

# Einführung in die Computerlinguistik und Sprachtechnologie

Vorlesung im WiSe 2018/19  
(B-GSW-12)

**Prof. Dr. Udo Hahn**

Lehrstuhl für Computerlinguistik  
Institut für Germanistische Sprachwissenschaft  
Friedrich-Schiller-Universität Jena

<http://www.julielab.de>

# Allgemeine Hinweise

- Vorlesung: Mi, 10-12h (Humboldt 8, SR 1)
- Übung zV: Fr, 8-10h (Fürstengrab. 1, SR 275)
  - beginnt am **19.10.**
- Vorlesungsmaterialien im Netz
  - <http://www.julielab.de/> ⇒ „Students“
- **B-GSW-12 besteht aus VL+ÜB und Seminar!**
- Sprechstunde: Mi, 12-13h (nA) (FG 30, R 004)
- Email: [udo.hahn@uni-jena.de](mailto:udo.hahn@uni-jena.de)
- URL: <http://www.julielab.de>
- Fachliteratur ist überwiegend in Englisch

# Bitte ...

- ... Handys/Smartphones ausschalten
- ... 90 Minuten ohne Mail- und Tweet-Check sind möglich  
„Digital detox“
- ... kein Picknick



# **Institut für Germanistische Sprachwissenschaft der FSU Jena**

- **Lehrstuhl für Theoretische Linguistik – Grammatiktheorie**
  - Prof. Dr. Peter Gallmann bzw. n.n.
- **Lehrstuhl für Angewandte Linguistik – Computerlinguistik**
  - Prof. Dr. Udo Hahn
- **Professur für Pragmatik**
  - Prof. Dr. Pia Bergmann
- **Professur für Phonetik & Sprechwissenschaft**
  - Prof. Dr. Adrian Simpson
- **Professur für Geschichte der deutschen Sprache**
  - Prof. Dr. Eckhard Meineke

# Computerlinguistik in Jena (1/2)

- **Institutionell: Teil der Germanistischen Sprachwissenschaft**
  - aber einzelsprachübergreifende Methodik
  - besondere Anwendungsdomänen:
    - Naturwissenschaften: Biologie + Medizin
    - Sozial- und Wirtschaftswissenschaft
    - Digital Humanities
- **Integration in die Informatik:**  
**Neben- bzw. Anwendungsfach für**
  - B.Sc.: Informatik, Angewandte Informatik
  - M.Sc.: Informatik, Computational Science

# Computerlinguistik in Jena (2/2)

- Aktive Forschergruppe
  - Lehrstuhl für Computerlinguistik = **Jena University Language & Information Engineering (JULIE) Lab**
    - Hohe internationale Visibilität (Publikationsdichte)
  - Deutsche Forschungsgemeinschaft (DFG)
    - Aktuell: (1/5) SFB 1076 **AquaDiva – Biodiversität in der Critical Zone**
    - Aktuell: Graduiertenkolleg **Modell ‚Romantik‘ [Digital Humanities]**
  - Bundesministerium für Bildung & Forschung (BMBF)
    - Aktuell: Nationale Förderinitiative „**Systemmedizin**“ (J – L – AC)
    - Frühere Projekte: Forschungs-Cluster **JenAge** – Nationaler Forschungskern, **StemNet**
  - Förderinitiativen der Europäischen Union
    - Frühere Projekte: **MANTRA (SA)**, **CALBC (SA)**, **BOOTStrep (STREP)**, ..
- Ausgründung von Start-up-Firmen
  - *Averbis, TexKnowlogy*
- **Jobs, Jobs, Jobs ... etwa als studentische Hilfskraft**
- **Themen, Themen, Themen ... BA- oder MA-Arbeit, Dissertation**

# Weitere Veranstaltungen

- Seminar zu B-GSW-12
  - SoSe 2019
- Vorlesung/Übung ASQ-DH
  - Einführung in Digital Humanities: Grundlagen der Informatisierung der Geisteswissenschaften
  - Di, 17-19, Humboldt 8, SR 3

**Computer (und Menschen!) tun  
sich schwer mit Sprache(n) ...**



# Die *pykka* Sprache

- Güney pykka-i tassas pel Criftek ut pykka-e coggy pons Criftek

– coggy	(1)
– Criftek	(2)
– Günny	(1)
– pel	(1)
– pons	(1)
– pykka-i	(1)
– pykka-e	(1)
– tassas	(1)
– ut	(1)

Lexikografische  
Ordnung

Häufigkeits-  
zählung

# Die *pykka* Sprache

- Günny pykka-i tassas pel Criftek ut pykka-e coggy pons Criftek
  - Perspektive des Computers/Menschen auf diese Äußerung:
    - uninterpretierbare Buchstaben-/Lautsequenz
  - Fehlt: Spezifikation von Wortbedeutung (Lexikon)
  - Fehlt: Regeln für Wortverknüpfung (Syntax)
  - Fehlt: Regeln für die Verbindung Syntax/Semantik
- Günny pykka-i tassas **aus** Criftek **und** pykka-e coggy **nach** Criftek
  - Pel → **aus**, ut → **und**, pons → **nach**
  - Lediglich ein Syntaxskelett

# Die *pykka* Sprache

- Günny pykka-i tassas **aus** Criftek **und** pykka-e coggy **nach** Criftek
- Deutsche Wortäquivalente:
  - { **Deutschland, Costa-Rica** }
  - { **exportieren, importieren** }
  - { **Optoelektronik, Banane** }
- **Deutschland importiert Bananen aus Costa-Rica und exportiert Optoelektronik nach Costa-Rica**

# Von *pykka* ins Deutsche I

- Günny pykka-i tassas **aus** Criftek **und** pykka-e coggy **nach** Criftek
- Deutsche Wortäquivalente:
  - [ **Deutschland** = Günny, **Costa-Rica** = Criftek ]
  - [ **importieren** = pykka-i, **exportieren** = pykka-e ]
  - [ **Banane(n)** = tassa(s), **Optoelektronik** = coggy ]
- Standard-Interpretation:

**Deutschland importiert Bananen aus Costa-Rica und exportiert Optoelektronik nach Costa-Rica**

# Von *pykka* ins Deutsche II

- Günny pykka-i tassas **aus** Criftek **und** pykka-e coggy **nach** Criftek
- Deutsche Wortäquivalente:
  - [ **Costa-Rica** = Günny, **Deutschland** = Criftek ]
  - [ **importieren** = pykka-i, **exportieren** = pykka-e ]
  - [ **Banane** = tassas, **Optoelektronik** = coggy ]
- Non-Standard-Interpretation:

**Costa-Rica importiert Bananen aus Deutschland und exportiert Optoelektronik nach Deutschland**

# Konstituenten der Analyse/ Produktion natürlicher Sprache

- Inventar von Wörtern (**Lexikon**) und ihrer Bedeutungen (lexikalische **Semantik**)
- Verknüpfungsregeln für Wörter (**Syntax**)
- Kompositionelle Ableitung der Bedeutung eines Satzes (Satz-**Semantik**) aus den lexikalischen Bedeutungen der Wörter und der Syntaxstruktur (**semantische Interpretation**)
- Evaluation der semantischen Interpretation auf der Basis von Hintergrundwissen (**Enzyklopädie, Alltagswissen** usw.)

# Computerlinguistik I

- Linguistik: Gegenstandsbereich sind (überwiegend) **natürliche Sprachen**
  - Deutsch, Englisch, Französisch, ...
- Beispiele für **formale Sprachen**
  - $L = \{a^n b^n, n \in \mathbb{N}\}$   
= {ab, aabb, aaabbb, aaaabbbb, ... }
  - jede Programmiersprache, Auszeichnungssprache
    - JAVA, C++, ..., XML, HTML, ...
  - jede Logik
    - Aussagenlogik, Prädikatenlogik, Typenlogik, ...
  - Differentialgleichungen, Integrale, Vektoren, ...

# Computerlinguistik II

- Beschreibungen und Formalisierungen entsprechen den Anforderungen, die sich aus der **Verarbeitung durch Computer** ergeben
  - keine natürlichsprachige Beschreibung (à la Duden oder Grammatik für Fremdsprachenerwerb), sondern **formalisiert** und damit explizit
  - explizite Spezifikation von Verfahrensbeschreibungen (**Algorithmen**), die von einer (abstrakten) Maschine ausgeführt werden können
  - Beachtung **formaler** (komplexitätstheoretischer) **Eigenschaften der Beschreibung**: Berechenbarkeit, Entscheidbarkeit, „Rechen-Kosten“ (Zeit, Speicher)



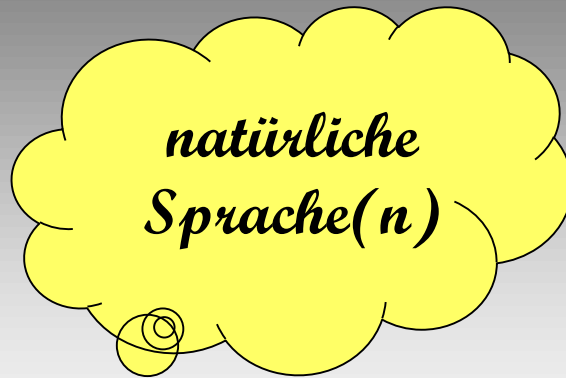
# Computerlinguistik III

- Fundierung computerlinguistischer **Beschreibungen** durch Bezug auf theoretische und methodische Prinzipien der **Linguistik und Informatik**
  - Linguistische Grammatikmodelle vs. formale Grammatikmodelle der Informatik
  - Automatenmodelle der Informatik als Grundlage des Parsings natürlicher Sprache
  - Lexikonmodelle und Suchverfahren in Lexika
  - Semantische Repräsentationsformalismen vs. Wissensrepräsentationssprachen (Beschreibungslogik)
- Notabene: die Relevanz der Informatik nimmt aktuell zu, die der Linguistik ab !

# Computerlinguistik IV

- Realisierung dieser Beschreibungen durch ihre **Implementation** in einem natürlichsprachlichen (Teil-)System entsprechend **informatischer Standards**
  - Computerlinguistik ist keine naiv „programmierte“ Linguistik
    - Programmiertechnologien (z.B. objekt-orientiert)
    - Daten(bank)technik (Speicher- und Zugriffsmethoden)
  - Software Engineering
    - Portierbarkeit (Domänenwechsel)
    - Wiederverwendbarkeit (Middleware: UIMA usw.)
    - Robustheit (NL ist ein sehr komplexes, nach wie vor nur partiell beschriebenes System)

# Verortung der Computerlinguistik



**Theoretische Linguistik**  
Phrasenstruktur-Grammatik  
Dependenzgrammatik  
Unifikationsgrammatik  
Konstruktionsgrammatik  
modelltheoretische Semantik  
strukturelle Semantik  
Frame-Semantik . . .

**Algebra**  
Formale Grammatiken & Sprachen  
Automatentheorie  
**Graphentheorie**  
**Logik**  
**Wahrscheinlichkeitstheorie**  
**Analysis, Numerik**

**Algorithmen & Datenstrukturen**  
**Programmierung**  
**Informationssysteme**  
**Künstliche Intelligenz**  
Maschinelles Lernen,  
Deduktionssysteme  
Wissensrepräsentation

**Deskription**

**Linguistik**

**Formalisierung**

**Mathematik**

**Algorithmisierung**  
**Programmierung**

**Informatik**

[www.ims.uni-stuttgart.de/info/SitesEurope.html#Germany](http://www.ims.uni-stuttgart.de/info/SitesEurope.html#Germany)



# Computerlinguistik-Standorte

[www.ims.uni-stuttgart.de/info/SitesEurope.html#Germany](http://www.ims.uni-stuttgart.de/info/SitesEurope.html#Germany)

24 [48]



U Saarbrücken (6)

U Stuttgart (3)

U Heidelberg (5)

RWTH Aachen

U München (2)

TU Darmstadt (4)

U Jena

U Tübingen (3)

U Bielefeld (4)

U Potsdam (2)

U Bremen

U Bochum (2)

U Erlangen-Nbg.

U Osnabrück (2)

U Hamburg (3)

KIT Karlsruhe

U Duisburg-Essen

U Leipzig

U Magdeburg

U Düsseldorf

U Gießen

U Hildesheim

U Koblenz

# Computerlinguistik-Standorte

[www.ims.uni-stuttgart.de/info/SitesEurope.html#Germany](http://www.ims.uni-stuttgart.de/info/SitesEurope.html#Germany)



U Saarbrücken (6)

U Stuttgart (3)

U Heidelberg (5)

RWTH Aachen

U München (2)

TU Darmstadt (4)

U Jena

U Tübingen (3)

U Bielefeld (4)

U Potsdam (2)

U Bremen

U Bochum (2)

U Erlangen-Nbg.

U Osnabrück (2)

U Hamburg (3)

KIT Karlsruhe

U Duisburg-Essen

U Leipzig

U Magdeburg

U Düsseldorf

U Gießen

U Hildesheim

U Koblenz

TU Darmstadt (2)

U Frankfurt/M. (2)

U Leipzig

U Bamberg

U Köln

U Passau

U Jena

HU Berlin

U Stuttgart

U Konstanz

U Dortmund

U Kassel

U Würzburg

U Göttingen

U Münster

U Regensburg

U Hildesheim

U Düsseldorf

U Dortmund

BU Weimar

U Bamberg

U Kaiserslautern

TU Dresden

<http://www.dig-hum.de/>

# Computerlinguistik-Stand

[www.ims.uni-stuttgart.de/info/SitesEurope.html#Germany](http://www.ims.uni-stuttgart.de/info/SitesEurope.html#Germany)

**+ 23 [25]**

**Texttechnologie**

**Digital  
Humanities**

**Informations-  
Wissenschaft /  
Information Retrieval**

U Saarbrücken (6)  
U Stuttgart (3)  
U Heidelberg (5)  
RWTH Aachen  
U München (2)  
TU Darmstadt (4)  
U Jena  
U Tübingen (3)  
U Bielefeld (4)  
U Potsdam (2)  
U Bremen  
U Bochum (2)  
U Erlangen-Nbg.  
U Osnabrück (2)  
U Hamburg (3)  
KIT Karlsruhe  
U Duisburg-Essen  
U Leipzig  
U Magdeburg  
U Düsseldorf  
U Gießen  
U Hildesheim  
U Koblenz

TU Darmstadt (2)  
U Frankfurt/M. (2)  
U Leipzig  
U Bamberg  
U Köln  
U Passau  
U Jena  
HU Berlin  
U Stuttgart  
U Konstanz  
U Dortmund  
U Kassel  
U Würzburg  
U Göttingen  
U Münster  
U Regensburg  
U Hildesheim  
U Düsseldorf  
U Dortmund  
BU Weimar  
U Bamberg  
U Kaiserslautern  
TU Dresden

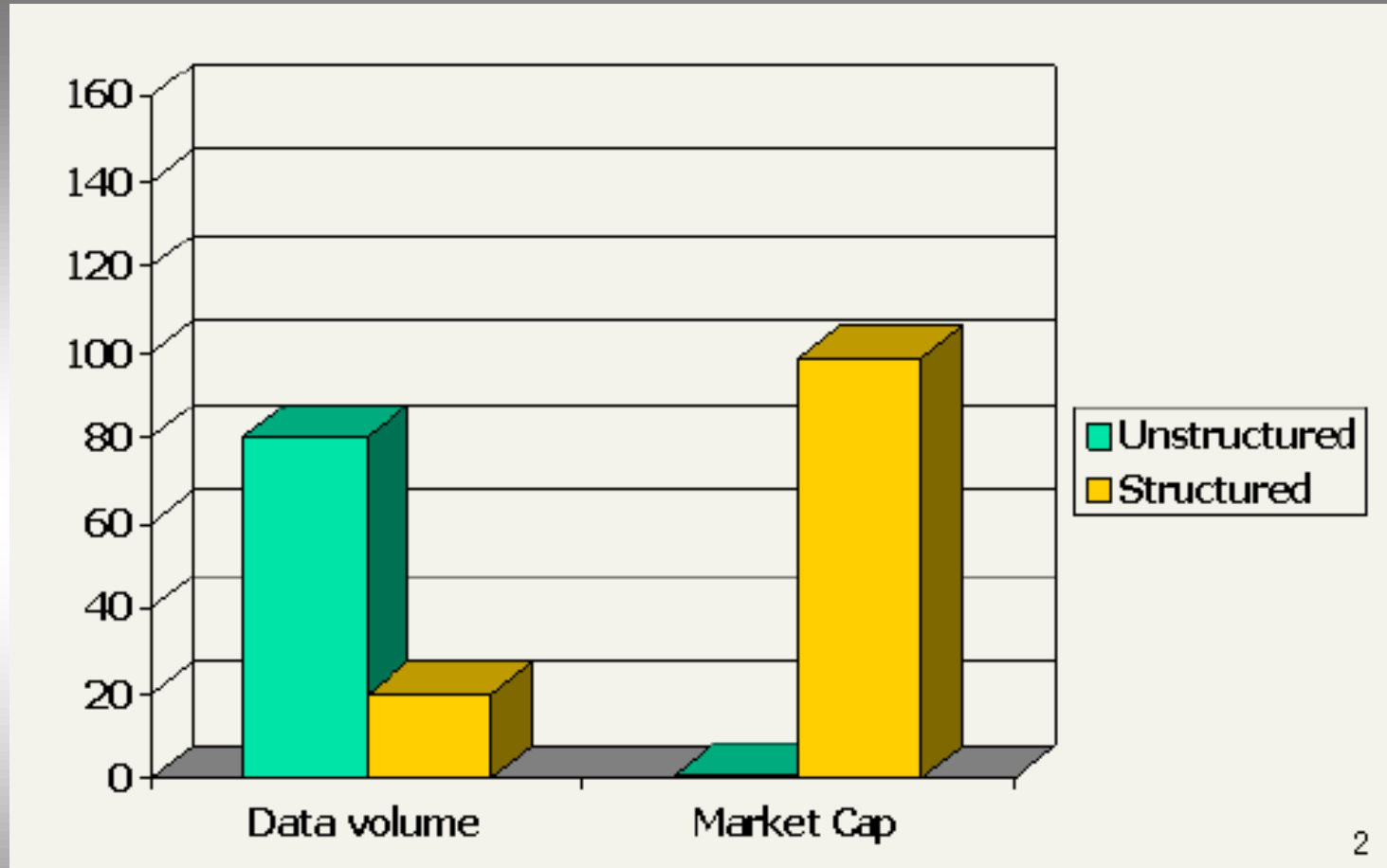
<http://www.dig-hum.de/>





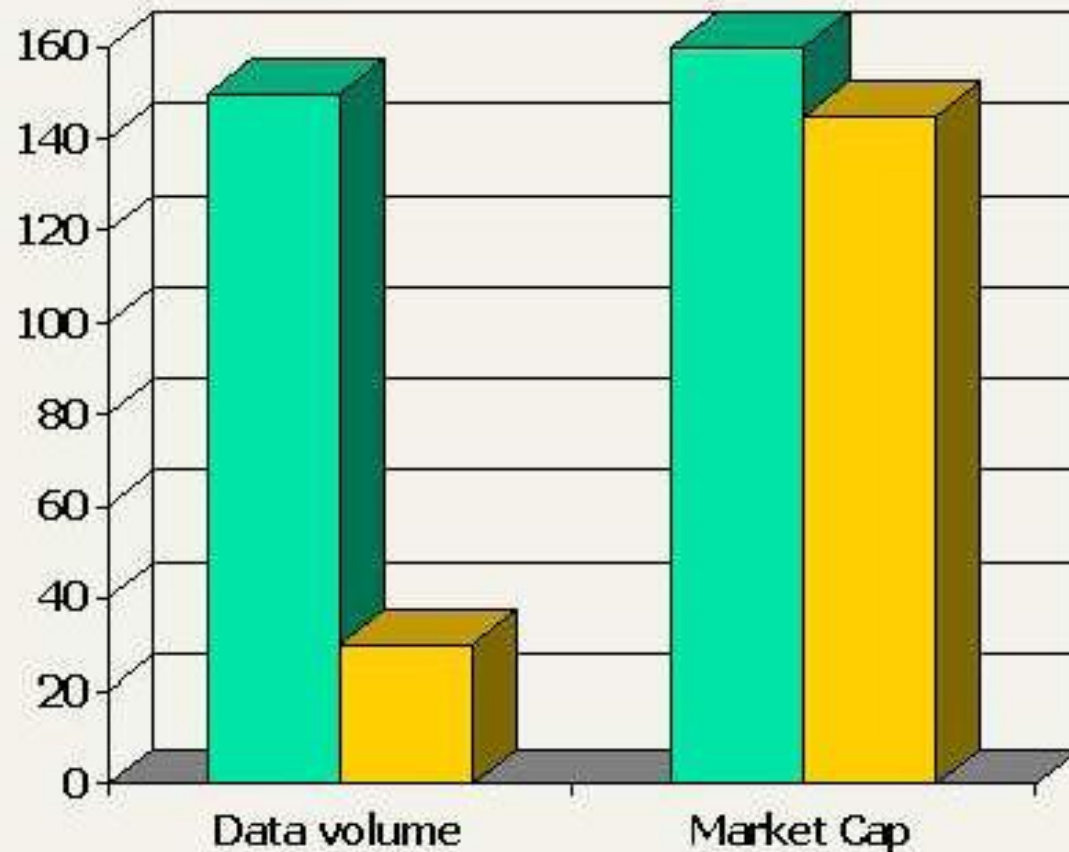


# Structured vs. Unstructured Data (1996)



2

# Structured vs. Unstructured Data (2006)



Google™

YAHOO!

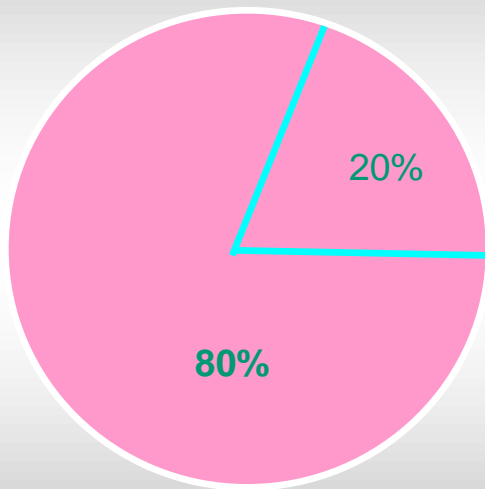
Unstructured  
Structured



Baidu 百度

# Die zunehmende Bedeutung von Sprachtechnologie

## 80:20 Argument



**Strukturierte Information  
(überwiegend numerisch kodiert)**

**Unstrukturierte oder semi-strukturierte Information  
(überwiegend verbal oder audio-visuell kodiert)**

Quelle: Oracle Corp.

# Sprachtechnologische Meilensteine

- Suchmaschinen
  - **Google** (Yahoo!, Bing, Baidu ...)
- Natürlichsprachliche Assistenten
  - Alltagsplanung: **Siri, Alexa**
  - Fragebeantwortung: **Watson, Ask (Jeeves)**
- Speech-to-Text-Systeme
  - **Dragon** Dictate, ...
- Übersetzungssysteme
  - **Google Translate, Microsoft Translator , Language Weaver, Lernout & Houspie, DeepL DeepL ...**
- Linguistisch sensitive Schreibprogramme
  - **Word**, ...

apc - Google-Suche - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.google.de/#hl=de&source=hp&q=apc&btnG=Goog

Google

LEO Thesaurus Aigaion DBIS Thulb Jena Coling Wiki ServerBelegung PHD Comics Desktop Application...

apc - Google-Suche

Web Bilder Videos Maps News Shopping E-Mail Mehr

Sucheinstellungen | Anmelden

Google

apc

Suche Erweiterte Suche

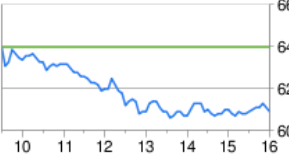
Suche: ☒ Das Web ☐ Seiten auf Deutsch ☐ Seiten aus Deutschland

Web Ergebnisse 1 - 10 von ungefähr 21.200.000 für apc. (0,07 Sekunden)

**APC-USV Shop** [www.battery-direct.de](#) USV Batterien und Anlagen, schnell sicher und günstig online bestellen

**APC Consumer Produkte** [www.apc-home.de](#) Stromversorgungslösungen für PC und Home-Entertainment Systeme

**APC** - Anadarko Petroleum Corporation (NYSE)  
[Google Finance](#) [Yahoo Finance](#) [MSN Money](#) [AOL Finance](#) [CNN Money](#) [Reuters](#)


**60.93 -3.05 (-4.77%)** 30. Okt. 16:01 ET  
**61.75 +0.82 (1.35%)** Nachbörslich  
 Eröffnung: 63.70 Volumen: 0  
 Hoch: 64.00 Durchschn. Vol: 5,577,000  
 Tief: 60.50 Marktkap.: 29.89B  
[Haftungsausschluss](#)

**APC Deutschland, USV, Serverraum, Rechenzentrum Kühlung, RZ ...**  
 Innovative, skalierbare Lösungen von **APC** für Datacenter. Infrastruktur für  
 gesundes Wachstum Die InfraStruXure Lösung überzeugte uns nicht nur aufgrund ...  
[apc.com/DE](#) - [Im Cache](#) - [Ähnlich](#)

**APC - Country Selection Page** - [ [Diese Seite übersetzen](#) ]  
 Manufacturer of a broad range of commercial uninterruptible power supplies,  
 including standby, line interactive and online technologies.  
[www.apc.com/](#) - [Im Cache](#) - [Ähnlich](#)

[+ Weitere Ergebnisse anzeigen von www.apc.com](#)

**A.P.C.** - [ [Diese Seite übersetzen](#) ]  
[A.P.C. Madras - Madras from A.P.C - Zip mini skirt](#)  
[www.apc.fr/](#) - [Im Cache](#) - [Ähnlich](#)

**PECL :: Package :: APC** - [ [Diese Seite übersetzen](#) ]  
 License, PHP License. Description, **APC** is a free, open, and robust framework for  
 caching and optimizing PHP intermediate code. ...  
[pecl.php.net/apc](#) - [Im Cache](#) - [Ähnlich](#)

**APC - Wikipedia**  
 15. Sept. 2009 ... Die Abkürzung **APC** steht für: A Perfect Circle, Progressive  
 Rock-Band; Adenomatous polyposis coli-Protein, ein Tumor-Suppressor-Gen und - ...  
[de.wikipedia.org/wiki/APC](#) - [Im Cache](#) - [Ähnlich](#)

# APC (1/2)

apc - Google-Suche - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.google.de/#hl=de&source=hp&q=apc&btnG=Goog

Google

apc - Google-Suche

including standby, line interactive and online technologies.  
[www.apc.com/](#) - [Im Cache](#) - [Ähnlich](#)

[Weitere Ergebnisse anzeigen von www.apc.com](#)

**A.P.C.** - [ [Diese Seite übersetzen](#) ]  
[A.P.C. Madras](#) - [Madras from A.P.C](#) - [Zip mini skirt](#)  
[www.apc.fr/](#) - [Im Cache](#) - [Ähnlich](#)

**PECL :: Package :: APC** - [ [Diese Seite übersetzen](#) ]  
License, PHP License. Description, **APC** is a free, open, and robust framework for caching and optimizing PHP intermediate code. ...  
[pecl.php.net/apc](#) - [Im Cache](#) - [Ähnlich](#)

**APC** - Wikipedia  
15. Sept. 2009 ... Die Abkürzung **APC** steht für: A Perfect Circle, Progressive Rock-Band; Adenomatous polyposis coli-Protein, ein Tumor-Suppressor-Gen und - ...  
[de.wikipedia.org/wiki/APC](#) - [Im Cache](#) - [Ähnlich](#)

**APC Propellers, props for model airplanes** - [ [Diese Seite übersetzen](#) ]  
**APC** propeller manufacturer of 2, 3 and 4 blade props intended for model aircraft. The design methodology used results in efficient, low noise props.  
[www.apcprop.com/](#) - [Im Cache](#) - [Ähnlich](#)

**APC-Resistenz - Onmeda: Medizin & Gesundheit**  
24. Sept. 2009 ... Hervorgerufen wird die **APC**-Resistenz durch ein verändertes Gen (Gene ... Die **APC**-Resistenz wird anhand einer sogenannten Mutationsanalyse ...  
[www.onmeda.de/krankheiten/apc\\_resistenz.html](#) - [Im Cache](#) - [Ähnlich](#)

**APC Schweiz, USV, Rechenzentrum und Serverraum Kühlung, RZ ...**  
Innovative, skalierbare Lösungen von **APC** für Datencenter. HEXAL AG ... **APC** by Schneider Electric ernannt Philippe Arsonneau zum Vice President (02-07-2009) ...  
[www.apcc.com/index.cfm?ISOCountryCode=CHDE](#) - [Im Cache](#) - [Ähnlich](#)

**APC - Support** - [ [Diese Seite übersetzen](#) ]  
The **APC** Customer Solutions Team regularly updates this database with information gathered from **APC** Engineers, Product Managers and Customers like yourself. ...  
[www.apcc.com/support/index.cfm](#) - [Im Cache](#) - [Ähnlich](#)

**Association for Progressive Communications | Internet for social ...** - [ [Diese Seite übersetzen](#) ]  
A global Internet community; site has links to civil society organizations in many countries working for environmental protection, human rights, development ...  
[www.apc.org/](#) - [Im Cache](#) - [Ähnlich](#)

Verwandte Suchvorgänge: **apc**

[apc usv](#)   [apc.php](#)   [apc ups](#)   [apc resistenz](#)  
[apc smart ups](#)   [apc powerchute](#)   [apc jeans](#)   [apc](#)

Goooooooooooooogle ▶  
1 2 3 4 5 6 7 8 9 10 Vorwärts

# APC (2/2)

exzessive  
lexikalische  
Ambiguität

Twease helps you find precisely what you're looking for, quickly. Search Medline™ abstracts in new ways and find the information you need. Type your search phrase here and click "Search" to get started, or, take a tour of Twease with our Tutorial.

### Search

Order results by ☐ Relevance ☒ Publication Date ☐ Oldest First

19116273



[Activated protein C ligation of apoer2 \( LRP8 \) causes Dab1 dependent signaling in U937 cells. \(2009/01/07\)](#)

of [activated protein C \( APC \)](#) to cells triggers multiple ... breakdown . One paradigm for **APC** s signaling emphasizes its binding ... apoer2 ) dependent signaling by **APC** and found that APC initiated ... by APC and found that **APC** initiated rapid phosphorylation of Tyr ... Ser 473 in Akt . **APC** also induced phosphorylation of Ser ... receptor family members , inhibited **APC** induced phosphorylation of Dab1 and ... apoer2 by using sirna ablated **APC** induced Dab1 phosphorylation , suggesting ... , suggesting that RAP sensitive **APC** induced signaling requires apoer2 . ... resonance equilibrium binding studies , **APC** bound with high affinity to ... lipoprotein receptor . RAP blocked **APC** binding to sapoer2 but not ... of U937 cells to immobilized **APC** . RAP also blocked APC ... APC . RAP also blocked **APC** s ability to inhibit endotoxin ... that ligation of apoer2 by **APC** signals via Dab1 phosphorylation and ... gsk3beta , thereby contributing to **APC** s beneficial effects on cells

[Related Articles](#)

19124733



[Kinetics of major histocompatibility class I antigen presentation in acute infection. \(2009/01/06\)](#)

between the size of the **APC** population and duration of Ag ... . In both infections , **APC** production was calculated to occur ... rate of dendritic cells . **APC** production was found to closely ... , in influenza infection , **APC** production ceased between the third ... self limiting infections generate the **APC** necessary to drive T cell

[Related Articles](#)

[\[APC = antigen-presenting cells\]](#)

19122503



[Argon plasma coagulation therapy. \(2009/01/05\)](#)

in [argon plasma coagulation \( APC \)](#) therapy in gastroenterology . ... gastroenterology . recent findings : **APC** has gained importance in gastrointestinal ... the last decade , and **APC** technology has improved over time ... series , have shown that **APC** can be used effectively in ... are low . summary : **APC** can be considered as a ... trials including a comparison of **APC** to other treatment modalities as ... long term follow up after **APC** treatment are required to clarify ... clarify the relative value of **APC** in gastroenterology .

[Related Articles](#)

19121182



[Sterility screening of platelet concentrates : questioning the optimal test strategy. \(2009/01/05\)](#)  
of [apheresis platelet concentrates \( APC \)](#) and pooled platelet concentrates

[Related Articles](#)

# Semantische Suchmaschine

## Semedico

- Ermöglicht einen tieferen, semantisch-ontologisch strukturierten Zugang zu biomedizinischen Fachtexten
- semantisch
  - Automatische Erkennung von Entitäten (=Begriffe, Konzepte)
  - Abbildung von Synonymen und Abkürzungen auf Konzept-Identifikatoren



# Semantischer Zugang



# Semantischer Zugang

Browser address bar: <http://www.semedico.org/app;jsessionid=5FC24D6C0E5F0D676448EC39CE82BC55>

 **semedico**  
search PubMed **beta**

Search input:

Select a term to search for synonyms and related terms. Press ESC to ignore suggestions.

Result 1-10 of **157870** (1266 ms)

BioMed Immunology Bibliography

▼ **Chemicals and Drugs** (128053)

- Enzymes and Coenzymes (56619)
- Chemical Actions and Uses (52644)
- Organic Chemicals (42905)

[more..](#)

▼ **Investigative Techniques** (117902)

- Clinical Laboratory Techniq... (45025)
- Immunologic Techniques (30227)
- Genetic Techniques (29169)

[more..](#)

▼ **Organisms** (117383)

- Homo sapiens (65319)
- Mus musculus (45501)
- Rattus (18999)

[more..](#)

These terms define your query. Click ☐ to remove a term.

apc ▶ is a synonym for 10 terms, [refine](#) ☐

[show review articles only](#) sort by:


Ohnishi Hiroshi, Miyahara Nobuaki, Gelfand Erwin

 [The Role of Leukotriene B\(4\) in Allergic Diseases.](#)

*Allergology international : official journal of the Japanese Society of Allergology.*  
2008 December 01; 57 : 4

- ...and that is rapidly generated from activated innate immune cells such as neutrophils, **macrophages**, and mast cells...
- .../**macrophages**, mast cells, **dendritic cells**, and more recently, effector T cells to inflamed tissues...

Yu Ying, Koike Tomonari, Kitajima Shuji, Liu Enqi, Morimoto Masatoshi, Shiomi Masashi, Hatakeyama Kinta, Asada Yujiro, Wang Ke-Yong, Sasaguri Yasuyuki, Watanabe Teruo, Fan Jianglin

 [Temporal and quantitative analysis of expression of metalloproteinases \(MMPs\) and their endogenous inhibitors in atherosclerotic lesions.](#)

*Histology and histopathology.* 2008 December 01; 23 (12) : 1503-16

# Semantischer Zugang

http://www.semedico.org/app?component=%24QueryPanel.%24DirectLink&page=HitList&service=direct&session=T&sp=Sapc



apc

Select a term to search for synonyms and related terms. Press ESC to ignore suggestions.

Result 1-10 of **157870** (1266 ms)

BioMed Immunology Bibliography

▼ **Chemicals and Drugs** (128053)

- Enzymes and Coenzymes (56619)
- Chemical Actions and Uses (52644)
- Organic Chemicals (42905)
- [more..](#)

▼ **Investigative Techniques** (117902)

- Clinical Laboratory Techniq... (45025)
- Immunologic Techniques (30227)
- Genetic Techniques (29169)
- [more..](#)

▼ **Organisms** (117383)

- Homo sapiens (65319)
- Mus musculus (45501)
- Rattus (18999)
- [more..](#)

These terms define your query. Click ☐ to remove a term.

What did you mean with "apc"?  
We found the following synonyms:

**Diseases / Pathological Processes:**

- ☒ Atrial Premature Complexes
- ☐ Adenomatous Polyposis Coli

**Genes and Proteins:**

- ☐ PIFC (Escherichia coli (strain K12))
- ☐ APC (any organism)

**Blood Cells:**

- ☐ Antigen-presenting cells

**Chemicals and Drugs:**

- ☐ N-(3-aminopropyl)cadaverine
- ☐ allylpyrocatechol
- ☐ anaphase-promoting complex
- ☐ Aphidicolin
- ☐ allophycocyanin

[don't refine "apc"](#)

# Semantischer Zugang

<http://www.semedico.org/app?component=%24QueryPanel.%24DisambiguationPanel.%24DirectLink&page=HitList&service=direct&session=T&sp=Sapc&sp=O&s>

### Diseases / Pathological Processes (266)

- Cardiovascular Diseases (266)
- Pathologic Processes (60)
- Nervous System Diseases (24)
- [more..](#)

### Investigative Techniques (199)

- Epidemiologic Methods (136)
- Catheterization (67)
- Evaluation Studies as Topic (15)
- [more..](#)

### Organisms (188)

- Homo sapiens (176)
- Canis familiaris (10)
- Rattus (4)
- [more..](#)

### Therapies and Treatments (115)

- Cardiac Pacing, Artificial (29)
- Complementary Therapies (13)
- Electric Stimulation Therapy (12)
- [more..](#)

### Chemicals and Drugs (113)

- Organic Chemicals (57)
- Chemical Actions and Uses (51)
- Pharmaceutical Preparations (28)
- [more..](#)

### Signs and Symptoms (12)

- Heart Murmurs (4)
- Edema (2)
- Signs and Symptoms, Respiratory (2)
- [more..](#)

### Genes and Proteins (11)

- GTPB1 (any organism) (1)
- CALR (any organism) (1)
- CO3 (any organism) (1)
- [more..](#)

### Gene Expression (3)

## Diseases / Pathological Processes: Cardiovascular Diseases ▶ Heart Diseases ▶ Arrhythmias, Cardiac ▶ Cardiac Complexes, Premature ▶ Atrial Premature Complexes

show review articles only      sort by: date and relevance ▼

---

Achen Sarah E, Miller Matthew W  
[ECG of the Month. Atrial premature complex.](#)  
*Journal of the American Veterinary Medical Association*. 2008 August 15; 233 (4) : 561-2

---

Yamada Takumi, Murakami Yoshimasa, Okada Taro, McElderry Hugh Thomas, Doppalapudi Harish, Epstein Andrew E, Plumb Vance J, Murohara Toyooka, Kay George Neal  
[Electroanatomic mapping in the catheter ablation of premature atrial contractions with a non-pulmonary vein origin.](#)  
*Europace : European pacing, arrhythmias, and cardiac electrophysiology : journal of the working groups on cardiac pacing, arrhythmias, and cardiac cellular electrophysiology of the European Society of Cardiology*. 2008 August 28; :

- Symptomatic **premature atrial contractions (PACs)** may be a target for catheter ablation...
- ... However, mapping of **PACs** with an atrial origin may not be easy because of erratic incidence and different sites...
- ...in stable arrhythmias, electroanatomic mapping of **PACs** in intermittent arrhythmias...

---

Gorenk Bulent, Parspour Afshin, Birdane Alparslan, Cavusoglu Yuksel, Goktekin Omer, Unalir Ahmet, Ata Necmi, Timuralp Bigin  
[Can short-long-short sequences predict atrial fibrillation in acute inferior myocardial infarction?](#)  
*International journal of cardiology*. 2008 July 04; 127 (2) : 260-1

- Short-long-short sequences (SLSS), related to **atrial ectopic beats (AEBs)**, predict the recurrence...
- AEBs** were more frequent. We concluded that the presence of frequent SLSS could be predictors...

---

Matsutani Hideyuki, Sano Tomonari, Kondo Takeshi, Morita Hitomi, Arai Takehiro, Sekine Takako, Takase Shinichi, Oida Akitsugu, Fukazawa Hiroshi, Sugata Masahiko, Kondo Makoto, Kodama Takahide, Orihara Tadaaki, Yamada Norikazu, Tsuyuki Masaharu, Narula Jagat  
[ECG-edit function in multidetector-row computed tomography coronary arteriography for patients with arrhythmias.](#)  
*Circulation journal : official journal of the Japanese Circulation Society*. 2008 July 01; 72 (7) : 1071-8

- ...**premature atrial contractions (PAC)**, 42 premature ventricular contractions (PVC), 3 **PAC + PVC**, 10 sinus arrhythmias...
- ...% for the 28 patients with Af; 56%, 36% and 8% for the 39 **PAC** patients,

# Ontologischer Zugang

- Ermöglicht einen tieferen, semantisch-ontologisch strukturierten Zugang zu Medline-Texten
- semantisch
  - Automatische Erkennung von Entitäten (=Begriffe, Konzepte)
  - Abbildung von Synonymen und Abkürzungen auf Konzept-Identifikatoren
- ontologisch
  - Taxonomien werden bei der Suche mit einbezogen
  - Ontologisch motivierte Facetten ermöglichen Browsing bei Suche

# Ontologischer Zugang

http://www.semedico.org/app?component=diseases.%24DirectLink&page=HitList&service=direct&session=T&sp=0

 **semedico**  
search PubMed beta

dementia search

Result 1-10 of 53154 (383 ms)

BioMed Immunology Bibliography

**Diseases / Pathological Processes** (53121)  
Nervous System Diseases  
Central Nervous System Diseases  
Brain Diseases  
Dementia  
Alzheimer Disease (39756)  
Creutzfeldt-Jakob Syndrome (3174)  
Pick Disease of the Brain (554)  
Aphasia, Primary Progressive (179)  
Kluver-Bucy Syndrome (29)  
[less..](#)

**Investigative Techniques** (36918)  
Epidemiologic Methods (20029)  
Clinical Laboratory Techniques (7541)  
Genetic Techniques (5829)  
[more..](#)

**Chemicals and Drugs** (34987)  
Chemical Actions and Uses (14914)  
Enzymes and Coenzymes (13007)  
Organic Chemicals (11860)  
[more..](#)

**Organisms** (33569)  
Homo sapiens (28290)  
Mus musculus (5027)  
Rattus (3745)  
[more..](#)

**Genes and Proteins** (25128)  
A4 (any organism) (8163)  
TAU (any organism) (3822)

These terms define your query. Click ☒ to remove a term.

**Diseases / Pathological Processes:** [Nervous System Diseases](#) ☒ [Central Nervous System Diseases](#) ☒ [Brain Diseases](#) ☒ [Dementia](#) ☒

[show review articles only](#) sort by:

McLean Catriona A  
[Review. The neuropathology of kuru and variant Creutzfeldt-Jakob disease.](#)  
*Philosophical transactions of the Royal Society of London. Series B, Biological sciences.* 2008 November 27; 363 (1510) : 3685-3687  
...presumptive route of infection was performed. Archival kuru and recent variant **Creutzfeldt-Jakob disease...**

Wadsworth Jonathan D F, Joiner Susan, Linehan Jacqueline M, Asante Emmanuel A, Brandner Sebastian, Collinge John  
[Review. The origin of the prion agent of kuru: molecular and biological strain typing.](#)  
*Philosophical transactions of the Royal Society of London. Series B, Biological sciences.* 2008 November 27; 363 (1510) : 3747-3753  
...cerebellar ataxia and, in sharp contrast to most cases of sporadic **Creutzfeldt-Jakob disease (CJD), dementia...**  
...**CJD**, which also has similar prodromal symptoms of sensory disturbance and joint pains in the legs...  
...with sporadic, iatrogenic and variant **CJD** in both transgenic and wild-type mice. These data...

Brandner Sebastian, Whitfield Jerome, Boone Ken, Puwa Anderson, O'Malley Catherine, Linehan Jacqueline M, Joiner Susan, Scaravilli Francesco, Calder Ian, P Alpers Michael, Wadsworth Jonathan D F, Collinge John  
[Central and peripheral pathology of kuru: pathological analysis of a recent case and comparison with other forms of human prion disease.](#)  
*Philosophical transactions of the Royal Society of London. Series B, Biological sciences.* 2008 November 27; 363 (1510) : 3755-3763  
...with those seen in classical (sporadic and iatrogenic) **Creutzfeldt-Jakob disease (CJD)** and variant **CJD**...  
...(v**CJD**). The neuropathological findings of the kuru patient showed all the

# Ontologischer Zugang

http://www.semedico.org/app?component=diseases.%24DirectLink&page=HitList&service=direct&session=T&sp=0

 **semedico**  
search PubMed beta

dementia

Select a term to search for synonyms and related terms. Press ESC to ignore suggestions.

Result 1-10 of 53154 (383 ms)

BioMed Immunology Bibliography

**Diseases / Pathological Processes** (53121)

- Nervous System Diseases
  - Central Nervous System Diseases
    - Brain Diseases
      - Dementia**
        - Alzheimer Disease (39756)
        - Creutzfeldt-Jakob Syndrome (3174)
        - Pick Disease of the Brain (554)
        - Aphasia, Primary Progressive (179)
        - Kluver-Bucy Syndrome (29)

[less..](#)

**Investigative Techniques** (36918)

- Epidemiologic Methods (20029)
- Clinical Laboratory Techniques (7541)
- Genetic Techniques (5829)

[more..](#)

**Chemicals and Drugs** (34987)

- Chemical Actions and Uses (14914)
- Enzymes and Coenzymes (13007)
- Organic Chemicals (11860)

[more..](#)

**Organisms** (33569)

- Homo sapiens (28290)
- Mus musculus (5027)
- Rattus (3745)

[more..](#)

**Genes and Proteins** (25128)

- A4 (any organism) (8163)
- TAU (any organism) (3822)

These terms define your query. Click ☒ to remove a term.

**Diseases / Pathological Processes:** [Nervous System Diseases](#) ☒ [Central Nervous System Diseases](#) ☒ [Brain Diseases](#) ☒ [Dementia](#)

[show review articles only](#) sort by:

McLean Catriona A

 [Review. The neuropathology of kuru and variant Creutzfeldt-Jakob disease.](#)

*Philosophical transactions of the Royal Society of London. Series B, Biological sciences.* 2008 November 27; 363 (1510) : 3685-3687

- ...presumptive route of infection was performed. Archival kuru and recent variant **Creutzfeldt-Jakob disease...**

Wadsworth Jonathan D F, Joiner Susan, Linehan Jacqueline M, Asante Emmanuel A, Brandner Sebastian, Collinge John

 [Review. The origin of the prion agent of kuru: molecular and biological strain typing.](#)

*Philosophical transactions of the Royal Society of London. Series B, Biological sciences.* 2008 November 27; 363 (1510) : 3747-3753

- ...cerebellar ataxia and, in sharp contrast to most cases of sporadic **Creutzfeldt-Jakob disease (CJD), dementia...**
- CJD**, which also has similar prodromal symptoms of sensory disturbance and joint pains in the legs...
- ...with sporadic, iatrogenic and variant **CJD** in both transgenic and wild-type mice. These data...

Brandner Sebastian, Whitfield Jerome, Boone Ken, Puwa Anderson, O'Malley Catherine, Linehan Jacqueline M, Joiner Susan, Scaravilli Francesco, Calder Ian, P Alpers Michael, Wadsworth Jonathan D F, Collinge John

 [Central and peripheral pathology of kuru: pathological analysis of a recent case and comparison with other forms of human prion disease.](#)

*Philosophical transactions of the Royal Society of London. Series B, Biological sciences.* 2008 November 27; 363 (1510) : 3755-3763

- ...with those seen in classical (sporadic and iatrogenic) **Creutzfeldt-Jakob disease (CJD)** and variant **CJD**...
- ...(vCJD). The neuropathological findings of the kuru patient showed all the

# Ontologischer Zugang



The screenshot displays the SemaMedico web application. The browser address bar shows <http://www.semedico.org/app>. The left sidebar contains a navigation menu with the following items:

- Gene Expression (2057)
- Signs and Symptoms (900)
  - Signs and Symptoms, Respira... (232)
  - Signs and Symptoms, Digestive (208)
  - Body Temperature Changes (81)
- [more..](#)
- [configure...](#)

The main content area displays several article snippets:

**Dukes Kyle D, Rodenberg Cassandra F, Lammi Robin K**  
[Monitoring the earliest amyloid-beta oligomers via quantized photobleaching of dye-labeled peptides.](#)  
*Analytical biochemistry*. 2008 November 01; 382 (1) : 29-34

- ...events in **Alzheimer's disease (AD)**. Contrary to earlier hypotheses, recent studies have identified...
- ...are paramount to understanding, preventing, and treating **AD**. We have exploited quantized...

**Gajdusek D Carleton**  
[Review. Kuru and its contribution to medicine.](#)  
*Philosophical transactions of the Royal Society of London. Series B, Biological sciences*. 2008 November 27; 363 (1510) : 3697-3700

- The solution of kuru led us to the solution of **Creutzfeldt-Jakob disease** and to the elucidation...

**Gillette-Guyonnet Sophie, Vellas Bruno**  
[Caloric restriction and brain function.](#)  
*Current opinion in clinical nutrition and metabolic care*. 2008 November 01; 11 (6) : 686-92

- ...neurodegenerative disorders such as **Alzheimer's disease**? Is there a future for caloric restriction interventions...
- ...may also prevent beta-amyloid neuropathology in **Alzheimer** transgenic models. Finally, both exercise...

**Collinge John, Whitfield Jerome, McKintosh Edward, Frosh Adam, Mead Simon, Hill Andrew F, Brandner Sebastian, Thomas Dafydd, Alpers Michael P**  
[A clinical study of kuru patients with long incubation periods at the end of the epidemic in Papua New Guinea.](#)  
*Philosophical transactions of the Royal Society of London. Series B, Biological sciences*. 2008 November 27; 363 (1510) : 3725-3739

- ...in the 1950s, the arrival of variant **Creutzfeldt-Jakob disease (vCJD)**, also thought to be transmitted...

**van der Zee Julie, Slegers Kristel, Van Broeckhoven Christine**  
[Invited article: the Alzheimer disease-frontotemporal lobar degeneration spectrum.](#)  
*Neurology*. 2008 October 07; 71 (15) : 1191-7

- **Alzheimer disease (AD)** and **frontotemporal lobar degeneration (FTLD)** are two frequent forms...
- ...of primary neurodegenerative **dementias**. Despite distinctive clinical diagnostic criteria...
- ...more about brain pathology and genetic makeup underlying these **dementia** disorders, evidence...



# Architektur eines sprachtechnologischen Systems: Textanalytik

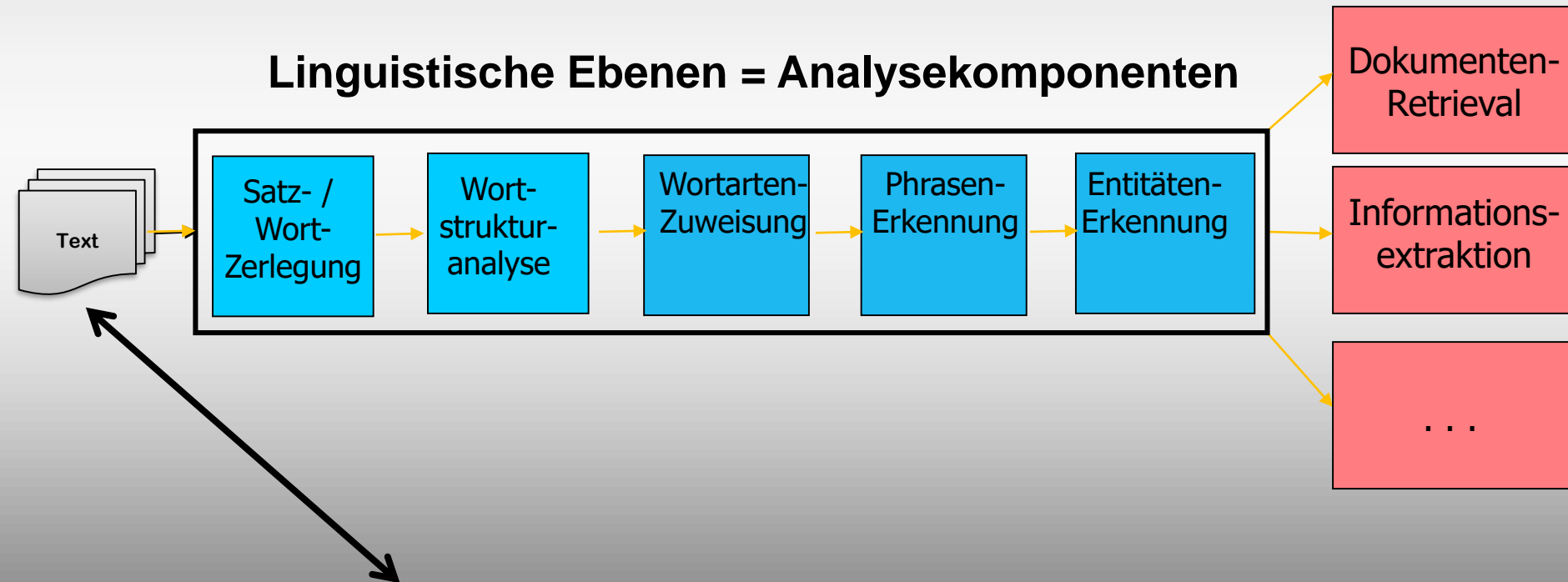


Unstructured Information Management Architecture

*An Apache Incubator Project*



## Linguistische Ebenen = Analysekomponenten



... may explain impaired transcriptional responses in leishmania-infected cells

# Architektur eines sprachtechnologischen Systems: Textanalytik

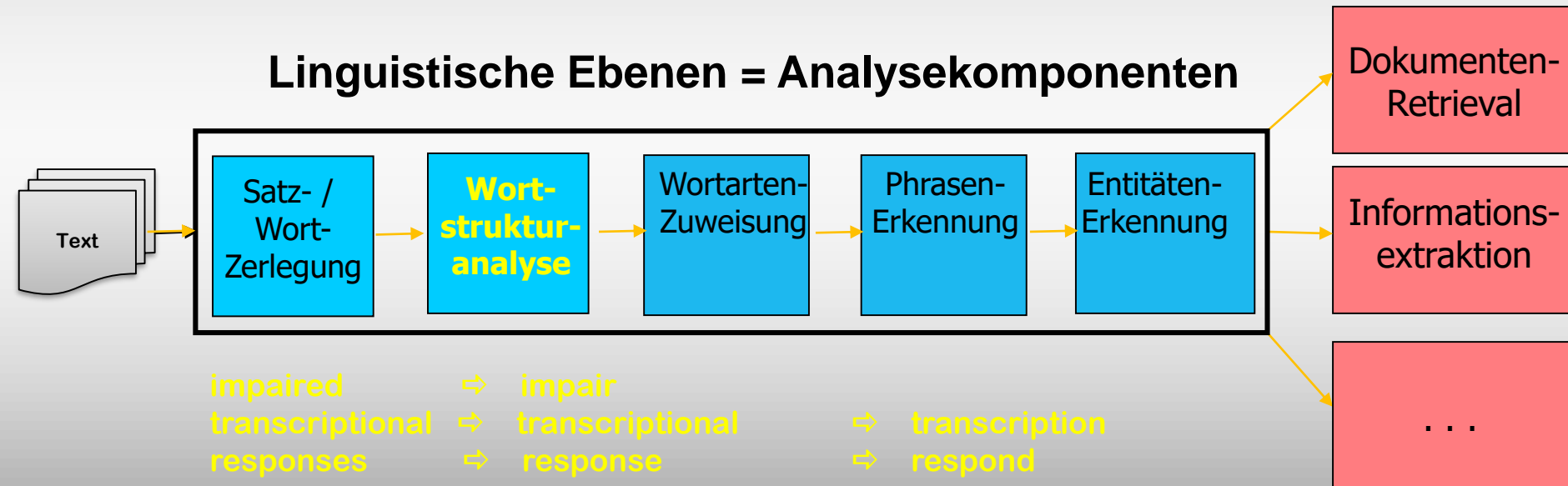


Unstructured Information Management Architecture

*An Apache Incubator Project.*



## Linguistische Ebenen = Analysekomponenten



... may explain impaired transcriptional responses in leishmania-infected cells

# Architektur eines sprachtechnologischen Systems: Textanalytik

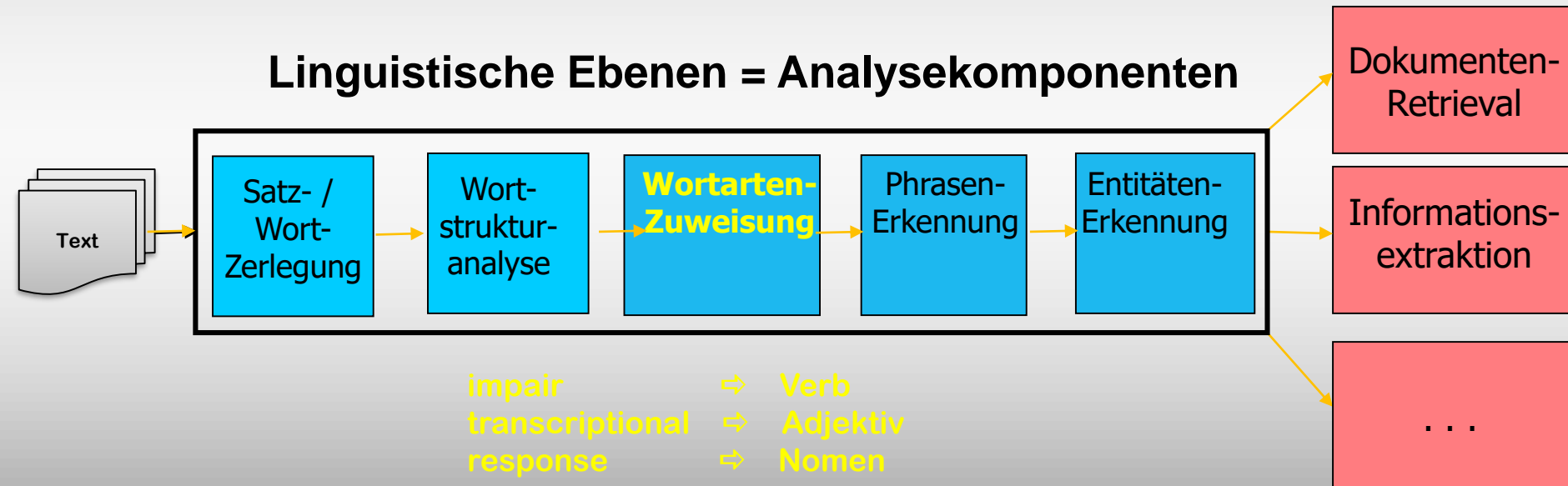


Unstructured Information Management Architecture

*An Apache Incubator Project*



## Linguistische Ebenen = Analysekomponenten



... may explain impaired transcriptional responses in leishmania-infected cells

# Architektur eines sprachtechnologischen Systems: Textanalytik

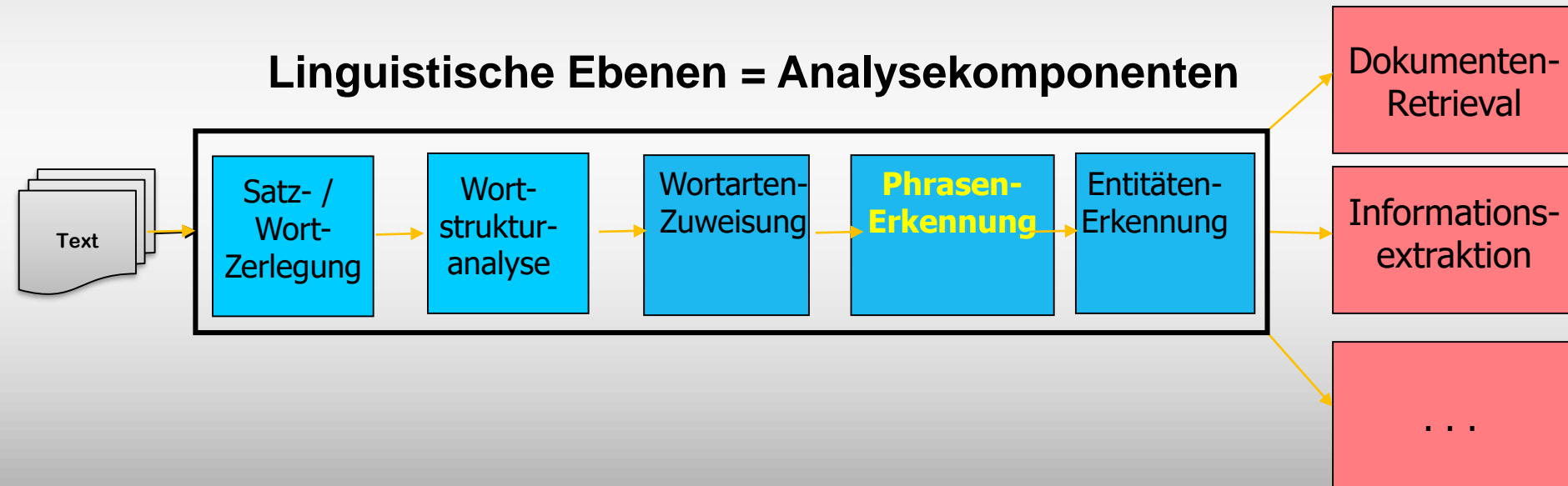


Unstructured Information Management Architecture

An Apache Incubator Project



## Linguistische Ebenen = Analysekomponenten



... may explain impaired transcriptional responses in leishmania-infected cells  
 ... MODAL VERB PARTIZIP ADJEKTIV NOMEN PRÄP PARTIZIP NOMEN  
 [ may explain ]<sub>VP</sub> [ impaired transcriptional responses ]<sub>NP</sub> [ in leishmania-infected cells ]<sub>PP</sub> ]<sub>NP</sub>

# Architektur eines sprachtechnologischen Systems: Textanalytik

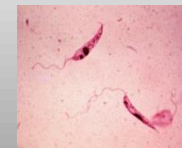
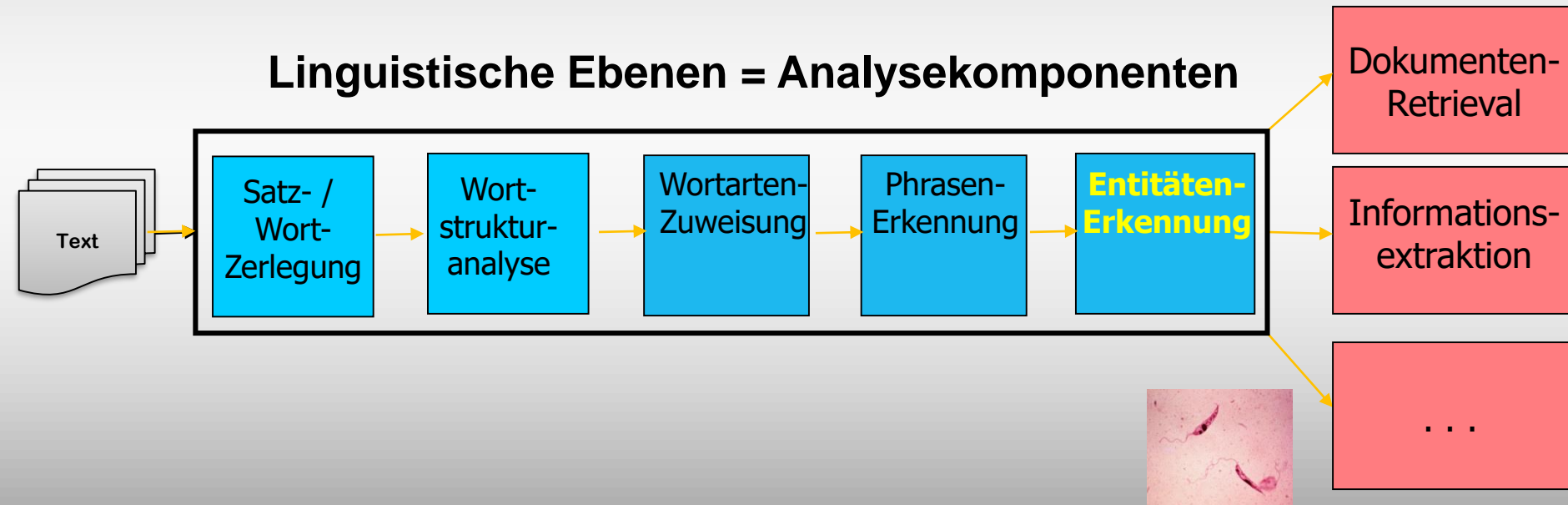


Unstructured Information Management Architecture

*An Apache Incubator Project*



## Linguistische Ebenen = Analysekomponenten



... may explain impaired transcriptional responses in leishmania-infected cells

[ ... ]<sub>VP</sub> [ [ impaired transcriptional responses ]<sub>NP</sub> [ in leishmania-infected cells ]<sub>PP</sub> ]<sub>NP</sub>

⇒ Negative\_Regulation [ako PPI] in cells with a disease

# Fazit: Semedico

## Suchmaschine

- Ermöglicht einen tieferen semantisch-ontologischen Zugang zu Medline-Texten
- Derzeit sind > 28 Millionen Medline-Dokumente (Abstracts) und 1,8 Mio. Volltexte prozessiert und indexiert
- Nutzerstudie an der Medizinischen Hochschule Hannover (Proc. SMC, 2007)
- <http://www.semedico.org>
- E.Faessler & U. Hahn (2017), ACL Demos.
- Youtube-Demo:  
<http://www.youtube.com/watch?v=f7lIPqGh-sc>

# Natürliche Sprache

## Ambiguität: lexikalisch-semantic

- Homografie, Polysemie
  - Konstanz liegt am Bodensee.
  - Bei Konstanz des Luftdrucks ...
  - I saw that gasoline can explode
    - [Ich sah diesen Benzinbehälter explodieren]
    - [Ich sah, dass Benzin explodieren kann]

# Natürliche Sprache

## Ambiguität: syntaktisch

- Skopus
  - die **alten** Männer und Frauen
    - die alten Männer und **[allgemein alle]** Frauen
    - die alten Männer und **alten** Frauen
- PP-Anbindung
  - Sie sahen den Mann mit dem Fernrohr
    - **Sie** sahen den Mann **mit Hilfe ihres**<sub>INSTRUM</sub> Fernrohrs
    - Sie sahen **den Mann**, der **sein**<sub>POSSESS</sub> Fernrohr **trug**



# Natürliche Sprache

## Ambiguität: syntaktisch

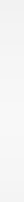
- PP-Anbindung & Homografie/Polysemie
  - They saw the man with the telescope.

# Natürliche Sprache

## Ambiguität: syntaktisch

- PP-Anbindung & Homografie/Polysemie
  - They saw the man with the telescope.

- Sie sahen den Mann mit Hilfe ihres<sub>INSTRUM</sub> Fernrohrs.



# Natürliche Sprache

## Ambiguität: syntaktisch

- PP-Anbindung & Homografie/Polysemie
  - They saw the man with the telescope.

- Sie sahen den Mann mit Hilfe ihres<sub>INSTRUM</sub> Fernrohrs.

- Sie sahen den Mann, der sein<sub>POSSESS</sub> Fernrohr trug.



# Natürliche Sprache

## Ambiguität: syntaktisch

- PP-Anbindung & Homografie/Polysemie

- They saw the man with the telescope.

- Sie sahen den Mann mit Hilfe ihres<sub>INSTRUM</sub> Fernrohrs.

- Sie sahen den Mann, der sein<sub>POSSESS</sub> Fernrohr trug.

- Sie zersägen den Mann mit Hilfe ihres<sub>INSTRUM</sub> Fernrohrs.

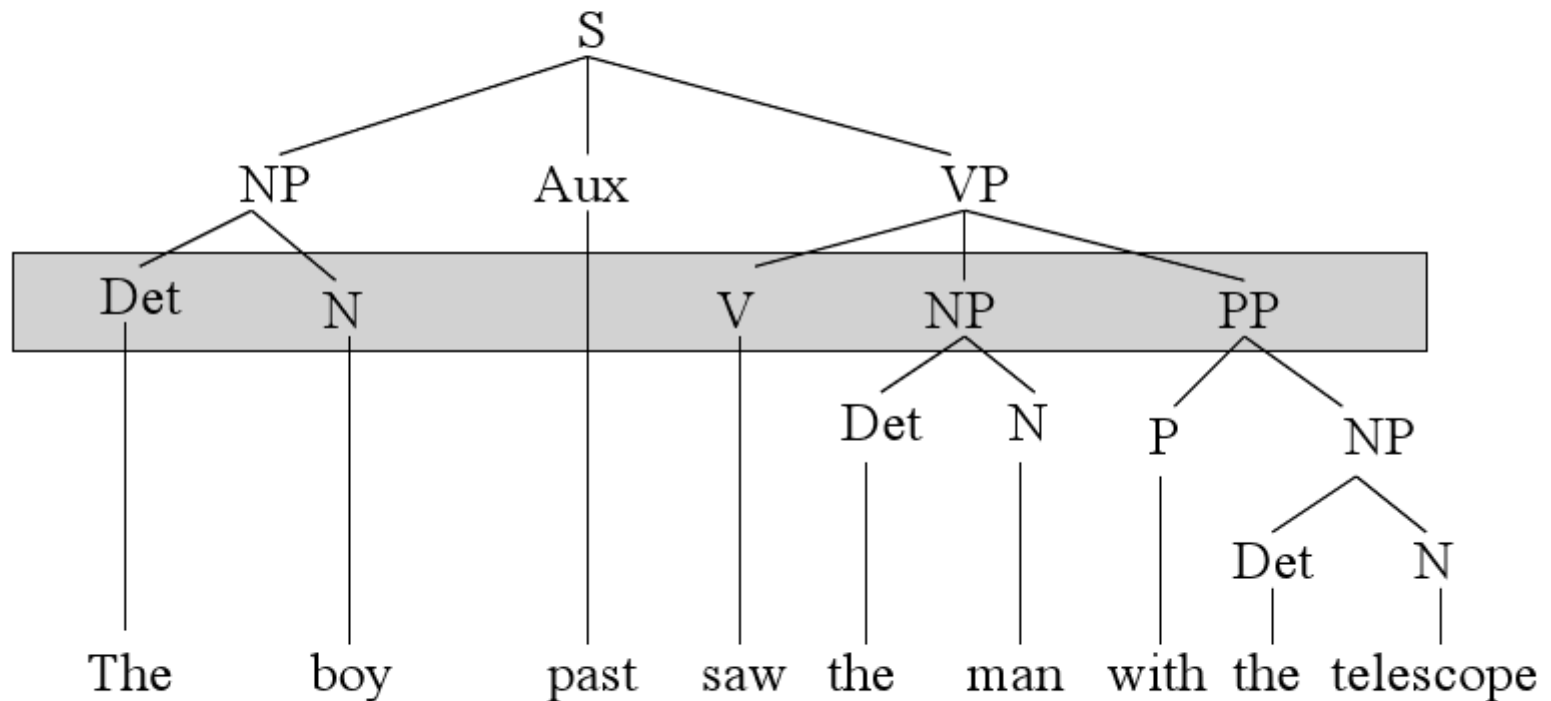
- Sie zersägen den Mann, der sein<sub>POSSESS</sub> Fernrohr trug.



# Structural Ambiguity (1)

The boy saw the man with the telescope

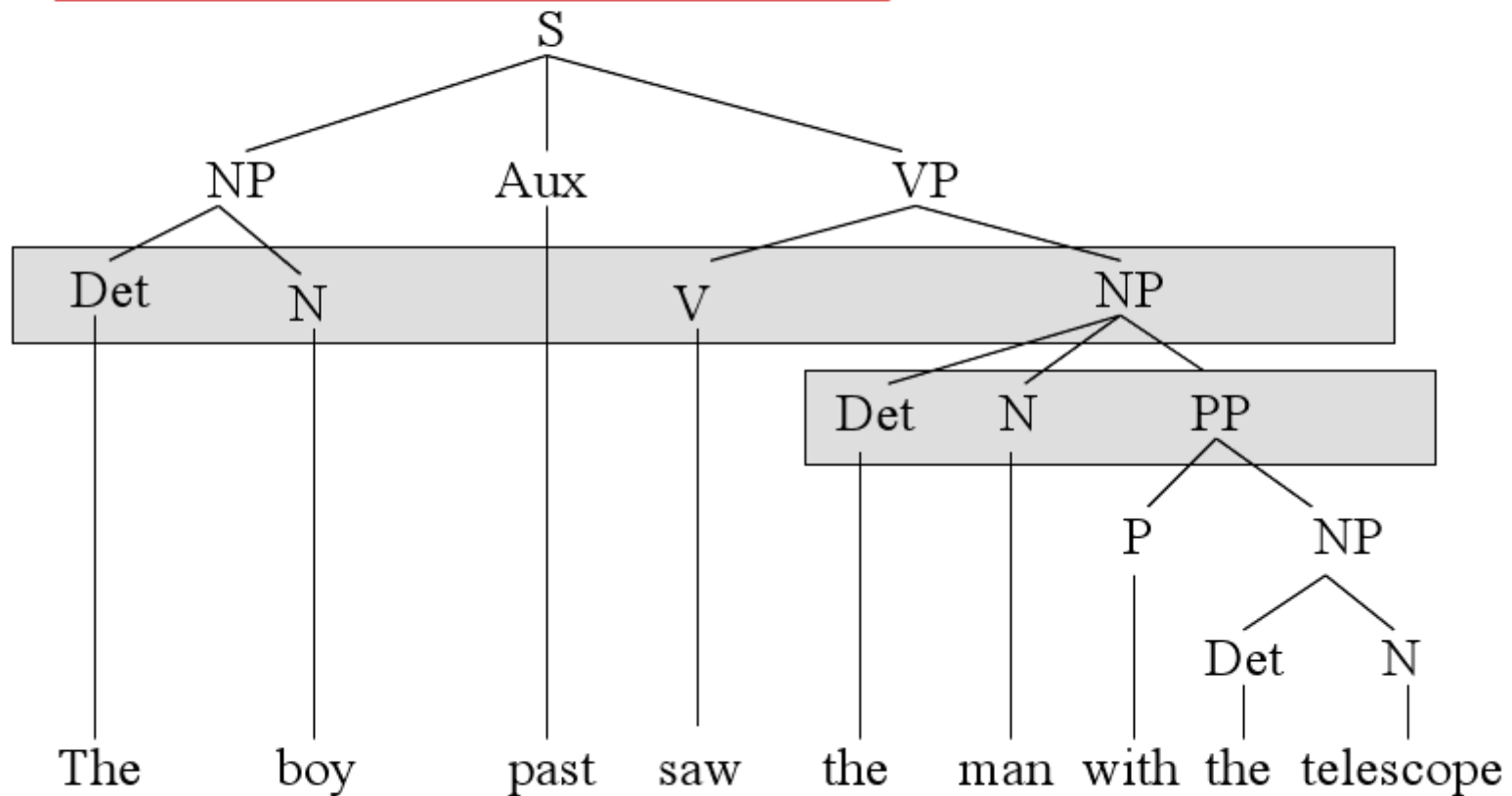
---



## Structural Ambiguity (2)

The boy saw the man with the telescope

---



# Natürliche Sprache

## Ambiguität: syntaktisch

- Wachstumsverhalten v. PP-Anbindungen
  - Die Tarifparteien haben sich [auf der Basis eines Kompromissentwurfs des neutralen Schlichters]<sub>1</sub> [am frühen Mittwoch Morgen]<sub>2</sub> [in Bad Nauheim]<sub>3</sub> [nach einem 16-stündigen Verhandlungsmarathon]<sub>4</sub> [auf einen Tarifabschluss]<sub>5</sub> [in Höhe]<sub>6</sub> [von 1,2 Prozent]<sub>7</sub> [für die 160.000 Beschäftigten]<sub>8</sub> [in der hessischen Bekleidungsindustrie]<sub>9</sub> geeinigt.

# Natürliche Sprache

## Ambiguität: syntaktisch

- Wachstumsverhalten v. PP-Anbindungen

- Die Tarifparteien haben sich [auf der Basis eines Kompromissentwurfs des neutralen

Schlicht

[in Bao

Verhar

Tarifak

[für die

hessisc

$$\begin{aligned} CAT_n &= \binom{2n}{n} - \binom{2n}{n-1} \\ &= \frac{(2n)!}{n! \times (2n-n)!} - \frac{(2n)!}{(n-1)! \times (2n-(n-1))!} \end{aligned}$$



# Natürliche Sprache

Ambiguität: syntaktisch

- Wachstumsverhalten

- Die Tarifparteien behaupten

eines Kompromisses

Schlicht

[in Baden

Verhar

Tarifak

[für die

hessisc

$$CAT_{n=9-1} = 1430$$

$$\begin{aligned} CAT_n &= \binom{2n}{n} - \binom{2n}{n-1} \\ &= \frac{(2n)!}{n! \times (2n-n)!} - \frac{(2n)!}{(n-1)! \times (2n-(n-1))!} \end{aligned}$$

**Binärbäume !!**

# Natürliche Sprache

## Ambiguität: semantisch

- Quantoren-Skopus
  - Welcher Mitarbeiter verdient mehr als **ein** Abteilungsleiter?
    - **Werner Ceusters**
      - Bezugsmaßstab sind **alle** Abteilungsleiter (globales Max)
    - **Werner Ceusters und Barry Smith**
      - Angestellte aus der Abteilung ihres jeweiligen Abteilungsleiters, die mehr als **dieser** Abteilungsleiter verdienen (lokales Max)
    - **Werner Ceusters, Barry Smith, Peter Fripp, ...**
      - Bezugsmaßstab ist **irgendein** Abteilungsleiter

# Natürliche Sprache

## Ambiguität: pragmatisch

- Kann ich mit **remove** **<file.txt>** die Datei **<file.txt>** löschen?
  - Nein!
    - **remove** ist ein falscher Kommandobezeichner
    - **remove** ist zwar richtiger Kommandobezeichner, aber der Benutzer hat keine Löschberechtigung
    - es existiert keine Datei mit dem Namen **<file.txt>**

# Dimensionen zur Beschreibung natürlicher Sprache

- **Sprachspezifisches Wissen (I)**
  - **Phonologie/Graphematik**
    - Laut- und Schriftzeichensystem
  - **Lexikologie**
    - Lexemsystem (Wörter, Kollokationen)
  - **Morphologie**
    - Flexion: Kombination des Lexemsystems mit grammatischen Informationen (Genus, Numerus, Kasus, Tempus, Modus)
    - Derivation, Komposition: lexikalisch-  
semantische Bedeutungsverschiebung durch  
Verknüpfung inhaltstragender Elemente

# Dimensionen zur Beschreibung natürlicher Sprache

- **Sprachspezifisches Wissen (II)**
  - **Syntax**
    - Prinzipien der linearen Reihung in Phrasen und Sätzen (Grammatikalität)
  - **Semantik**
    - Lexikalische Semantik
    - Kompositionale Satzsemantik (literale Lesarten) und Figürlichkeit (Metonymie, Metapher)
  - **Pragmatik (Diskurs, Text)**
    - Textsemantik (Textualität): Kohäsion, Kohärenz
    - Situativer, intentionaler usw. Kontext

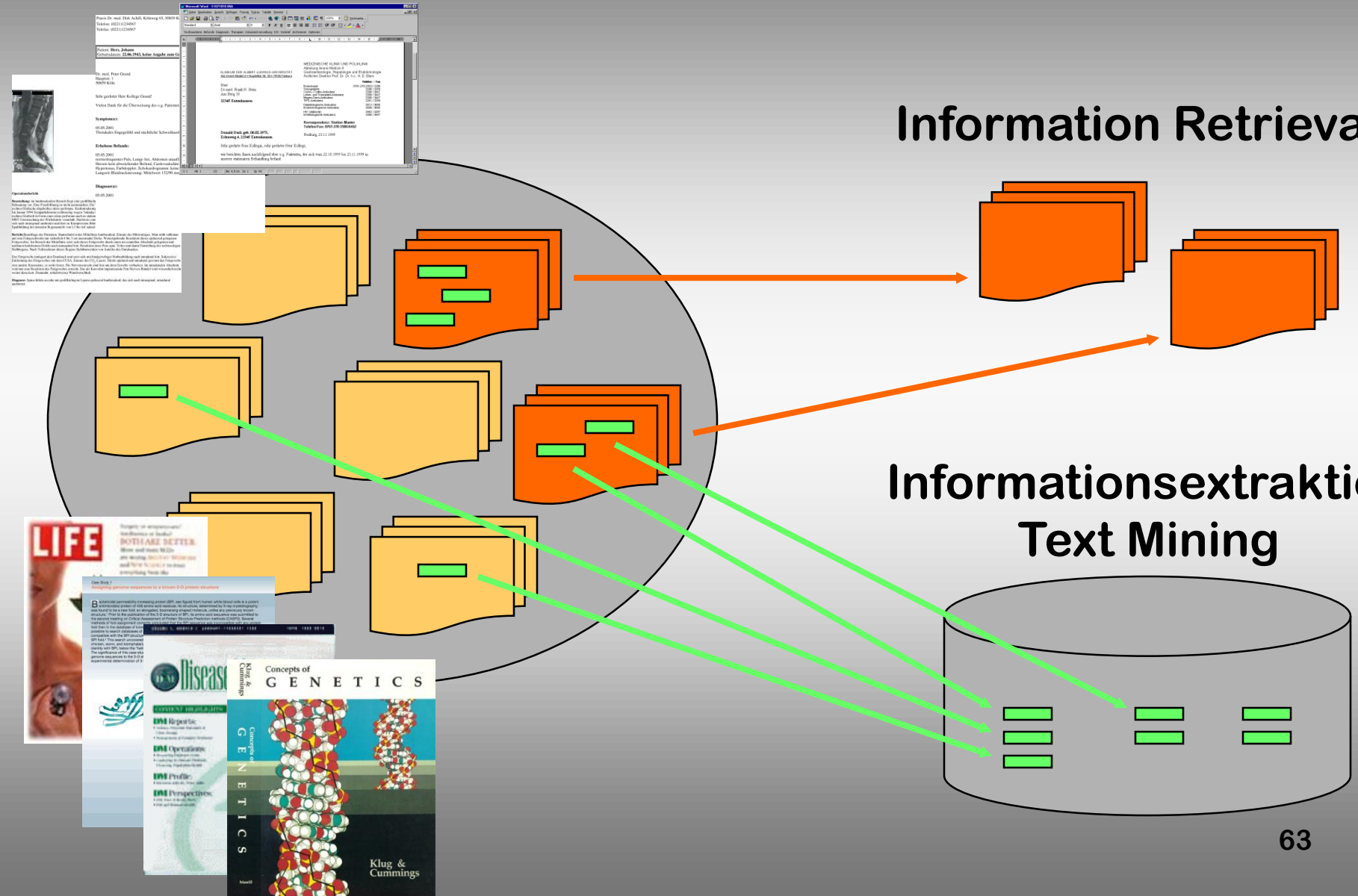
# Dimensionen zur Beschreibung natürlicher Sprache

- Funktionaler Sprachgebrauch
  - Suche nach relevanten Texten
  - Faktenextraktion aus Texten
  - Textzusammenfassung
  - Übersetzung
  - Frage-Beantwortung

# Information Retrieval & Extraktion

Information Retrieval

Informationsextraktion  
Text Mining



# Indexing

Thalidomide was found to be highly effective in managing the cutaneous manifestations of leprosy (erythema nodosum leprosum) and even to be superior to aspirin (acetylsalicylic acid) in controlling leprosy-associated fever

Thalidomide

leprosy

Aspirin



# Informationsextraktion

Thalidomide was found to be highly effective in managing the cutaneous manifestations of leprosy (erythema nodosum leprosum) and even to be superior to aspirin (acetylsalicylic acid) in controlling leprosy-associated fever

Disease: leprosy

Drug: Thalidomide

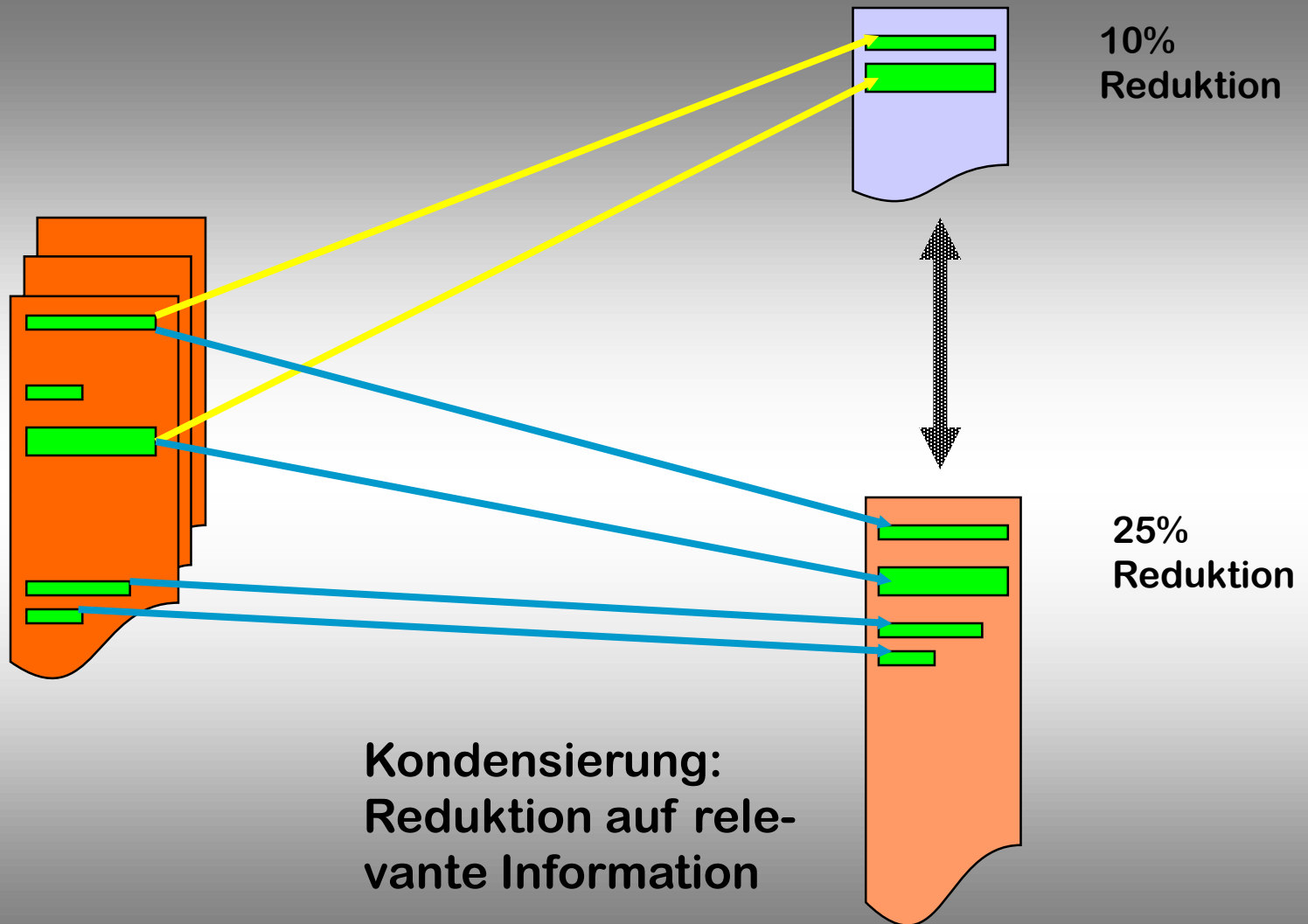
Effective-for: Thalidomide, cutaneous manifestations of leprosy

Disease: leprosy-associated fever

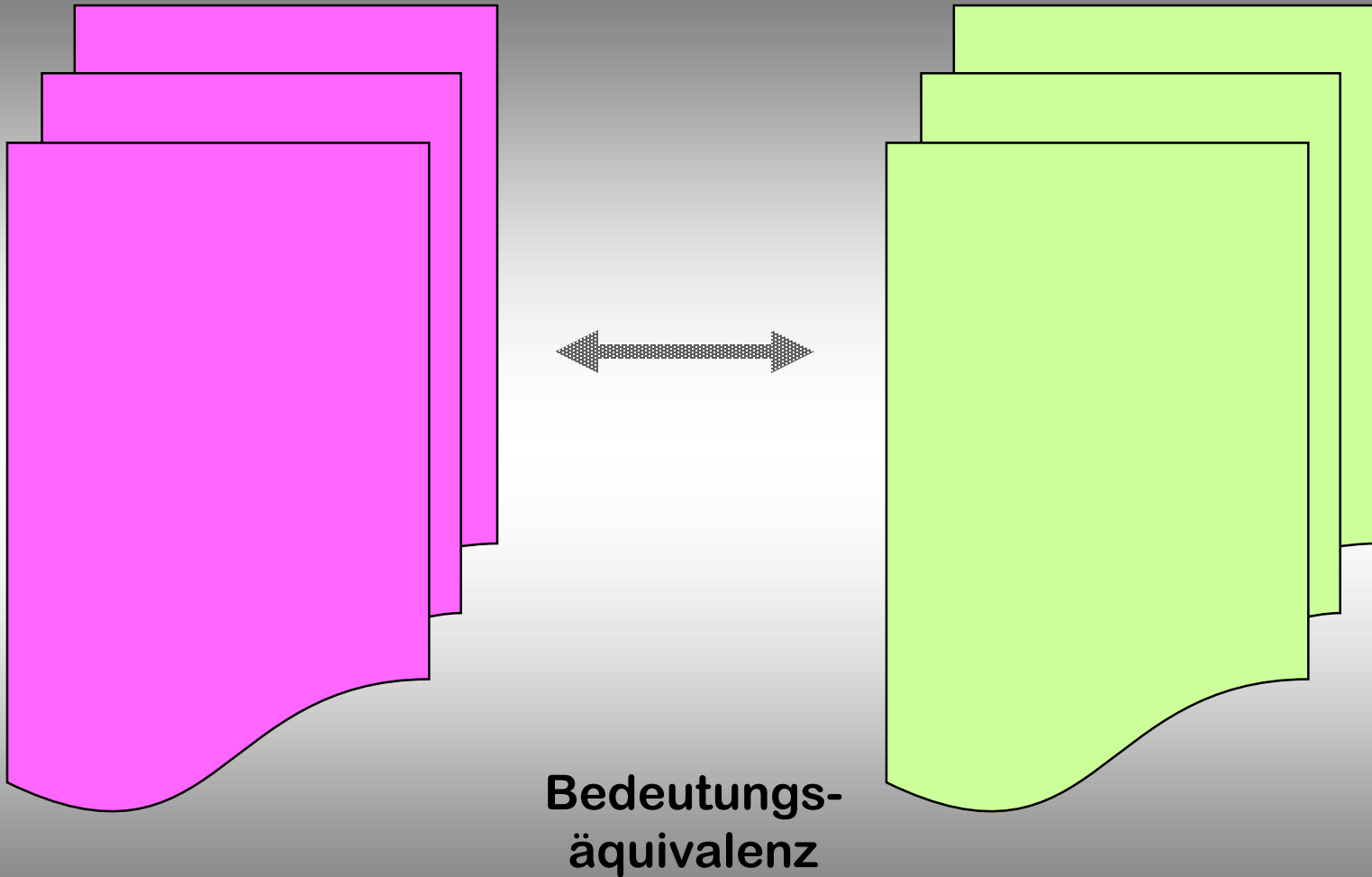
Drug: Thalidomide, Aspirin

Effective-for: [ Thalidomide > Aspirin ], leprosy-associated fever

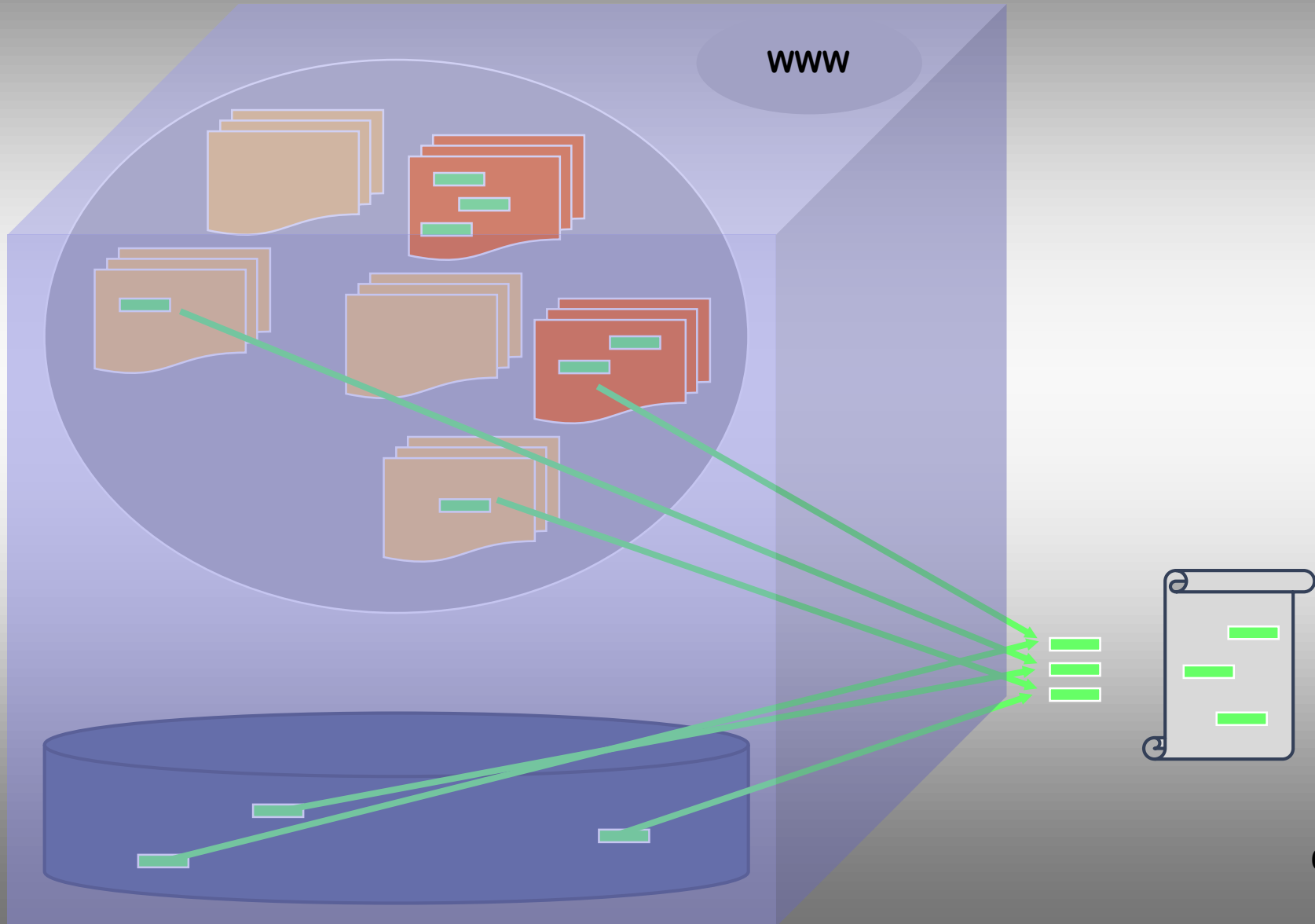
# Textzusammenfassung



# Maschinelle Übersetzung



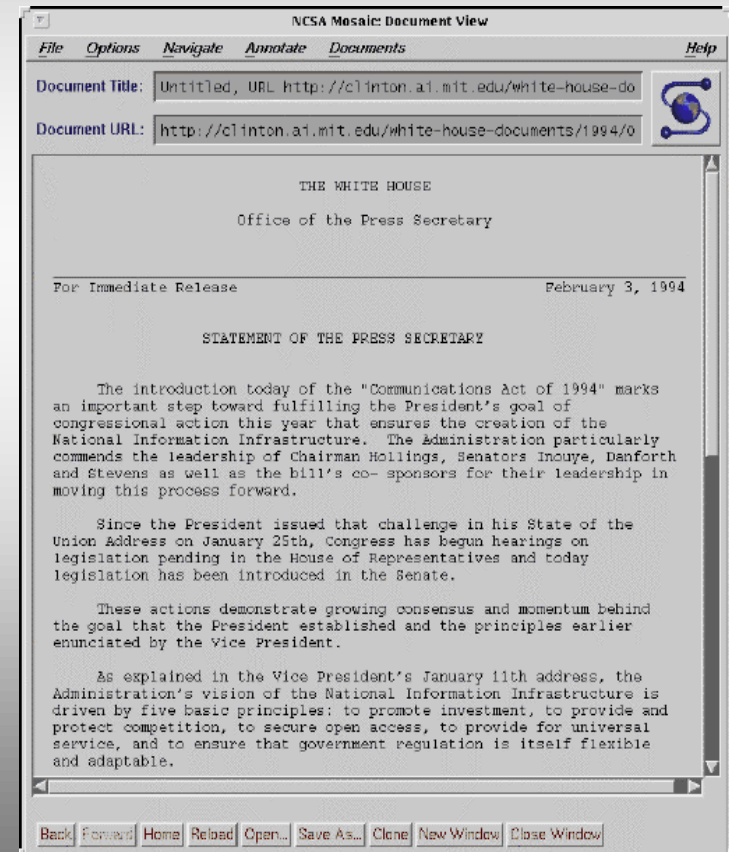
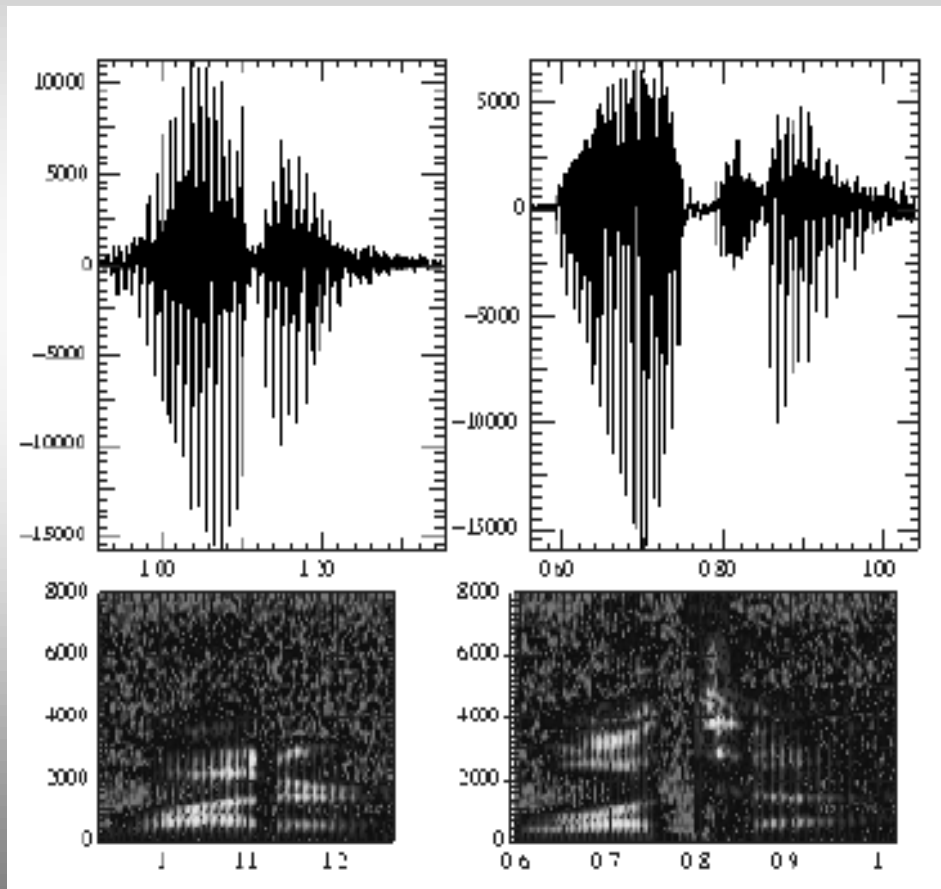
# Frage-Beantwortung



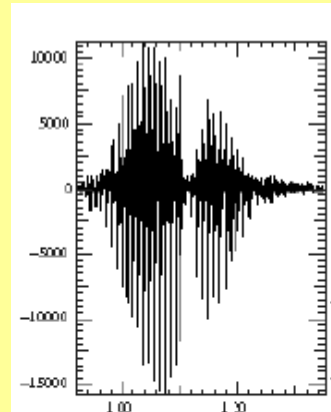
# Die zwei Modi natürlicher Sprache

... in *gesprochener* Form  
(Spontansprache)

... in *geschriebener* Form  
(Dokumente, Texte)



# Speech2Text



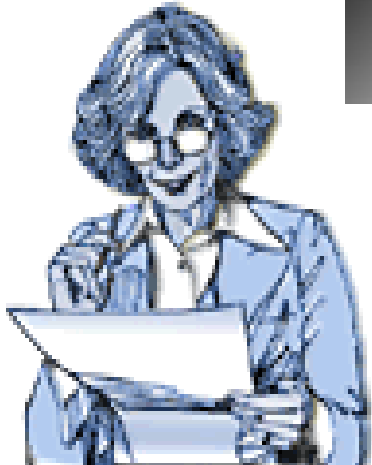
Speech Signal Detection  
(Waveform spectrogram)

Signal Analysis:  
Acoustic, Lexical &  
Language Models for  
phonetic/prosodic  
decoding

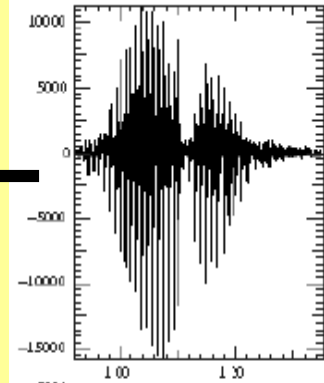
NLP  
Application

Signal Processing:  
Digitization, Noise  
Filtering, Vector  
Quantization

# Text2Speech



Speech Signal Production  
(Waveform spectrogram)



Signal Processing:  
Digitization, Speech  
Signal Synthesis:  
duration, amplitude,  
spectrum generation

Signal Synthesis:  
Acoustic, Lexical &  
Language Models:  
phonetic/prosodic  
encoding

NLP  
Application

# Natürlichsprachliche Systeme

- implementieren sprachliche Verstehensleistungen (Text- oder Dialogmodus)
- zeigen die Emergenz sprachlichen und außersprachlichen Wissens an komplexen kognitiven Leistungen
  - isolierte Betrachtungsweise generiert häufig unproduktive Zuordnungsdebatten (Syntax/Semantik)
- sind stets funktional orientiert (Informationsgewinnung, Sprachtransfer usw.)
- sind die Grundlage für Sprachtechnologie



# Alternative Explananda linguistischer Modellierung

- Theoretische Linguistik
  - Natürliche Sprache(n) als **Kompetenz**system(e):  
Rekonstruktion d. **idealen Sprachsystems/Sprechers**
  - Natürliche Sprache(n) als **Performanz**system(e):  
**Korpuslinguistik: Empirie des Sprachgebrauchs**
- Computerlinguistik (e. Angewandte Linguistik)
  - Natürliche Sprache(n) als **Performanz**system(e):  
Rekonstruktion **funktionalen Sprachgebrauchs** aus  
**sprachtechnologischer** Perspektive (Automaten)
- Psycholinguistik (e. a. Angewandte Linguistik)
  - Natürliche Sprache(n) als **Performanz**system(e):  
Rekonstruktion **menschlichen Sprachgebrauchs** aus  
**kognitiver Simulation**sperspektive (Primaten)

# Meilensteine der Computerlinguistik

1950 – 1960	<b>Theorie formaler Sprachen</b> / Automatentheorie vs. Struktur natürlicher Sprachen ( <i>Noam Chomsky</i> ) Anwendungen: <b>autom. Indexing, masch. Übersetzung</b>
1966	ALPAC-Report
1966	Der erste <b>Chatbot</b> : ELIZA ( <i>J. Weizenbaum</i> )
2017	Der aktuellste <b>Chatbot</b> : Replika

# Meilensteine der

Col

1950 – 1960

1966

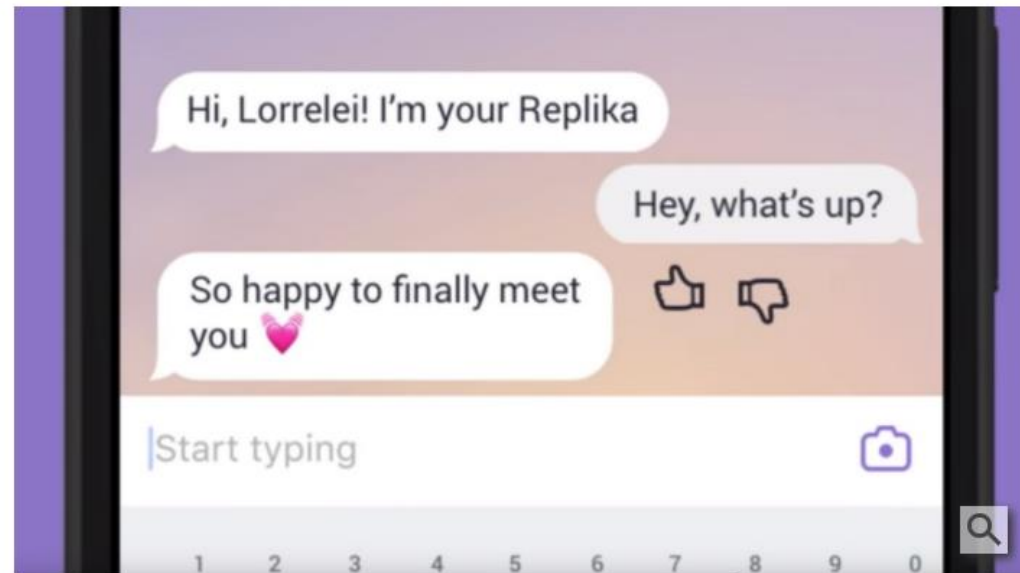
1966

2017

Theor  
Strukt  
Anwe  
ALPA  
Der er  
Der al

21. November 2017, 18:57 Uhr Replika

## Diese App redet mit dem Nutzer wie mit einem engen Freund



Replika schaut aus wie ein normaler Chatbot in einem Messenger. Die Fragen und Antworten werden aber schnell sehr persönlich. (Foto: Screenshot / Replika)

- Die Smartphone-App Replika will eine Beziehung zu seinem Nutzer aufbauen.
- Neuronale Netzwerke sollen das Verhalten von Menschen imitieren.
- Replika stellt persönliche Fragen und übernimmt die Ausdrucksweise des Nutzers.
- Die Entwicklerin wollte einen Computer erschaffen, "der schön genug ist, dass eine Seele darin wohnen wollen würde".

# Meilensteine der Computerlinguistik

- 1950 – 1960    **Theorie formaler Sprachen** / Automatentheorie vs. Struktur natürlicher Sprachen (*Noam Chomsky*)  
Anwendungen: **autom. Indexing, masch. Übersetzung**
- 1966    ALPAC-Report
- 1966    Der erste **Chatbot**: ELIZA (*J. Weizenbaum*)
- 1965 – 1970    **Syntax I**: Generative Transform.grammatik (*N. Chomsky*)  
vs. Augmented Transition Networks (*William Woods*)  
**Frage-Antwort-Systeme**: LUNAR (*William Woods*)