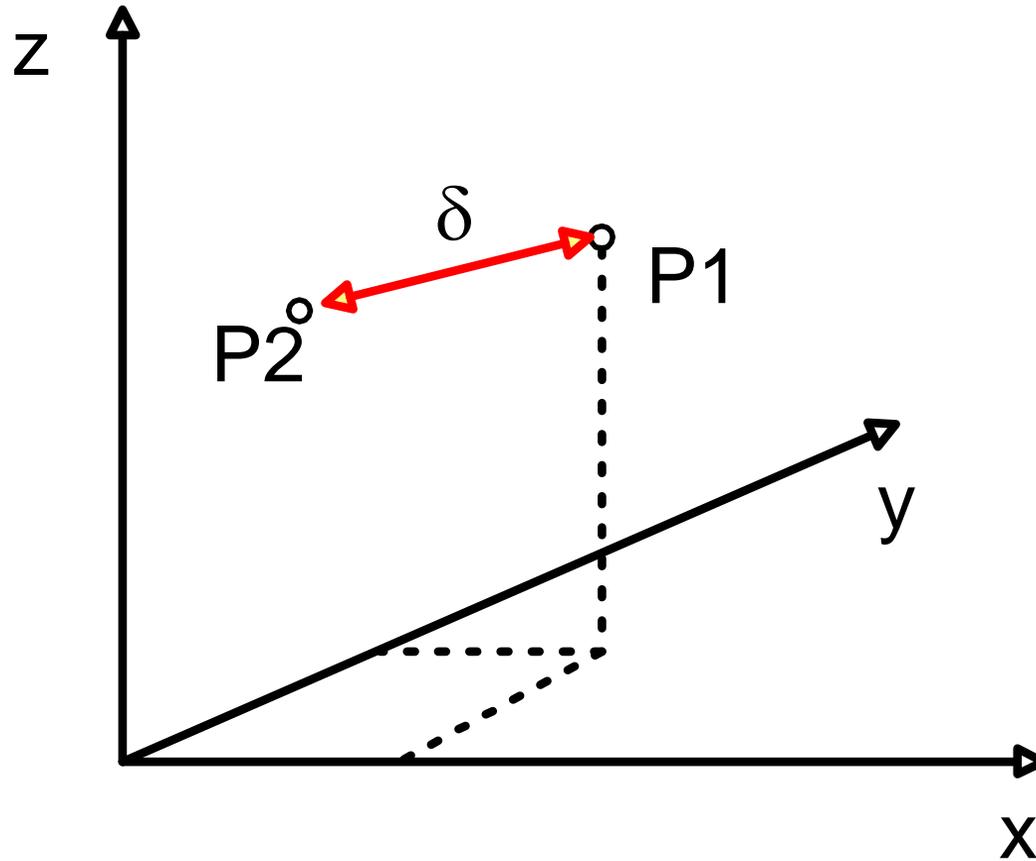
1. Stringvergleich in Thesauren

- Basis aller konventioneller Kodiertools
- Wegen kombinatorischer Explosion unvollständig
- Stringvergleich liefert viele mögliche Codes
- User muss am Ende aus Liste selber wählen

→ Keine eigenständige Kodierung durch Tool



2. Koordinatenvergleich



2. Koordinatenvergleich

- Keine komplexe Modellierung möglich
- Für Diagnosekodierung ungeeignet
- Von keinem Kodiertool auf dem Markt verwendet



3. Kaskadierte Abfrage

- Sehr präzise Kodierung möglich
- Kodierlaie wird geführt
- Problem: Wo beginnen? (Adam+Eva?)
- Deshalb **nur in Kombination** mit echt semantischer Kodierung sinnvoll



4. Echt semantische Kodierung



4. Echt semantische Kodierung

→ One-Step-Coding

→ 1 Eingabe = 1 Kodierantwort



1 Eingabe – 1 Kodierantwort

Semfinder CH 1.1 (Einzelplatz)

CHOP Code T Bisher A OK Abbruch

Semfinder CH liefert folgendes Kodierresultat:

C44.6	Bösartige Neubildung der Haut der oberen Extremität, einschließlich Schulter

OK



Mehrfachkodes – in einem Schritt

Semfinder CH 1.1 (Einzelplatz)

CHOP Code T Bisher A OK Abbruch

Semfinder CH liefert folgendes Kodiereresultat:

B37.1+	Kandidose der Lunge
J17.2*	Pneumonie bei Mykosen

OK



Einfache Fragen → korrekte Kodierung

Semfinder CH 1.1 (Einzelplatz)

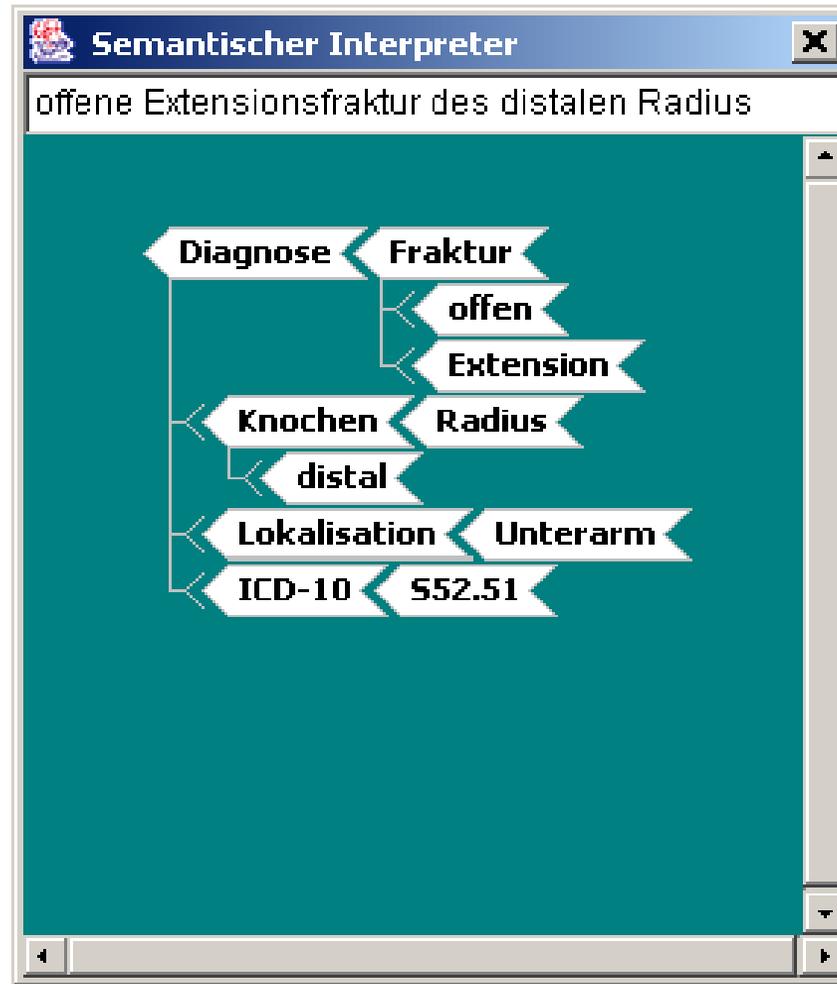
CHOP Code T Bisher A OK Abbruch

Frage

maligne
tuberkulös
chylös
bei Pleuritis
bei Pneumonie
bei Rechtsherzinsuffizienz
bei Linksherzinsuffizienz
bei globaler Herzinsuffizienz
bei Lungeninfarkt
bei anderer Ursache
unklare Ursache



4. Echt semantische Kodierung



← Eingabe

← Interne Repräsentation

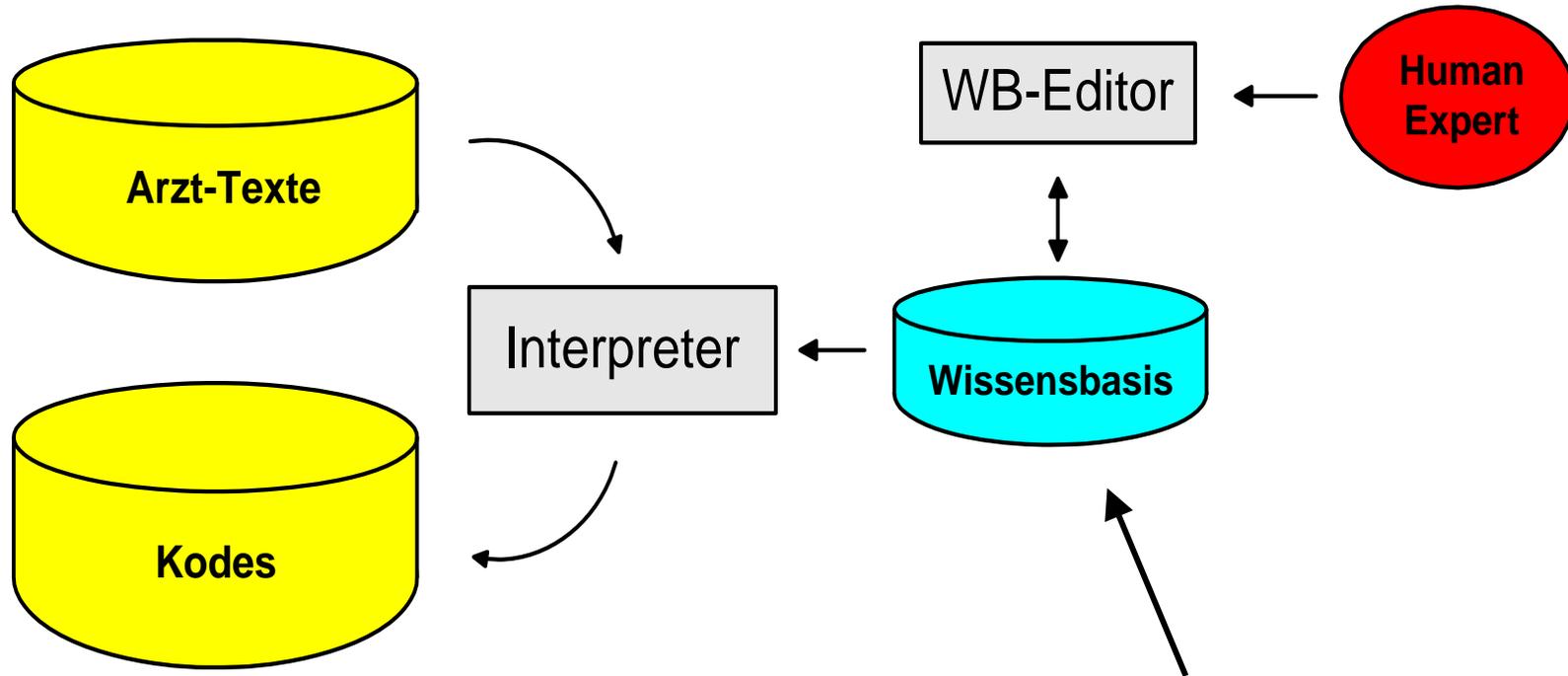


4. Echt semantische Kodierung

- Sehr präzise Kodierung möglich
- Kombinatorische Explosion wird abgefangen
- Kodierlaie wird geführt
- Eine Eingabe → Eine Kodierantwort
- Doppelkodierung in einem Schritt
- Gezielte Hinweise und Präzisierungsfragen
- Jedoch: Umfangreiche Wissensbasis nötig



4. Echt semantische Kodierung



Ein wissensbasiertes System

Themenliste

1. Was bedeutet Kodierung?
 - Informationsreduktion
 - Zwei Arten von Kodierung
2. Semantik und Information
 - Bits und Bedeutung
 - Das semiotische Dreieck
3. Vom Text zum Kode: Probleme der automatisierten Kodierung
4. Begriffsarchitekturen:
Hierarchisch, multidimensional, multifokal
5. Vier Methoden der Kodierunterstützung:
Vom Stringvergleich zur Semantik
6. Semantische Interoperabilität
7. Live-Demo einer semantischen Kodierung



Semantische Interoperabilität

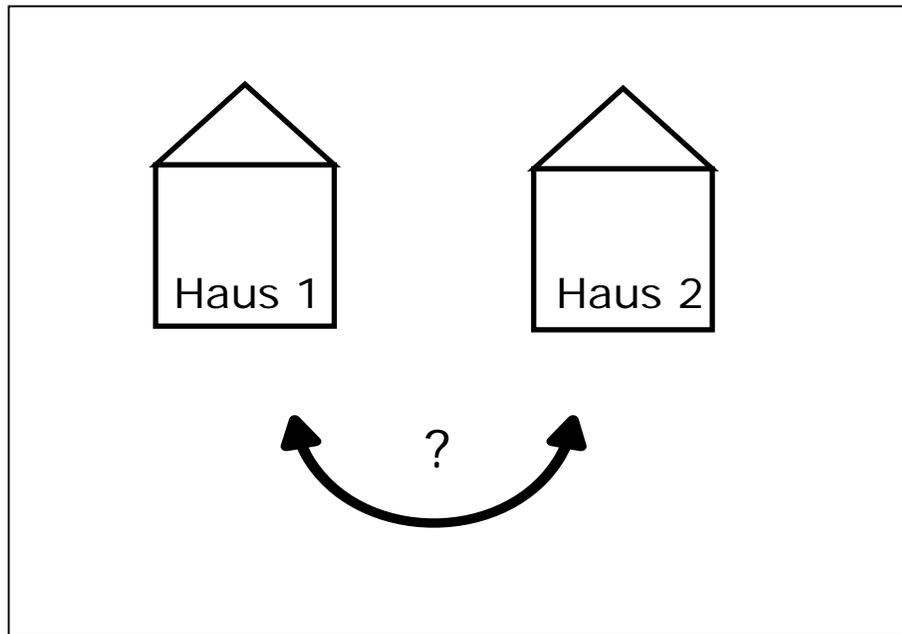
Fragen, die Sie sich vor den nächsten Folien stellen können:

- Was ist der Unterschied zwischen formaler und semantischer Interoperabilität?
- Was hat semantische Interoperabilität mit Kodierung zu tun?
- Für die Kodierung wird ein beträchtlicher Aufwand betrieben. Könnte davon neben der Abrechnung auch die Klinik profitieren?
- Die Information im ICD-10 Kode ist minimal. Könnte mehr Information strukturiert (=auswertbar) abgelegt werden?
- Was ist nötig um semantische Interoperabilität zu erreichen?



Wozu semantische Interoperabilität?

Wenn Daten unterschiedlicher Quellen (Spezialitäten, KIS, usw.) verglichen werden, müssen sie nicht nur formal, sondern auch inhaltlich interoperabel sein



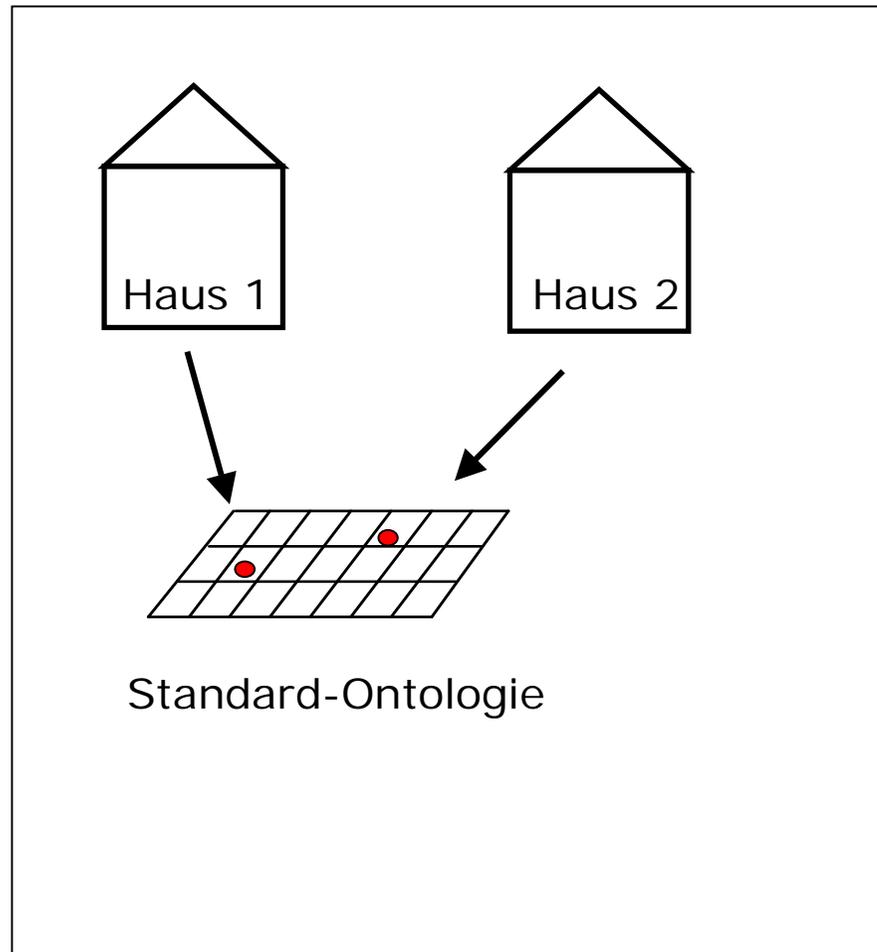
- Gleiches gleich bezeichnen
- aber Achtung:
 - Quasisynonyme
 - Teilaspekte
 - Ober- / Unterbegriffe
 - Summe-Summanden-P.
 - Pars-pro-toto
 - Medizinerdialekte
 - Kontext

Semantische Interoperabilität ist Bedingung für einen Austausch von medizinischen Daten



Statischer Ansatz

Alle Daten werden in EINE Standardontologie abgebildet (z.B. SNOMED CT)



Vorteil

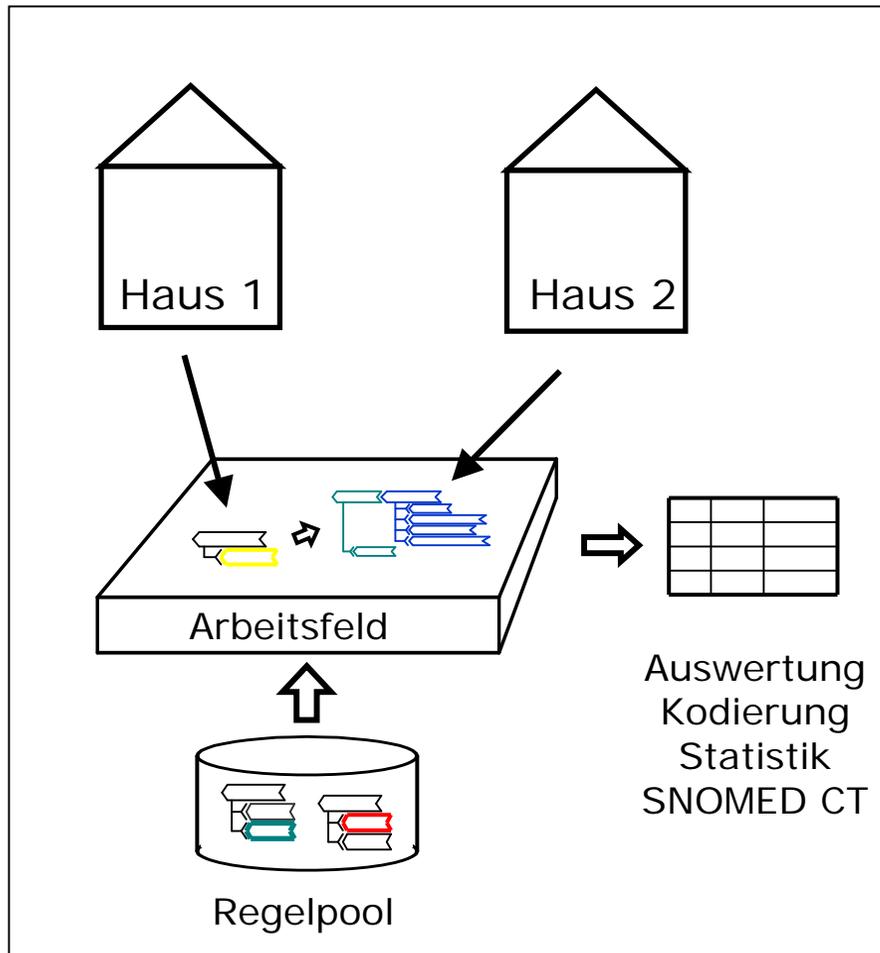
- Einigung auf klares terminologisches System

Nachteile

- Es gibt nur ein System
 - 1 System für alle ist vermutlich illusorisch
 - 1 System für immer ist sicher illusorisch
- Detailgenauigkeit lässt System "platzen" (kombinatorische Explosion. Trade-off zwischen Präzision und Wartbarkeit)
- Systematische Probleme bleiben ungelöst
 - Summe/Summanden-Problem (SSP)
 - Klassifikationen
 - Dateneingabe
 - Auswertungen, Kodierungen

Dynamischer Ansatz

Nicht vordefiniertes Gitter, sondern Methodenset (Repräsentation + Regelpool)



Der dynamische Ansatz unterscheidet:

- 1) Semantisches Netz
multidimensional, multifokal, multipunktuell
- 2) Zustände (Input, Output, u.a.)
= Ausschnitte aus Netz = Daten
- 3) Regeln
 - bearbeiten Input automatisch
 - vergleichen unterschiedliche Ontologien
 - integrieren neues Wissen (Ontologien)
 - erlauben Freitexteingaben
 - kodieren automatisch
 - erlauben automatisierte Auswertungen nach beliebigen Gesichtspunkten

Weitere Informationen

<http://www.semfinder.com/publikationen.html>

