

Computerlinguistik I

Vorlesung im WiSe 2018/19
(M-GSW-09)

Prof. Dr. Udo Hahn

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

<http://www.julielab.de>

Allgemeine Hinweise

- Vorlesung: Do, 10-12h (Fürstengrb.1, SR 275)
- Übung zV: Mo, 8-10h (Fürstengrb.1, SR 275)
 - beginnt am **22.10.**
- Vorlesungsmaterialien im Netz
 - <http://www.julielab.de/> ⇒ „Students“
- **M-GSW-09 besteht aus VL+ÜB und Seminar!**
- Sprechstunde: Mi, 12-13h, bA (FG 30, 004)
- Email: udo.hahn@uni-jena.de
- URL: <http://www.julielab.de>
- Fachliteratur ist überwiegend in Englisch

Bitte ...

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



Institut für Germanistische Sprachwissenschaft der FSU Jena

- **Lehrstuhl für Theoretische Linguistik – Grammatiktheorie**
 - Prof. Dr. Peter Gallmann – 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: (1/5) Graduiertenkolleg **Modell ‚Romantik‘ [Digital Humanities]**
 - Bundesministerium für Bildung & Forschung (BMBF)
 - Aktuell: (1/7/26) 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 M-GSW-09
 - Machine Reading – WWW-skalierbares automatisches Textverstehen
 - Do, 16-18, Fürstengraben 1, SR 164

Kleiner Exkurs zum Thema “Wissenschaftliche Exzellenz”

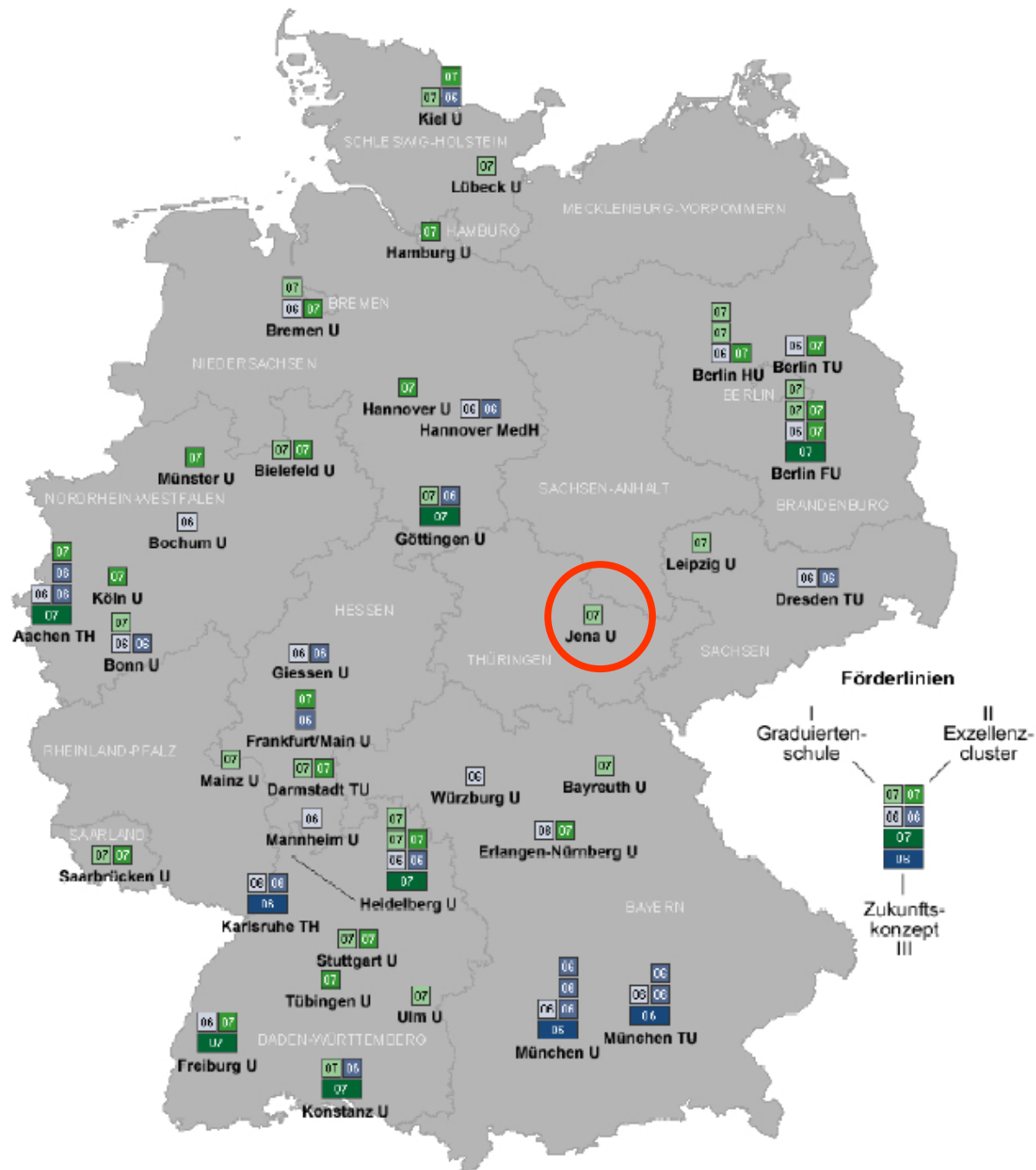
1. Exzellenzinitiative (2006-07)

87 deutsche Universitäten

44 in Förderlinien

FSU Jena: 1 Graduiertenschule:
Jena School for
Microbial Communication

9 Elite-Universitäten (I)
(FUB, FR, GÖ, HD, KA, KN,
MUM, TUM, RWTH AC)



2. Exzellenzinitiative (2010-12)

87 deutsche Universitäten

11+45+43 = 89 in Förderlinien

FSU Jena: 1 Graduiertenschule:
Jena School for
Microbial Communication

9 Elite-Universitäten (I)
(FUB, FR, GÖ, HD, KA, KN,
MUM, RWTH AA, TUM)

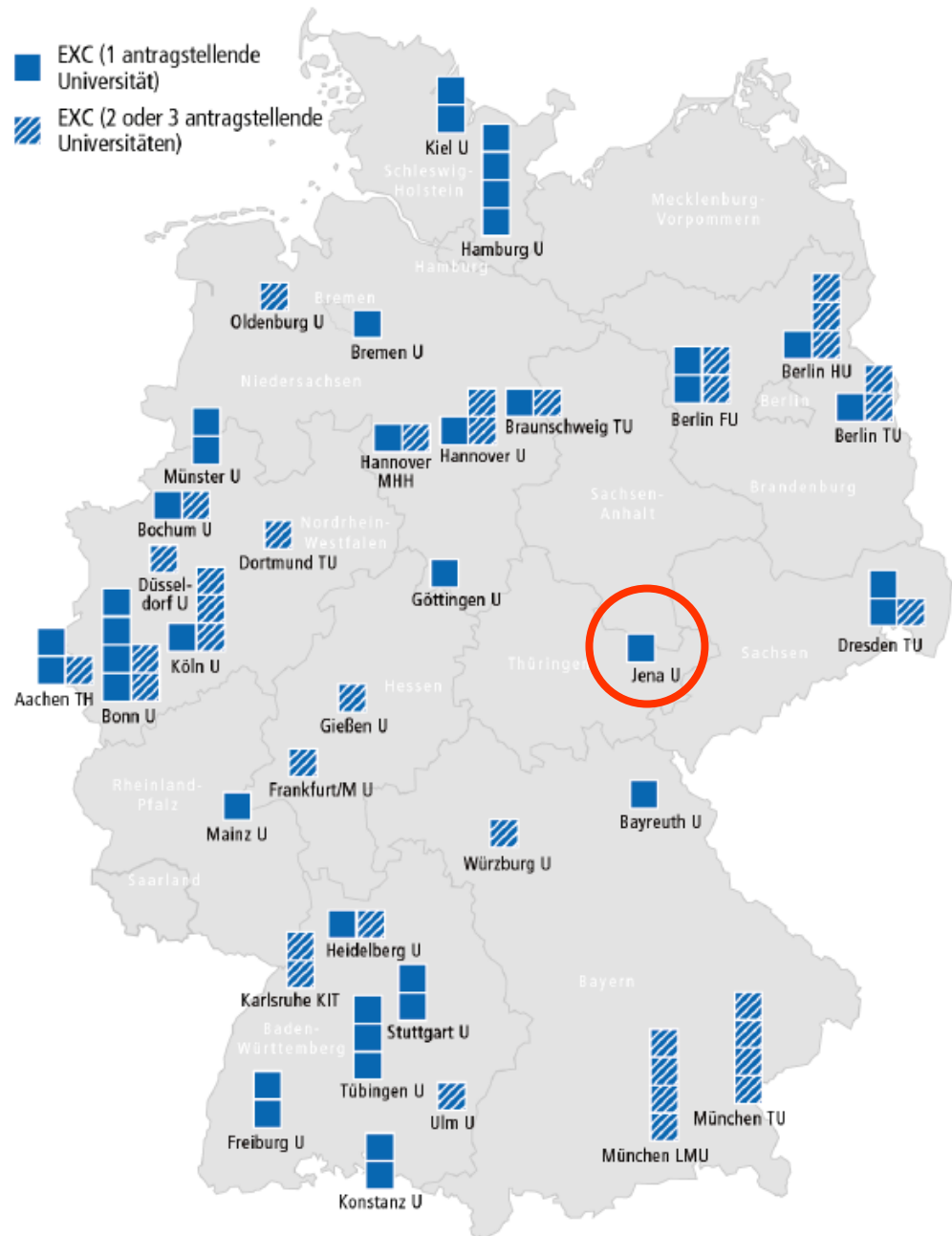
11 Elite-Universitäten (II)
(TUDD, FUB, HB, HUB, HD, K, KN,
MUM, TUM, RWTH AC, TÜ)



3. Exzellenzinitiative (2017-22)

Förderung der neuen Exzellenzcluster (EXC) ab 1. Januar 2019

Entscheidung der Exzellenzkommission vom 27. September 2018



Woher kommt Exzellenz ?

- (High-impact-)Publikationen
 - Wissenschaftspreise
 - Drittmiteleinwerbungen
 - SFBs, Graduiertenschulen ...
 - Zukunftsentwürfe
-
- Im internationalen Kontext weltweit sichtbar sein (visibility)

Ein Beispiel für den Nachweis wissenschaftlicher Exzellenz

- **Semantik**
 - Bedeutung von Sprache
- **Semantische Textanalytik**
 - Inhaltliche Analyse von Texten
- **Informationsbeschaffung für Biologen und Mediziner**
 - Medline/PubMed: mehr als 27M Dokumente
- **„Weltmeisterschaft“ für semantische Textanalytik**
 - Wo ist Jena (JULIE Lab) ?

<Semantische Textanalytik>

- **Natürlichsprachliche Semantik**
 - Lexikalische Semantik, Satzsemantik
- **Term-Semantik**
 - Termvarianten: Synonyme, Akronyme, Abkürzungen
- **Typen-Semantik**
 - Generalisierung auf Klassen
- **Propositionale Semantik**
 - Prädikationen : $p(a_1, \dots, a_n)$, a_i kann Term sein, aber auch eine Prädikation

Propositionale Semantik

Annotation Results for 7591091.xmi in C:\Users\jwermt\ Desktop\HahnBOOTStrip

Attenuation of **gamma interferon**-induced tyrosine phosphorylation in mononuclear phagocytes infected with *Leishmania donovani*: selective inhibition of signaling through Janus kinases and **Stat1**.

The induction of gene transcription in response to **gamma interferon** is impaired in mononuclear phagocytes infected with *Leishmania donovani*, and the mechanisms involved are not fully understood. The changes in gene expression brought about by **gamma interferon** are thought to **involve** transient **increases** in the activities of cellular protein tyrosine kinases, including the Janus kinases **Jak1** and **Jak2**, **leading** to tyrosine phosphorylation of the transcription factor **Stat1**. To investigate the mechanisms accounting for the impaired responses to **gamma interferon**, a model system for examining overall changes in protein tyrosine phosphorylation, **activation** of **Jak1** and **Jak2** and **phosphorylation** of **Stat1** was developed in phorbol 12-myristate 13-acetate-differentiated U-937 cells. Analysis of whole-cell lysates by antiphosphotyrosine immunoblotting showed that incubation with **gamma interferon** brought about specific increases in phosphotyrosine labeling of several proteins. Increased labeling of these proteins occurred to similar extents in control cells and in cells that had been infected with *L. donovani* for 16 h. **Jak1**, **Jak2**, and **Stat1** were **immunoprecipitated** from control and interferon-treated cells, and tyrosine phosphorylation of these proteins, detected by antiphosphotyrosine immunoblotting was used to measure their activation. Tyrosine **phosphorylation** of **Jak1**, **Jak2**, and **Stat1** **increased** markedly, in a dose-dependent manner, in U-937 cells incubated with **gamma interferon**. In contrast, in cells infected with *L. donovani*, tyrosine **phosphorylation** of **Jak1**, **Jak2**, and **Stat1** was markedly **impaired**. This effect was dependent upon the duration of exposure to *L. donovani* and was maximal and complete at 16 h. Results similar to those observed with U-937 cells were also obtained with human peripheral blood monocytes. These findings indicate that infection of human mononuclear phagocytes with *L. donovani* leads to **impaired gamma interferon-mediated** tyrosine **phosphorylation** and selective effects on the **Jak-Stat1** pathway. Unresponsiveness to **gamma interferon** for activation of this pathway may explain impaired transcriptional responses in *leishmania*-infected cells.

Legend

☒ EventMention

☒ Gene

Propositionale Semantik

Annotation Results for 7591091.xml in C:\Users\jwermt\Desktop\HahnBOOTStrip

Attenuation of **gamma interferon**-induced tyrosine phosphorylation in mononuclear phagocytes infected with *Leishmania donovani*: selective inhibition of signaling through Janus kinases and **Stat1**. The induction of gene transcription in response to **gamma interferon** is impaired in mononuclear phagocytes infected with *Leishmania donovani*, and the mechanisms involved are not fully understood. The changes in gene expression brought about by **gamma interferon** are thought to **involve** transient **increases** in the activities of cellular protein tyrosine kinases, including the Janus kinases **Jak1** and **Jak2**, **leading** to tyrosine phosphorylation of the transcription factor **Stat1**. To investigate the mechanisms accounting for the impaired responses to **gamma interferon**, a model system for examining overall changes in protein tyrosine phosphorylation, **activation** of **Jak1** and **Jak2** and **phosphorylation** of **Stat1** was developed in phorbol 12-myristate 13-acetate-differentiated U-937 cells. Analysis of whole-cell lysates by antiphosphotyrosine immunoblotting showed that incubation with **gamma interferon** brought about specific increases in phosphotyrosine labeling of several proteins. Increased labeling of these proteins occurred to similar extents in control cells and in cells that had been infected with *L. donovani* for 16 h. **Jak1**, **Jak2**, and **Stat1** were **immunoprecipitated** from control and interferon-treated cells, and tyrosine phosphorylation of these proteins, detected by antiphosphotyrosine immunoblotting was used to measure their activation. Tyrosine **phosphorylation** of **Jak1**, **Jak2**, and **Stat1** **increased** markedly, in a dose-dependent manner, in U-937 cells incubated with **gamma interferon**. In contrast, in cells infected with *L. donovani*, tyrosine **phosphorylation** of **Jak1**, **Jak2**, and **Stat1** was markedly **impaired**. This effect was dependent upon the duration of exposure to *L. donovani* and was **maximal and complete at 16 h**. **Results similar to those observed** with U-937 cells were also obtained with human peripheral blood monocytes. These findings indicate that infection of human mononuclear phagocytes with *L. donovani* leads to **impaired gamma interferon-mediated** tyrosine **phosphorylation** and selective effects on the **Jak-Stat1** pathway. Unresponsiveness to **gamma interferon** for activation of this pathway may explain impaired transcriptional responses in leishmania-infected cells.

Legend

☒ EventMention ☒ Gene

Click In Text to See Annotation Detail

Annotations

- EventMention
 - EventMention ("impaired")
 - begin = 1794
 - end = 1802
 - confidence = null
 - componentId = de.julielab.jules.JREX
 - id = E921
 - specificType = Negative_regulation
 - ref = null
 - resourceEntryList = null
 - textualRepresentation = null
 - arguments = FSArray
 - arguments = ArgumentMention ("phosphorylation")
 - begin = 1740
 - end = 1755
 - confidence = null
 - componentId = null
 - id = null
 - ref = EventMention ("phosphorylation")
 - begin = 1740
 - end = 1755
 - confidence = null
 - componentId = de.julielab.jules.JREX
 - id = E619
 - specificType = Phosphorylation
 - ref = null
 - resourceEntryList = null
 - textualRepresentation = null
 - arguments = FSArray
 - arguments = ArgumentMention ("Jak2")
 - begin = 1765
 - end = 1769
 - confidence = null
 - componentId = null
 - id = null
 - ref = Gene ("Jak2")
 - role = Theme

Challenge Competitions



- ParsEval, SemEval, RTE, ...
- MUC, ACE, TAC, SUMMAC
- BioCreative I, II. II.5, III, IV, LLL, NLPBA
- TREC (Genomics), CLEF eHealth, i2b2
- **BioNLP'09 Shared Task on Event Extraction**
 - <http://www-tsujii.is.s.u-tokyo.ac.jp/GENIA/SharedTask/>
- CALBC, MANTRA

Challenge Competition (1/3)

1. (vertrauenswürdiger, fairer, objektiver)
Ausrichter konstituiert sich
 - Thematik des Challenge festlegen
 - Textauswahl, Formate etc.
 - Wettbewerbssoftware bereitstellen
2. Anfertigung des Goldstandards (*ground truth*)
 - Aufspaltung in
 - Training-Set (70/90)
 - Test-Set (30/10)

Challenge Competition (2/3)

3. Freigabe des Training-Set (Dauer: 3-6 W)
 - Teilnehmer trainieren ihr System am Training-Set
 - Vergleich eigener Ergebnisse gegen Goldstandard
 - Teilnehmer fixiert am Ende der Trainingsphase *n* optimale Systemzustände (*frozen system*)
4. Freigabe des Test-Set (Dauer: 2-3 T)
 - Frozen system operiert auf Test-Set

Challenge Competition (3/3)

5. Abgabe der Ergebnisse beim Ausrichter
6. Auswertung der Ergebnisse des Test-Set-Laufs beim Ausrichter
 - Vergleich eigener Ergebnisse gegen Goldstandard
 - Standardisierte Metriken für Qualitätsmessung (precision, recall, F-score)
7. Vergleich und Ranking aller Teilnehmer durch Ausrichter
 - anonym (bei Bedarf)

And the winner is ...

Final Evaluation Results on ALL-TOTAL events by Approximate Span & Recursive Matching

<u>Team</u>	<u>gold (match)</u>	<u>answer (match)</u>	<u>recall</u>	<u>precision</u>	<u>fscore</u>
U Turku (FIN)	3182 (1487)	2541 (1486)	46.73	58.48	51.95
FSU Jena/JULIELab (GER)	3182 (1458)	3068 (1458)	45.82	47.52	46.66
Concordia U/CLaC (CAN)	3182 (1113)	1807 (1113)	34.98	61.59	44.62
U Tokyo+DBCLS (JAP)	3182 (1174)	2110 (1173)	36.90	55.59	44.35
Ghent U/VIB(BEL)	3182 (1063)	2062 (1063)	33.41	51.55	40.54
U Tokyo/Tsujii Lab (JAP)	3182 (895)	1671 (895)	28.13	53.56	36.88
U New South Wales (AUS)	3182 (898)	1957 (896)	28.22	45.78	34.92
U Zurich (SWI)	3182 (883)	1895 (883)	27.75	46.60	34.78
Arizona SU+HUB+BU(USA)	3182 (688)	1106 (688)	21.62	62.21	32.09
U Cambridge (UK)	3182 (672)	1181 (672)	21.12	56.90	30.80
U Antwerp/CNTSLTG (BEL)	3182 (716)	1501 (716)	22.50	47.70	30.58
U Manchester (UK)	3182 (702)	1444 (702)	22.06	48.61	30.35
SCAI Fraunhofer Inst (GER)	3182 (826)	2278 (826)	25.96	36.26	30.26
UAveiro (POR)	3182 (666)	1351 (666)	20.93	49.30	29.38
Team 24 (???)	3182 (722)	1778 (721)	22.69	40.55	29.10
U Szeged (HUN)	3182 (685)	1852 (685)	21.53	36.99	27.21
NICTA/U Melbourne (AUS)	3182 (555)	1388 (555)	17.44	39.99	24.29
CNB Madrid (ESP)	3182 (911)	4362 (911)	28.63	20.88	24.15
U Colorado/BTMG (USA)	3182 (428)	596 (428)	13.45	71.81	22.66
Arizona SU/CIPS (USA)	3182 (725)	3809 (725)	22.78	19.03	20.74
U Michigan (USA)	3182 (968)	6859 (968)	30.42	14.11	19.28
Sirma/Ontotext (BUL)	3182 (358)	538 (358)	11.25	66.54	19.25
Team 09 (???)	3182 (372)	1184 (372)	11.69	31.42	17.04
KoreaU (KOR)	3182 (299)	485 (299)	9.40	61.65	16.31

Post-competition Results I

Final Evaluation Results on ALL-TOTAL events by Approximate Span & Recursive Matching

<u>Team</u>	<u>gold (match)</u>	<u>answer (match)</u>	<u>recall</u>	<u>precision</u>	<u>fscore</u>
U Turku (FIN)	3182 (1487)	2541 (1486)	46.73	58.48	51.95
FSU Jena/JULIELab (GER)	3182 (1458)	3068 (1458)	45.82	47.52	46.66

Evaluation Results on ALL-TOTAL events by Approximate Span & Recursive Matching
after System Overhaul and further Tuning

<u>Team</u>	<u>gold (match)</u>	<u>answer (match)</u>	<u>recall</u>	<u>precision</u>	<u>fscore</u>
U Turku (FIN)					52.86
FSU Jena/JULIELab (GER)					51.10

In:
Computational Intelligence
Vol. 27, 2011, No.4, pp.610-44.

Post-Competition Results II

Participant	Rank in F1 score	Total			Localization	Binding	Gene expression	Transcription	Protein catabolism	Phosphorylation	Regulation	Positive Regulation	Negative Regulation	BioNLP '09 ST Total Evaluation
	#	F1	PR	RC	F1	F1	F1	F1	F1	F1	F1	F1	F1	F1
JULIE Lab JReX [1]	1	51.09	57.69	45.85	61.60	49.24	72.48	42.99	80.00	81.99	31.20	40.39	38.47	46.66
UTurku [2]	2	49.91	56.32	44.81	55.85	45.43	71.67	50.21	50.00	79.70	33.97	38.66	36.28	51.95
EventMine [3]	3	48.20	64.00	38.65	63.20	39.86	72.63	50.00	60.87	81.29	28.77	28.25	32.62	36.88
BExtract [4]	4	44.48	61.56	34.82	51.45	26.97	65.14	24.71	60.00	80.69	32.21	35.83	33.27	44.62
VIBGhent [7]	5	42.44	59.05	33.12	51.79	34.42	69.57	57.14	68.97	76.23	19.39	23.34	26.67	40.54
TheBeast [8]	6	37.19	48.15	30.30	48.98	34.50	59.28	17.48	72.00	72.79	29.96	29.57	27.32	44.35
UMich [9]	7	36.34	35.57	37.15	53.47	31.75	66.00	30.06	58.06	77.15	14.29	21.50	26.61	19.28
Moara [6, 10]	8	29.50	31.99	27.31	44.19	28.36	58.79	26.40	50.00	52.88	10.83	14.68	13.16	24.15
CCP-BTMG [13]	9	22.03	70.03	13.07	17.80	20.92	51.07	22.93	40.00	33.33	5.79	6.69	4.01	22.66

In:
BMC Bioinformatics
 Vol. 12, 2011, No.481

Merkmale von Challenge Competitions

- Internationaler Ideen-Wettbewerb
- Intersubjektive Bewertung
- Saubere Vergleichsmaßstäbe: Metriken
- Experimente
- Trennung Experimentator/Entwickler
- „sportlicher“ Aspekt
- Offenlegung der Methoden
 - Treiber für Methodenfortschritt
 - Latente Gefahr des Methodenkonservatismus‘
- Exemplarischer Fall empirischer Wissenschaft: Rationalität

</Exkurs>

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, ...

Formale Sprachen

- **Konstruiert**
 - Rein definitorischer (konstruktiver) Ansatz
- **Möglichst non-ambig**
 - Eindeutige syntaktische wie semantische Strukturen
- **Statisch**
 - zum Definitionszeitpunkt komplett fixiert
 - Endliches Vokabular
- **„Einfache“ Beschreibung**
 - Wenige Regeln, wenige Axiome
 - meist wenige Elemente umfassendes Vokabular („Lexikon“)
 - Wenige Schichten: Syntax, Semantik; keine Pragmatik
- **striktes Wohlgeformtheitskriterium**
 - Außer-definitorische Strukturen sind nicht wohlgeformt
 - ... und damit nicht prozessierbar

Natürliche Sprachen

- Konventionalisiert durch ‚sozialen Vertrag‘ einer Sprechergemeinschaft
 - Ausübung des Sprechens unterliegt sozialen Normen, Gewohnheiten und (impliziten) Übereinkünften (Regelkonformität)
- Hochgradig ambig
 - Mehrdeutige lexikalische, syntaktische, semantische, pragmatische Strukturen
- Dynamisch
 - Sprache verändert sich im Laufe der Zeit (Lexikon, Syntax)
 - Unendliches Vokabular (Komposition, Derivation)
- Komplexe Beschreibungen
 - Viele Regeln, viele Axiome
 - Sehr großes Vokabular („Lexikon“)
 - Starke Schichtung von Beschreibungsebenen
- laxe Wohlgeformtheitskriterien
 - Außer-definitorische Strukturen sind zwar nicht wohlgeformt, werden aber (bis zu einem gewissen Grad) verstanden

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 System)

Computerlinguistik-Standorte

www.ims.uni-stuttgart.de/info/SitesEurope.html#Germany



Computerlinguistik-Standorte

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

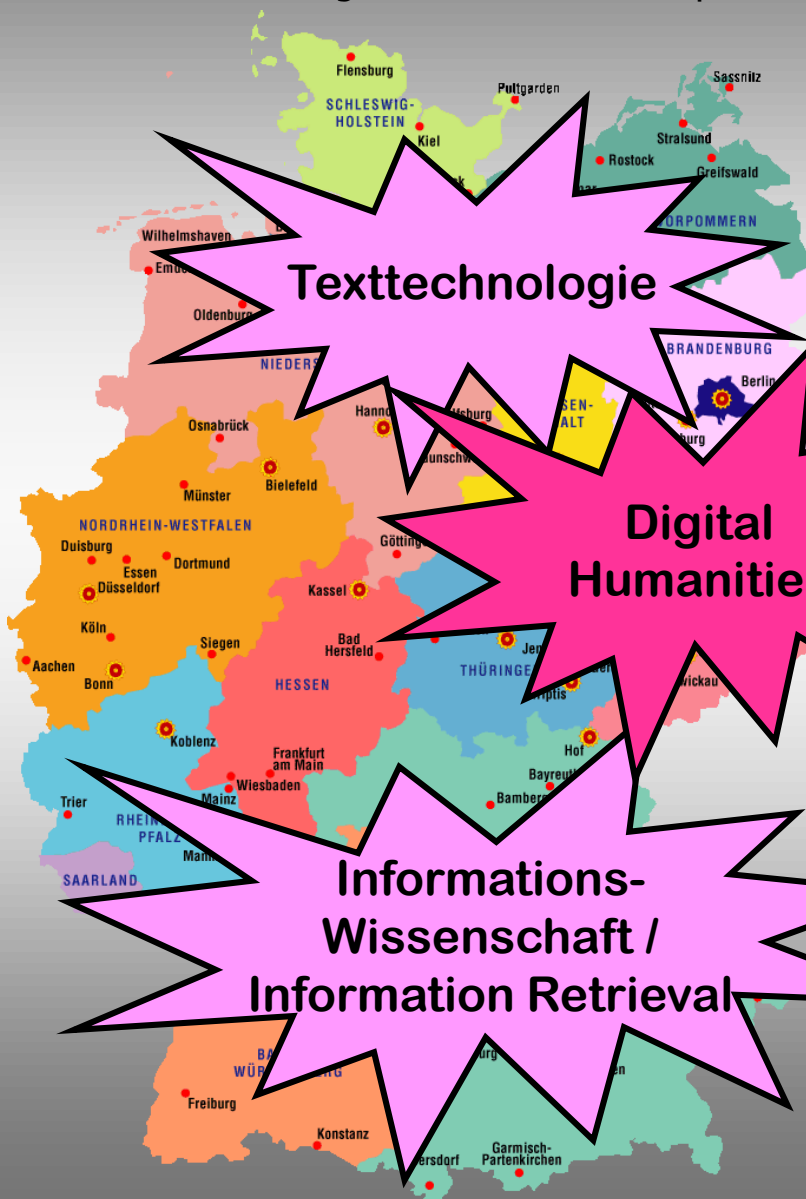
U Gießen

U Hildesheim

U Koblenz

Computerlinguistik-Standards

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

24 [48]

U 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

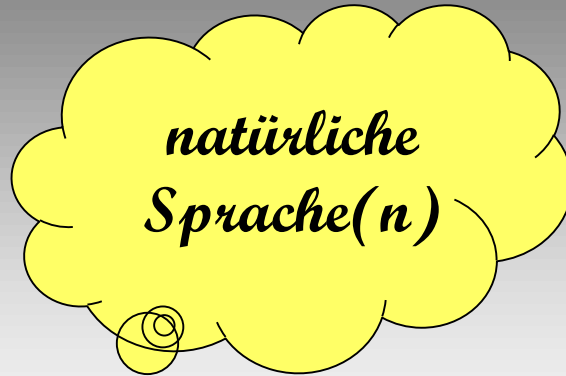
U Bamberg

U Kaiserslautern

TU Dresden

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

Verortung der Computerlinguistik



Theoretische Linguistik
Generative Grammatik
Dependenzgrammatik
Unifikationsgrammatik
Konstruktionsgrammatik
modelltheoretische oder
strukturelle Semantik
Frame-Semantik . . .

Algebra
Formale Grammatiken
Formale Sprachen
Automatentheorie
Graphentheorie
Logik
Wahrscheinlichkeitstheorie

Algorithmen & Datenstrukturen
Programmierung
Mustererkennung
Informationssysteme
Künstliche Intelligenz
Maschinelles Lernen,
Automatisches Schließen

Deskription

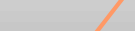
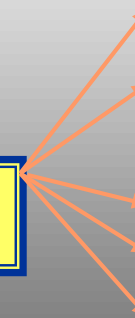
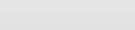
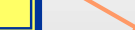
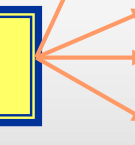
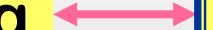
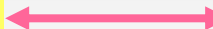
Linguistik

Formalisierung

Mathematik

**Algorithmisierung
Programmierung**

Informatik

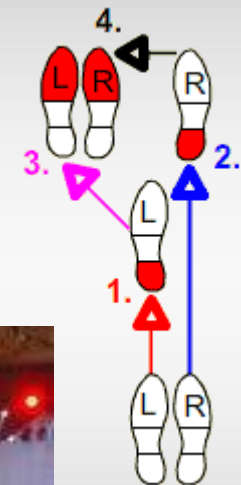
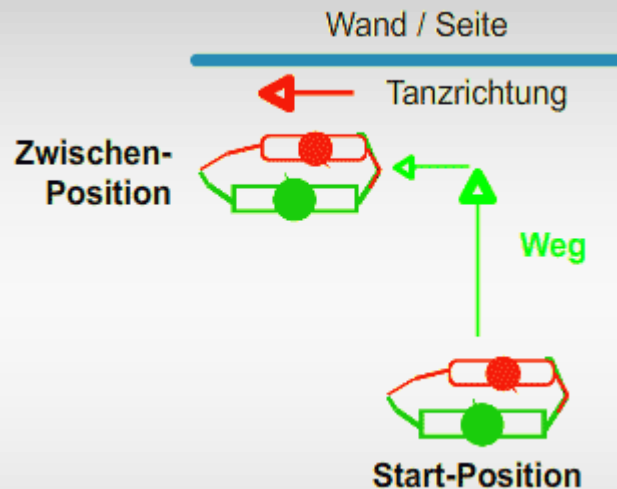


Keine natürlichen, aber doch auch Sprachen (1/6)

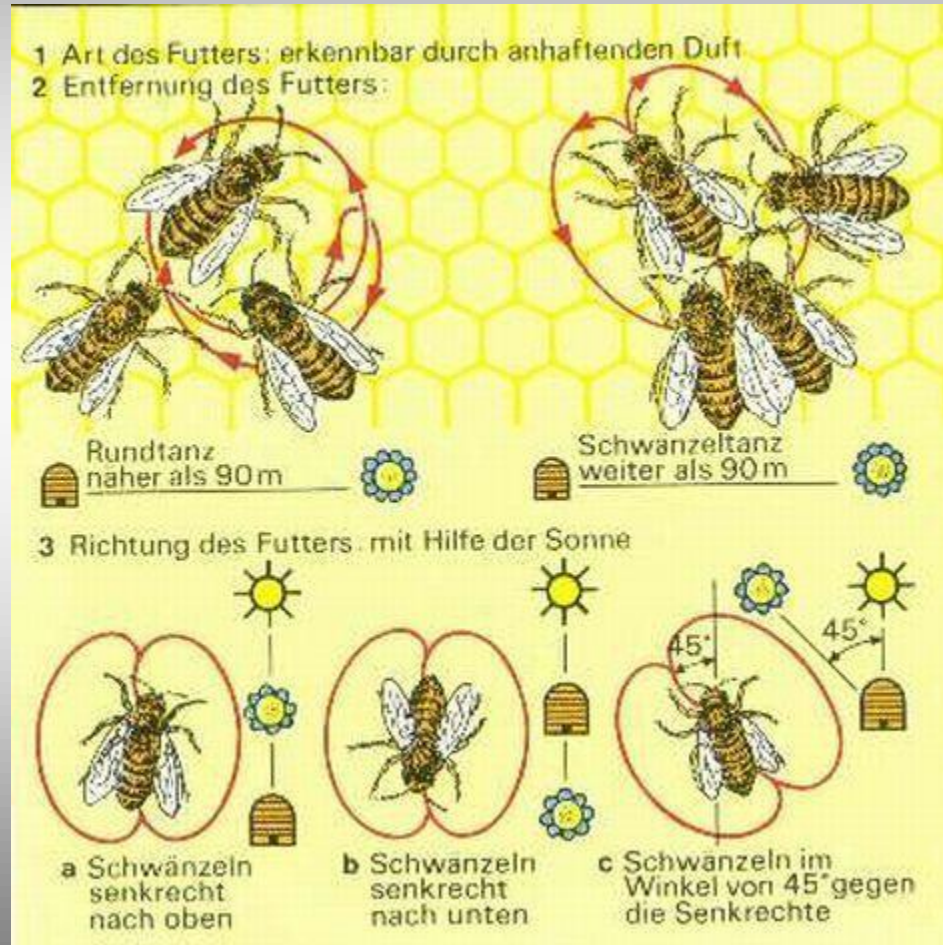
Allegro assai.

The musical score is presented in four systems, each with a piano (p) and vocal (v) staff. The key signature is G minor (three flats) and the time signature is 12/8. The tempo is marked 'Allegro assai.' at the beginning. The dynamics are indicated by 'pp' (pianissimo), 'p' (piano), 'mp' (mezzo-piano), and 'f' (forte). The score includes various musical notations such as notes, rests, and fingerings. The vocal line is introduced in the third system with the lyrics 'poco ritar - dan - do a tempo'. The score is a transcription of the first 16 measures of the first movement of Beethoven's Piano Sonata No. 23, 'Appassionata'.

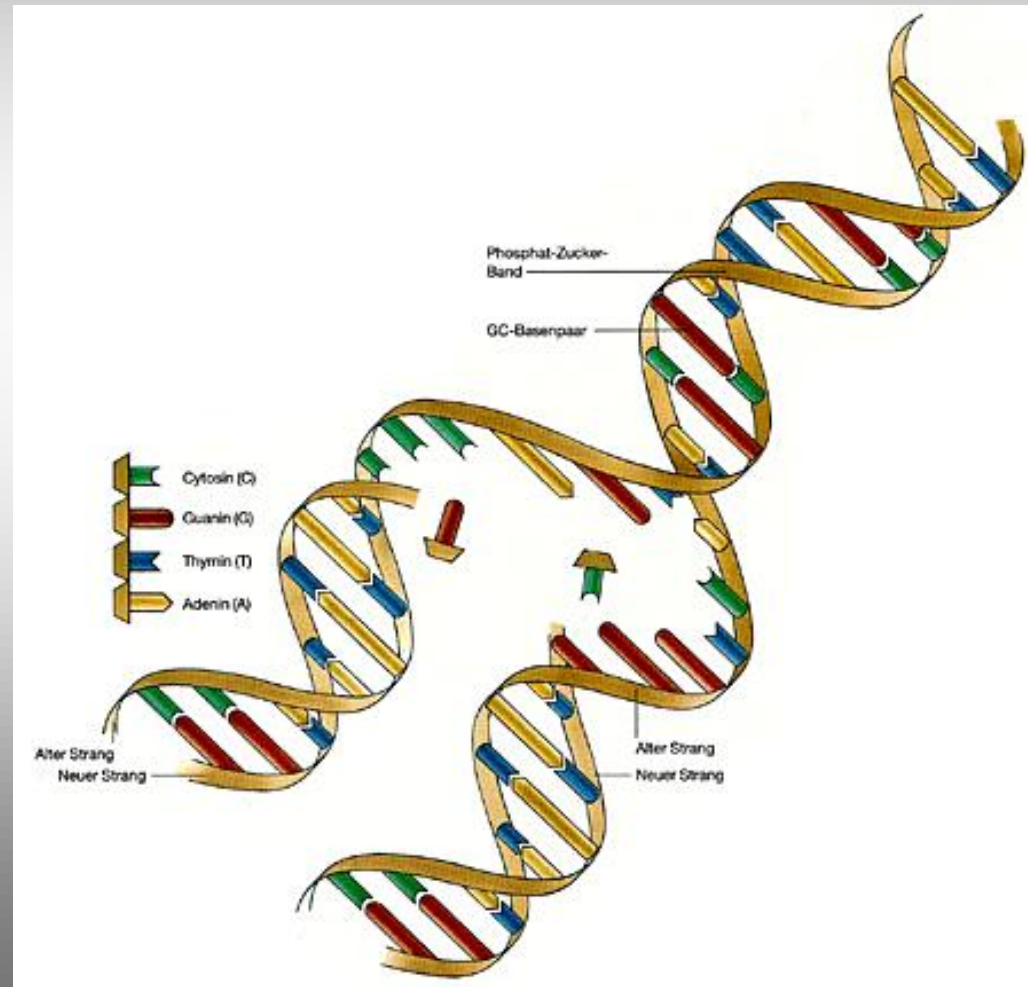
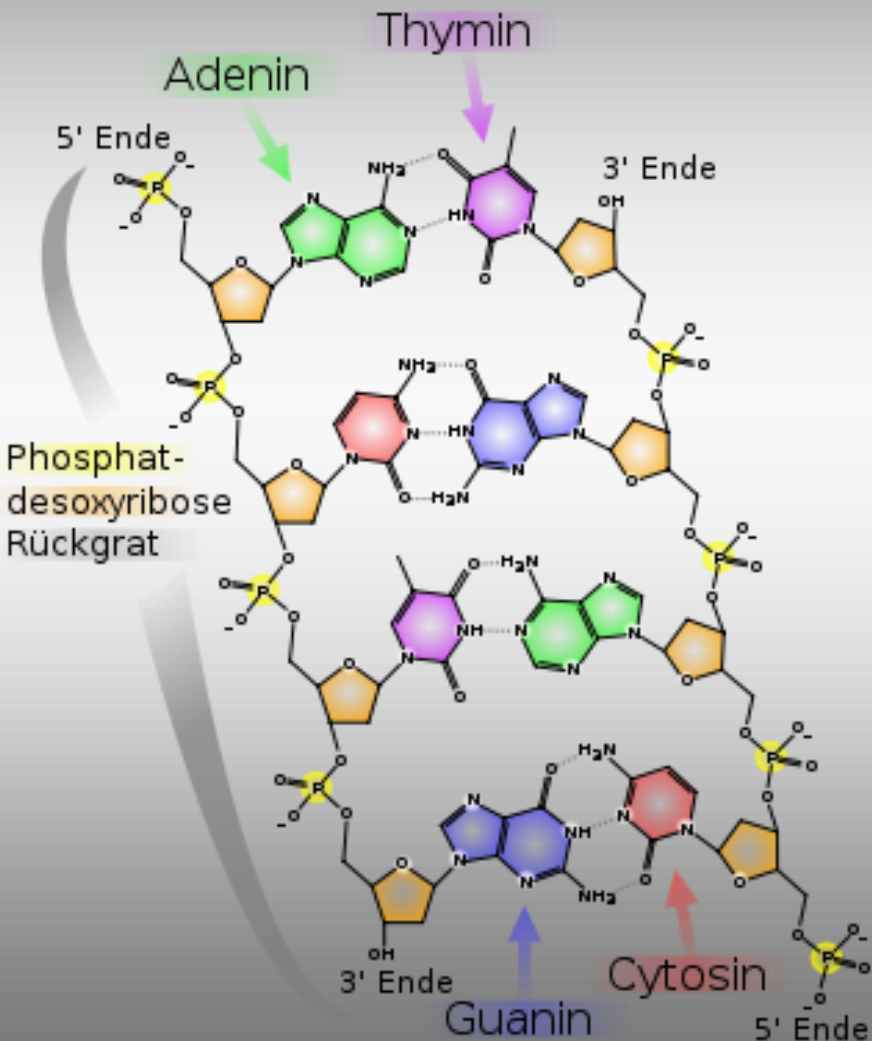
Keine natürlichen, aber doch auch Sprachen (2/6)



Keine natürlichen, aber doch auch Sprachen (3/6)



Keine natürlichen, aber doch auch Sprachen (4/6)



Keine natürlichen, aber doch auch Sprachen (5/6)



Keine natürlichen, aber doch auch Sprachen (6/6)



Zur Phänomenologie natürlicher Sprachen

- Linguistische Ebenen
- Produktivität
- Kontext
- Paraphrasen
- Ambiguität
- Graduierung von Korrektheit & Verstehbarkeit

Natürliche Sprache

verschiedene Approximationsstufen

- hciltsinatsemdnEre!eSgnaf
- fan gSe! erEn dmest anis tlich
- fangSe!erEndmestanistisch
- fang Se ! er End mest an ist lich
- Endlich ist Semesteranfang!

Symbolmengen (Vokabular)

(konventionelle) lineare Reihung

Natürliche Sprachen – verschiedene Schriftarten

لكبار. يُمْسِكُ بِيَدَيْ الدارس المبتدئ الذي لا يعرف الكفاية، يتيح له فهم اللغة، واستعمالها في الحياة سلة القراءة في الكتب العربية. وهو مكون من ثلاث ج التراكيب النحوية، تقديمًا وظيفيًا تطبيقيًا، وراعى على تدريبات الانماط.

أشيقا، ويلبي حاجات الكبار غير الناطقين بالعربية نظر الى اللغة على أنها مجموعة من المهارات العامة ريق التواصل معه، وإشراكه في اكتشافها ثم إدراكها بالثقتين.

पाचन की दृष्टि से राई, गेहूँ, जौ आदि का सेवन मुझे माफिक नहीं आता । ये सब चीजें मेरे स्वास्थ्य के अनुकूल नहीं हैं । मैं खास कर के आटा, रोटी या ब्रेडचूर्ण युक्त पदार्थ सेमियाँ, नूडल आदि से बने खाद्य पदार्थ बिल्कुल नहीं खा सकती ।

मेरे लिए कुछ ऐसे भोज्य पदार्थ परोसें जिनमें उपरोक्त पोषीयक दवाइन पदार्थों का मिश्रण न हो ।

धन्यवाद । [Hindi]

«Если хочешь сделать зловоние, возьми человеческий кал и мочу, вонючую лебеду, если же у тебя её нет, капусту и свеклу, и вместе положи в стеклянную бутылку, хорошо закупоренную,

21 'Et est notitia deo auxiliante:
в течен Pro annonis et capitu pro tempore praefecti
am Africam auri libras centum.
consiliariorum auri libras viginti.
cancellariorum auri libras septem.
sius ita:
mo hominibus decem pro annonis
pitu XIIIS, fiunt solidi CXLVIIS.
o annonis VI annona solidorum V
I capitus solidorum IIII, fiunt so-
lo pro annonis III annona solidorum
I capitus solidorum IIII, fiunt so-
pro annonis II annona solidorum
S capitus solidorum IIII, fiunt so-
quinto et sexto ad annonas IS an-
et ad caputem I capitus solidi-
solidi XXVIII, reliquis quattuor
ona solidorum V et ad caputem S
capitus solidorum IIII, fiunt solidi XXVIII.

Symbolmengen (Vokabular)

(konventionelle) lineare Reihung

Natürliche Sprache

Linguistische Ebenen: Lexikologie

- **Vollformen**

- rede
- redest
- reden
- redet
- Rede
- Reden
- Redner
- Redners

Natürliche Sprache

Linguistische Ebenen: Lexikologie

- **Vollformen Grundformen**

- rede
- redest
- reden
- redet
- Rede
- Reden
- Redner
- Redners
- reden [V]
- Rede [N]
- Redner [N]

Natürliche Sprache

Linguistische Ebenen: Lexikologie

- **Vollformen**
 - rede
 - redest
 - reden
 - redet
 - Rede
 - Reden
 - Redner
 - Redners
 - Grundformen**
 - reden [V]
 - Rede [N]
 - Redner [N]
 - Stämme**
 - RED
- Granularität
linguistischer
Einheiten
(Primitive, Atome)*

Natürliche Sprache

Linguistische Ebenen: Lexikologie

- Lexikoneintrag

- Redner

- Sprache: deutsch
 - Wortart: Nomen
 - Genus: maskulin
 - Numerus: (SG, PL)
 - Deklinationsklasse: D4 (SG:-s, PD:-n)
 - Bedeutung: jmd., der redet
jmd., der eine Rede hält

Natürliche Sprache

Linguistische Ebenen: Syntax

- Er schrieb ein erfolgreiches Buch.
- Schrieb er ein erfolgreiches Buch?
- Schrieb er [ein erfolgreiches Buch]?
- Schrieb er [es]? *Gruppierung*
- * Schrieb er ein [es]? *(linguistische Phrase)*
konventionelle lineare Reihung
(auf Satzebene)
- * Er Buch ein schrieb erfolgreiches.
- ** Er hucB nie chriseb eresreilgchfo.

Natürliche Sprache

Linguistische Ebenen: Semantik

- Er schrieb ein Buch.
- Er schrieb **kein** Buch.
- Er schrieb ein **Buch**.
- Er schrieb einen **Brief**.
- *Er schrieb einen **Berg**.
- ****Die Zündkerze** schrieb einen **Berg**.

Natürliche Sprache

Linguistische Ebenen: Semantik

- Satzsemantik: Kompositionalität
 - Er **gibt** mir sein Auto.
 - Sie **beendete** ihr **Arbeitsverhältnis**.
- „Feste“ Phrasen: Kollokationen
 - Er **stellt** mir sein Auto **zur Verfügung**.
 - Sie **gab** ihren **Posten auf**.
- Metonymie
 - Er fährt einen **[von der Firma] Ferrari [gebauten Sportwagen]**.
[**producer-for-product**]
- Metapher
 - Ich **gebe keinen Pfifferling** für dieses Team.
 - Dieses Team **hat keine Aussicht auf Erfolg**.

Natürliche Sprache

Linguistische Ebenen: Pragmatik

- Er schrieb ein Buch über Napoleon.
- *Er schrieb ein Buch über **den jetzigen_[t=2018] Kaiser von Frankreich.**
- Können Sie mir die Uhrzeit sagen?
 - 12.35 Uhr!
 - ***Ja!**

Natürliche Sprache

Linguistische Ebenen: Diskurs/Text

- **Das belastende Recherchematerial fehlte. Der Journalist öffnete den Safe. Aber das war jetzt ohne Belang. Er saß in der Falle. Sein Geld war noch da.**

Natürliche Sprache

Linguistische Ebenen: Diskurs/Text

*„logische“ lineare Reihung
(auf Textebene)*

- Das belastende Recherchematerial fehlte. Er saß in der Falle. Aber das war jetzt ohne Belang. Sein Geld war noch da. Der Journalist öffnete den Safe.
- **Der Journalist öffnete den Safe. Das belastende Recherchematerial fehlte. Sein Geld war noch da. Aber das war jetzt ohne Belang. Er saß in der Falle.**

Natürliche Sprache

Linguistische Ebenen: Diskurs/Text

*unterschiedlichste
Bezeichner(phrasen)
für einen Referenten*

- Referenz (Kohäsion)

- Angela Merkel rüffelte ihren Finanzminister. Olaf Scholz hatte ihr neueste Haushaltsdaten verschwiegen. Die Kanzlerin erfuhr dies auf ihrem Rückflug vom Weltwährungsgipfel. Der schmallippige Geldhüter ist für solche Überraschungen schon bekannt. Gut möglich, dass der hanseatische Haushaltsvorstand sich für höhere Aufgaben profiliert. In Berlin werden schon die ersten Namen als Nachfolger des sozialdemokratischen Ministers gehandelt.

Natürliche Sprache

Linguistische Ebenen: Diskurs/Text

*argumentative
Makrostruktur*

- Textsemantik: Kohärenzrelationen
 - Angela Merkel **rüffelte** ihren Finanzminister. Olaf Scholz hatte ihr neueste Haushaltsdaten **verschwiegen**. Die Kanzlerin **erfuhr dies** auf ihrem Rückflug vom Weltwährungsgipfel. Der schmallippige Geldhüter ist für solche Überraschungen schon bekannt. Gut möglich, dass der hanseatische Haushaltsvorstand sich für **höhere Aufgaben** profiliert. In Berlin werden schon **die ersten Namen als Nachfolger** des sozialdemokratischen Ministers **gehandelt**.
 - [**Begründung** – **Elaboration** – **Evidenz**]

Zur Phänomenologie natürlicher Sprachen

- Linguistische Ebenen
- **Produktivität**
- Kontext
- Paraphrasen
- Ambiguität
- Graduierung von Korrektheit & Verstehbarkeit

Natürliche Sprache

Produktivität: lexikalisch-semantic

Wörter können nicht

- Derivation *erschöpfend „aufgezählt“ werden*
 - grün: grünlich, begrünen, Grün
- Komposition
 - grün ● gelb, Grün ● anlage , Grün ● streifen
- Neologismen *Wissensintensiv, regelaffin, kreativ*
 - Schweine-Grippe, Jamaika-Koalition
 - googlen, simsen, chatten, whatsappen,
 - Handy, Jazz, Meme

Natürliche Sprache

Produktivität: syntaktisch

- Einbettung

- Das Buch verkauft sich gut.
- Das Buch, das X geschrieben hatte, verkauft sich gut.
- Das Buch, das X, der auch Autor von Y war, geschrieben hatte, verkauft sich gut.

- Koordination

- Er schrieb an X.
- Er schrieb an X und an Y.
- Er schrieb an X, an Y und an Z.

*Sätze können nicht
erschöpfend „aufgezählt“ werden*

Zur Phänomenologie natürlicher Sprachen

- Linguistische Ebenen
- Produktivität
- **Kontext**
- Paraphrasen
- Ambiguität
- Graduierung von Korrektheit & Verstehbarkeit

Natürliche Sprache

Kontext

- Morphosyntax
 - dies**es** interessante **Buch**Ø
 - die neu**en** Büch**er**
- Syntax
 - Heute **geht** die Sonne um 7.05 Uhr ... **auf**.
 - **Das Buch** von X, **das** sich gut verkaufte ...

Natürliche Sprache

Kontext

- Lexikalische Semantik
 - [+human,+schreibkundig] *schreiben* [Schriftstück]
 - Der Journalist schreibt einen Leitartikel.
 - Der Komponist schreibt [den Notentext für] eine Ballade.
 - (*)Der Pygmäe schreibt einen Protestbrief.
 - *Der Journalist schreibt eine Sahnetorte.
 - *Der Walzstahl schreibt einen Leitartikel.
 - **Der Walzstahl schreibt eine Sahnetorte.

Natürliche Sprache

Kontext

- Referenzieller Diskurskontext
 - Der Chefredakteur hatte die Kolumne geschrieben. Sie war ihm besonders gelungen. [syntaktisch-grammatisch]
 - Der Chefredakteur hatte den Leitartikel geschrieben. Er war ihm besonders gelungen. [semantisch]
Er war mit ihm zufrieden. [semantisch]
*Er war mit ihm zufrieden. *[semantisch]
Er war mit sich zufrieden. [semantisch]
- Konzeptueller Diskurskontext
 - Der Chefredakteur hatte den Leitartikel geschrieben. Der Titel war dem Journalisten besonders gelungen.
- Situationeller Diskurskontext (Schemata)
 - Der Journalist wusste den Code. Er öffnete den Safe, aber das belastende Recherchematerial fehlte. 63

Zur Phänomenologie natürlicher Sprachen

- Linguistische Ebenen
- Produktivität
- Kontext
- **Paraphrasen**
- Ambiguität
- Graduierung von Korrektheit & Verstehbarkeit

Natürliche Sprache

Paraphrasen: monolingual

- Syntax
 - Seine Amtszeit geht **in diesem Jahr** zu Ende.
 - **In diesem Jahr** geht seine Amtszeit zu Ende.
- Lexikalische Semantik
 - Seine Amtszeit **geht** in diesem Jahr **zu Ende**.
 - Seine Amtszeit **endet** in diesem Jahr.
 - Seine Amtszeit **läuft** in diesem Jahr **ab**.
- Referenzielle Semantik
 - Seine Amtszeit geht **in diesem Jahr** zu Ende.
 - Seine Amtszeit geht **2018** zu Ende.

Natürliche Sprache

Paraphrasen: multilingual

- Auf Wiedersehen, Herr Präsident!
- So long, Mr. President!
- Au revoir, Monsieur le président!
- Ciao, signore presidente!

Zur Phänomenologie natürlicher Sprachen

- Linguistische Ebenen
- Produktivität
- Kontext
- Paraphrasen
- **Ambiguität**
- Graduierung von Korrektheit & Verstehbarkeit

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**_{INSTRUM} Fernrohrs
 - Sie sahen **den Mann**, der **sein**_{POSSESS} Fernrohr **trug**

Natürliche Sprache

Ambiguität: syntaktisch

- PP-Anbindung & Homografie/Polysemie

- They saw the man with the telescope.

- Sie sahen den Mann mit Hilfe ihres_{INSTRUM} Fernrohrs.

- Sie sahen den Mann, der sein_{POSSESS} Fernrohr trug.

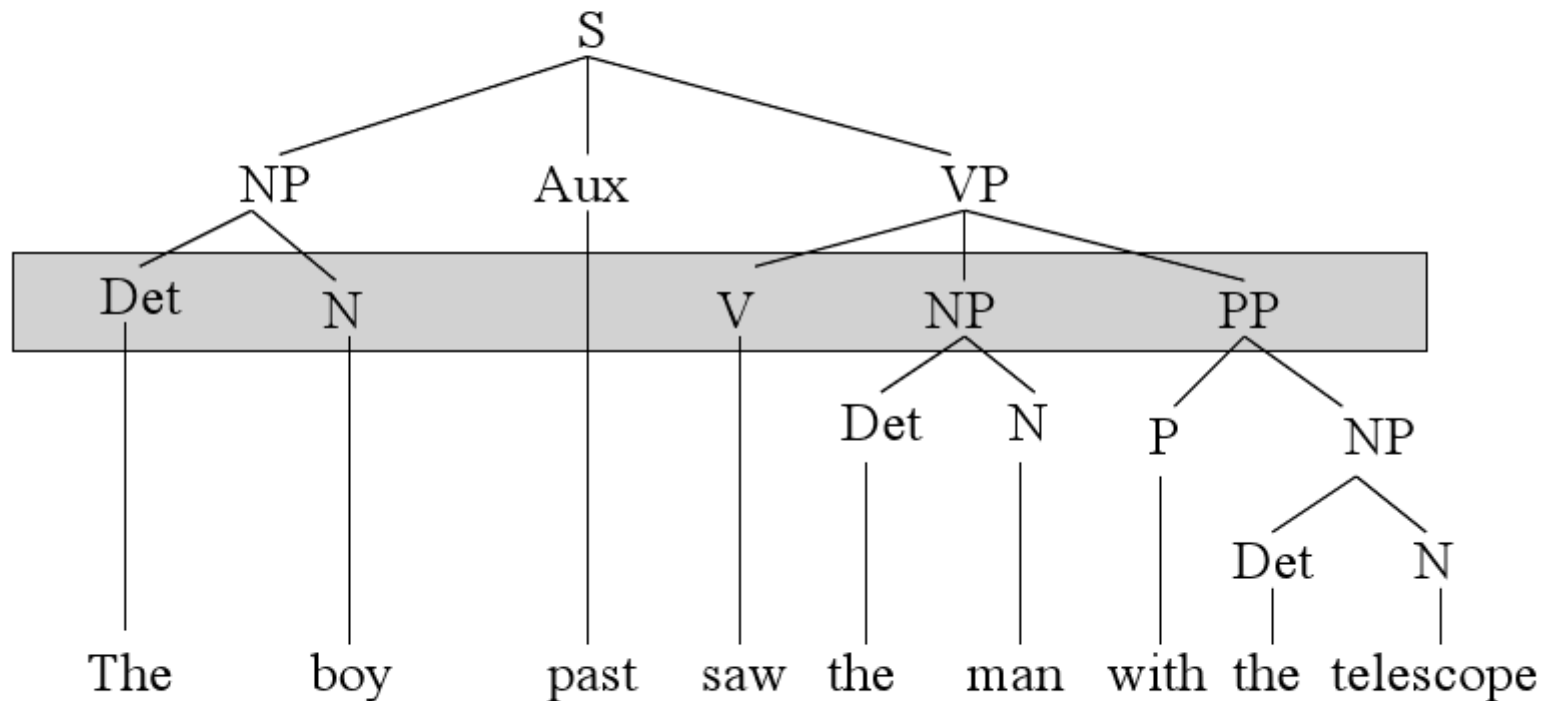
- Sie zersägen den Mann mit Hilfe ihres_{INSTRUM} Fernrohrs.

- Sie zersägen den Mann, der sein_{POSSESS} Fernrohr trug.



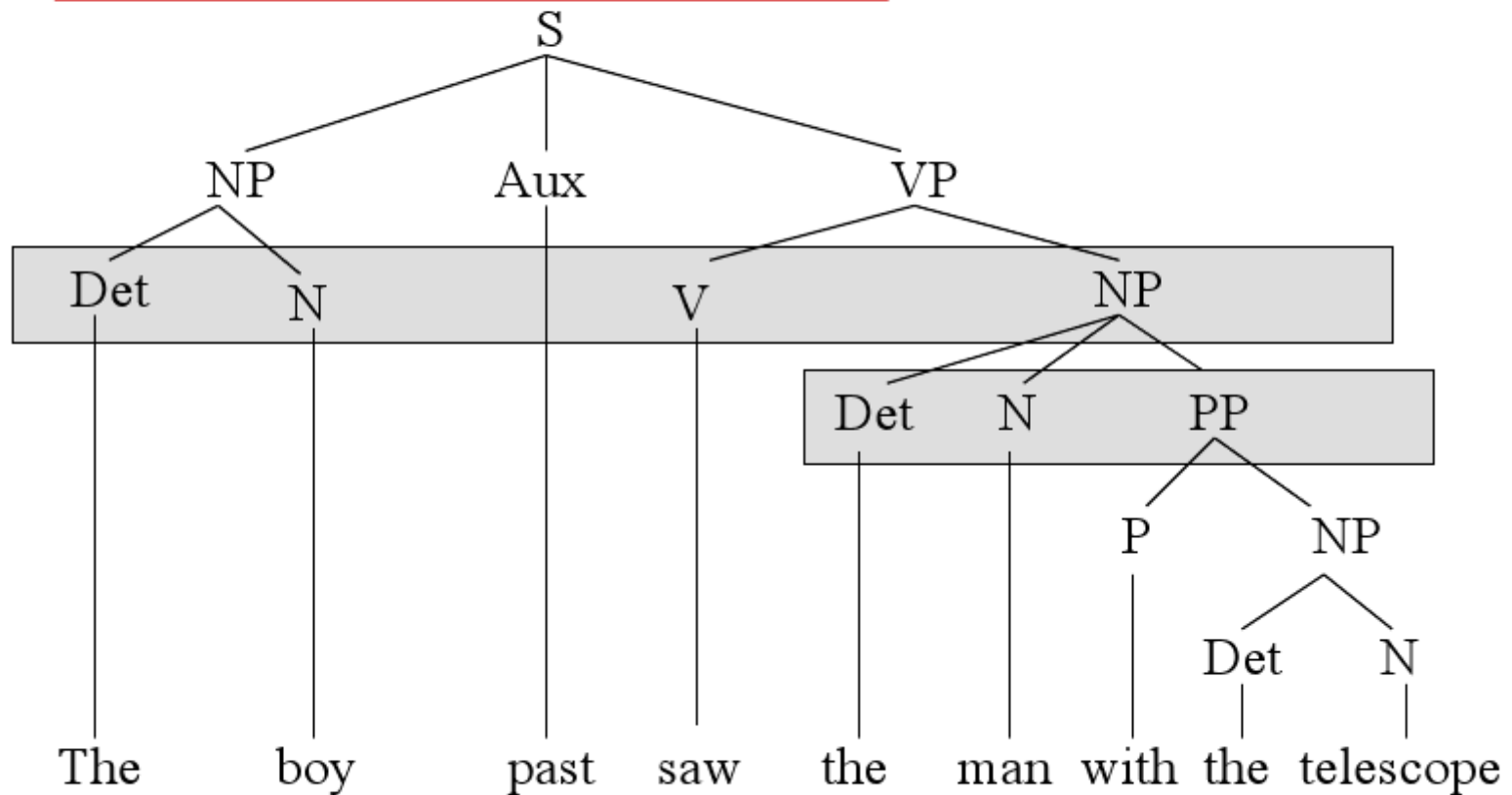
Structural Ambiguity (1)

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]₁ [am frühen Mittwoch Morgen]₂ [in Bad Nauheim]₃ [nach einem 16-stündigen Verhandlungsmarathon]₄ [auf einen Tarifabschluss]₅ [in Höhe]₆ [von 1,2 Prozent]₇ [für die 160.000 Beschäftigten]₈ [in der hessischen Bekleidungsindustrie]₉ geeinigt.

Natürliche Sprache

Ambiguität: syntaktisch

- Wachstumsverhalten

- Die Tarifparteien behaupten

eines Kompromisses

Schlichtung

[in Baden-Württemberg]

Verhandlungen

Tarifabschluss

[für die öffentlichen

Bediensteten]

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

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

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>**

Zur Phänomenologie natürlicher Sprachen

- Linguistische Ebenen
- Produktivität
- Kontext
- Paraphrasen
- Ambiguität
- Graduierung von Korrektheit & Verstehbarkeit

Natürliche Sprache

Graduierung von Korrektheit, Verstehbarkeit

- Reihenfolgenverletzung (Grammatikalität)
 - ****irun gizu kles ken jebtna wam jainezb?
 - ***inklusenbe jajezzna warbtkeinumgi?
 - **keinenwarumjazz jenain esklubgibt?
 - *in klub gibt es keinen jena warum jazz?
 - warum gibt es in Jena keinen Jazzklub?

Natürliche Sprache

Graduierung von Korrektheit, Verstehbarkeit

- Basisformen, Lücken
 - **dinner night restaurant?
 - *have dinner tonight in restaurant?
 - can we have dinner tonight in this restaurant?

Textverstehen

Originaltext

Am Freitagabend kam es auf der Bundesstraße 69 in der Gemarkung Ampermoching zu einem schwerwiegenden Verkehrsunfall. Ein Auto brach in Folge überhöhter Geschwindigkeit in einer lang gezogenen Rechtskurve nach links aus, schleuderte einen Abhang hinab und überschlug sich anschließend mehrfach auf einem angrenzenden Maisfeld. Drei der vier Insassen wurden aus dem Wagen heraus geschleudert und wurden dabei schwer verletzt. Der Fahrer des Wagens, ein 19-jähriger Klempner aus Erding, konnte von der rasch eintreffenden Ortsfeuerwehr zwar noch lebend aus seinem Wagen mit der Rettungsschere geborgen werden, erlag aber seinen schweren Verletzungen auf dem Weg in die Klinik. Der Unfallwagen, ein getunter Golf, wurde vollständig zerstört. Der Schaden beläuft sich auf ca. 15.000 Euro.

Textverstehen

Originaltext und Paraphrase (I)

Am Freitagabend kam es auf der Bundesstraße 69 in der Gemarkung Ampermoching zu einem schwerwiegenden Verkehrsunfall. Ein Auto brach in Folge überhöhter Geschwindigkeit in einer lang gezogenen Rechtskurve nach links aus, schleuderte einen Abhang hinab und überschlug sich anschließend mehrfach auf einem angrenzenden Maisfeld. Drei der vier Insassen wurden aus dem Wagen heraus geschleudert und wurden dabei schwer verletzt. Der Fahrer des Wagens, ein 19-jähriger Klempner aus Erding, konnte von der rasch eintreffenden Ortsfeuerwehr zwar noch lebend aus seinem Wagen mit der Rettungsschere geborgen werden, erlag aber seinen schweren Verletzungen auf dem Weg in die Klinik. Der Unfallwagen, ein getunter Golf, wurde vollständig zerstört. Der Schaden beläuft sich auf ca. 15.000 Euro.

Ein schlimmer Verkehrsunfall ereignete sich am Freitagabend auf der Bundesstraße 69 in der Gemarkung Ampermoching. Wegen zu hoher Geschwindigkeit verlor der Fahrer in einer lang gezogenen Rechtskurve die Herrschaft über seinen Wagen, rutschte einen Abhang hinab und landete auf einem Maisacker. Von den vier Insassen erlitten drei schwere Verletzungen. Der 19-jährige Fahrer, wohnhaft in Erding, konnte zwar noch mit schwerem Rettungsgerät aus den Trümmern seines Wagens befreit werden, verstarb aber während des Krankentransports. Sein auf hohe Geschwindigkeiten getrimmtes Auto besitzt nur noch Schrottwert. Der entstandene Sachschaden wird auf 15.000 € geschätzt.

Textverstehen

Originaltext und Paraphrase (II)

Am Freitagabend kam es auf der Bundesstraße 69 in der Gemarkung Ampermoching zu einem schwerwiegenden Verkehrsunfall. Ein Auto brach in Folge überhöhter Geschwindigkeit in einer lang gezogenen Rechtskurve nach links aus, schleuderte einen Abhang hinab und überschlug sich anschließend mehrfach auf einem angrenzenden Maisfeld. Drei der vier Insassen wurden aus dem Wagen heraus geschleudert und wurden dabei schwer verletzt. Der Fahrer des Wagens, ein 19-jähriger Klempner aus Erding, konnte von der rasch eintreffenden Ortsfeuerwehr zwar noch lebend aus seinem Wagen mit der Rettungsschere geborgen werden, erlag aber seinen schweren Verletzungen auf dem Weg in die Klinik. Der Unfallwagen, ein getunter Golf, wurde vollständig zerstört. Der Schaden beläuft sich auf ca. 15.000 Euro.

A severe car accident occurred Friday evening on National Route 69 near Ampermoching. A speeding car swerved off a right-hand bend. The vehicle catapulted down a hill and rolled over several times on a nearby corn field. Three of the passengers were thrown out off the car and fell on the ground heavily injured. The car's driver, a 19-year old plumber from Erding, was freed out of the damaged car by a rescue team but finally died on his way to the hospital. The vehicle, a sports-tuned Golf car, was totally destroyed. The estimated damage comes close to 15,000 Euros.

Textverstehen

Faktenbezogene Fragebeantwortung

Am Freitagabend kam es auf der Bundesstraße 69 in der **Gemarkung Ampermoching** zu einem schwerwiegenden **Verkehrsunfall**. Ein Auto brach in Folge überhöhter Geschwindigkeit in einer lang gezogenen Rechtskurve nach links aus, schleuderte einen Abhang hinab und überschlug sich anschließend mehrfach auf einem angrenzenden Maisfeld. Drei der vier Insassen wurden aus dem Wagen heraus geschleudert und wurden dabei schwer verletzt. Der Fahrer des Wagens, ein 19-jähriger Klempner aus Erding, konnte von der rasch eintreffenden Ortsfeuerwehr zwar noch lebend aus seinem Wagen mit der Rettungsschere geborgen werden, erlag aber seinen schweren Verletzungen auf dem Weg in die Klinik. Der Unfallwagen, ein getunter Golf, wurde vollständig zerstört. Der Schaden beläuft sich auf ca. 15.000 Euro.

F: In welcher Gemarkung geschah der Verkehrsunfall?

A: **Ampermoching!**

Textverstehen

Faktenbezogene Fragebeantwortung

Am Freitagabend kam es auf der Bundesstraße 69 in der **Gemarkung Ampermoching** zu einem schwerwiegenden **Verkehrsunfall**. Ein Auto brach in Folge überhöhter Geschwindigkeit in einer lang gezogenen Rechtskurve nach links aus, schleuderte einen Abhang hinab und überschlug sich anschließend mehrfach auf einem angrenzenden Maisfeld. Drei der vier Insassen wurden aus dem Wagen heraus geschleudert und wurden dabei schwer verletzt. Der Fahrer des Wagens, ein 19-jähriger Klempner aus Erding, konnte von der rasch eintreffenden Ortsfeuerwehr zwar noch lebend aus seinem Wagen mit der Rettungsschere geborgen werden, erlag aber seinen schweren Verletzungen auf dem Weg in die Klinik. Der **Unfallwagen**, ein getunter Golf, wurde **vollständig zerstört**. Der Schaden beläuft sich auf ca. 15.000 Euro.

F: In welcher Gemarkung geschah der Verkehrsunfall?

A: **Ampermoching!**

F: Wurde der Unfallwagen völlig zerstört?

A: Ja!

Textverstehen

Faktenbezogene Fragebeantwortung

Am Freitagabend kam es auf der Bundesstraße 69 in der **Gemarkung Ampermoching** zu einem schwerwiegenden **Verkehrsunfall**. Ein Auto brach in Folge überhöhter Geschwindigkeit in einer lang gezogenen Rechtskurve nach links aus, schleuderte einen Abhang hinab und überschlug sich anschließend mehrfach auf einem angrenzenden Maisfeld. Drei der vier Insassen wurden aus dem Wagen heraus geschleudert und wurden dabei schwer verletzt. Der **Fahrer des Wagens**, ein **19-jähriger** Klempner aus Erding, konnte von der rasch eintreffenden Ortsfeuerwehr zwar noch lebend aus seinem Wagen mit der Rettungsschere geborgen werden, erlag aber seinen schweren Verletzungen auf dem Weg in die Klinik. Der **Unfallwagen**, ein getunter Golf, wurde **vollständig zerstört**. Der Schaden beläuft sich auf ca. 15.000 Euro.

F: In welcher Gemarkung geschah der Verkehrsunfall?

A: **Ampermoching!**

F: Wurde der Unfallwagen völlig zerstört?

A: Ja!

F: Wie alt war der Fahrer?

A: **19 Jahre!**

Textverstehen

Faktenbezogene Fragebeantwortung

Am Freitagabend kam es auf der Bundesstraße 69 in der **Gemarkung Ampermoching** zu einem schwerwiegenden **Verkehrsunfall**. Ein **Auto** brach in Folge überhöhter Geschwindigkeit in einer lang gezogenen Rechtskurve nach links aus, schleuderte einen Abhang hinab und überschlug sich anschließend mehrfach auf einem angrenzenden Maisfeld. Drei der vier Insassen wurden aus dem **Wagen** heraus geschleudert und wurden dabei schwer verletzt. Der **Fahrer des Wagens**, ein **19-jähriger Klempner** aus Erding, konnte von der rasch eintreffenden Ortsfeuerwehr zwar noch lebend aus seinem Wagen mit der Rettungsschere geborgen werden, erlag aber seinen schweren Verletzungen auf dem Weg in die Klinik. Der **Unfallwagen**, ein **getunter Golf**, wurde **vollständig zerstört**. Der Schaden beläuft sich auf ca. 15.000 Euro.

F: In welcher Gemarkung geschah der Verkehrsunfall?

A: **Ampermoching!**

F: Wurde der Unfallwagen völlig zerstört?

A: Ja!

F: Wurde das *Fahrzeug* völlig zerstört?

A: Nein! oder Weiß nicht!

F: Wie alt war der Fahrer?

A: **19 Jahre!**

F: War der Fahrer *Handwerker*?

A: **Nein!** oder **Weiß nicht!**



Textverstehen

Inferenzielle Fragebeantwortung

Am Freitagabend kam es auf der Bundesstraße 69 in der Gemarkung Ampermoching zu einem schwerwiegenden Verkehrsunfall. Ein Auto brach **in** Folge überhöhter Geschwindigkeit in einer lang gezogenen Rechtskurve nach links aus, schleuderte einen Abhang hinab und überschlug sich anschließend mehrfach auf einem angrenzenden Maisfeld. Drei der vier Insassen wurden aus dem Wagen heraus geschleudert und wurden dabei schwer verletzt. Der Fahrer des Wagens, ein 19-jähriger **Klempner** aus Erding, konnte von der rasch eintreffenden Ortsfeuerwehr zwar noch lebend aus seinem Wagen mit der Rettungsschere geborgen werden, erlag aber seinen schweren Verletzungen auf dem Weg in die Klinik. Der Unfallwagen, ein getunter Golf, wurde vollständig zerstört. Der Schaden beläuft sich auf ca. 15.000 Euro.

F: War der Fahrer Handwerker?

A: Ja! [Ein Klempner ist (IS-A) ein Handwerker]

Textverstehen

Inferenzielle Fragebeantwortung

Am Freitagabend kam es auf der Bundesstraße 69 in der Gemarkung Ampermoching zu einem schwerwiegenden Verkehrsunfall. Ein Auto brach in Folge überhöhter Geschwindigkeit in einer lang gezogenen Rechtskurve nach links aus, schleuderte einen Abhang hinab und überschlug sich anschließend mehrfach auf einem angrenzenden Maisfeld. Drei der vier Insassen wurden aus dem Wagen heraus geschleudert und wurden dabei schwer verletzt. Der Fahrer des Wagens, ein 19-jähriger Klempner aus Erding, konnte von der rasch eintreffenden Ortsfeuerwehr zwar noch lebend aus seinem Wagen mit der Rettungsschere geborgen werden, erlag aber seinen schweren Verletzungen auf dem Weg in die Klinik. **Der Unfallwagen, ein getunter Golf, wurde vollständig zerstört.** Der Schaden beläuft sich auf ca. 15.000 Euro.

F: War der Fahrer Handwerker?

A: Ja! [Ein Klempner ist (IS-A) ein Handwerker]

F: Wurde das Fahrzeug völlig zerstört?

A: Ja! [ein Golf ist (IS-A) ein Fahrzeug &
ein Fahrzeug ist (IS-A) ein Wagen &
ein Wagen CONTEXT-SYN ein Unfallwagen]

Textverstehen

Inferenzielle Fragebeantwortung

Am Freitagabend kam es auf der Bundesstraße 69 in der Gemarkung Ampermoching zu einem schwerwiegenden Verkehrsunfall. Ein Auto brach in Folge überhöhter Geschwindigkeit in einer lang gezogenen Rechtskurve nach links aus, schleuderte einen Abhang hinab und überschlug sich anschließend mehrfach auf einem angrenzenden Maisfeld. Drei der vier Insassen wurden aus dem Wagen heraus geschleudert und wurden dabei schwer verletzt. Der **Fahrer des Wagens**, ein 19-jähriger **Klempner** aus Erding, **konnte** von der rasch eintreffenden Ortsfeuerwehr zwar **noch lebend** aus seinem Wagen mit der Rettungsschere **geborgen werden, erlag aber seinen schweren Verletzungen auf dem Weg in die Klinik**. Der Unfallwagen, ein getunter Golf, wurde vollständig zerstört. Der Schaden beläuft sich auf ca. 15.000 Euro.

F: War der Fahrer am Unfallort bereits tot?

A: Nein!

[Ereignisablaufferkennung („erlag ... auf dem Weg“;
Commonsense-Prozesslogik zu Unfalltransporten]

Textverstehen

Inferenzielle Fragebeantwortung

Am Freitagabend kam es auf der Bundesstraße 69 in der Gemarkung Ampermoching zu einem schwerwiegenden Verkehrsunfall. Ein Auto brach **in Folge überhöhter Geschwindigkeit** in einer lang gezogenen Rechtskurve nach links aus, schleuderte einen Abhang hinab und überschlug sich anschließend mehrfach auf einem angrenzenden Maisfeld. Drei der vier Insassen wurden aus dem Wagen heraus geschleudert und wurden dabei schwer verletzt. Der **Fahrer des Wagens**, ein 19-jähriger **Klempner** aus Erding, **konnte** von der rasch eintreffenden Ortsfeuerwehr zwar **noch lebend** aus seinem Wagen mit der Rettungsschere **geborgen werden, erlag aber seinen schweren Verletzungen auf dem Weg in die Klinik**. Der Unfallwagen, ein getunter Golf, wurde vollständig zerstört. Der Schaden beläuft sich auf ca. 15.000 Euro.

F: Wer trägt die Unfallschuld?

A: Der Fahrer des Unfallwagens!

Textverstehen

Inferenzielle Fragebeantwortung

Am Freitagabend kam es auf der Bundesstraße 69 in der Gemarkung Ampermoching zu einem schwerwiegenden Verkehrsunfall. Ein Auto brach in Folge überhöhter Geschwindigkeit in einer lang gezogenen Rechtskurve nach links aus, schleuderte einen Abhang hinab und überschlug sich anschließend mehrfach auf einem angrenzenden Maisfeld. Drei der vier Insassen wurden aus dem Wagen heraus geschleudert und wurden dabei schwer verletzt. Der Fahrer des Wagens, ein 19-jähriger Klempner aus Erding, konnte von der rasch eintreffenden Ortsfeuerwehr zwar noch lebend aus seinem Wagen mit der Rettungsschere geborgen werden, erlag aber seinen schweren Verletzungen auf dem Weg in die Klinik. Der Unfallwagen, ein getunter Golf, wurde vollständig zerstört. Der Schaden beläuft sich auf ca. 15.000 Euro.

F: Stürzte der Unfallwagen auf einen Acker?

A: Ja!

[ein Maisfeld ist (IS-A) ein Acker]

[ein Wagen SYN ein Auto &

ein in den Unfall verwickeltes Auto CONT-SYN Unfallwagen]

[aus Straßenverlauf ausbrechen & schleudern &
überschlagen Hat-Resultat/Impliziert stürzen]

Textverstehen

Inferenzielle Fragebeantwortung

Am Freitagabend kam es auf der Bundesstraße 69 in der **Gemarkung Ampermoching** zu einem schwerwiegenden **Verkehrsunfall**. **Ein Auto brach** in Folge überhöhter Geschwindigkeit in einer lang gezogenen Rechtskurve nach links **aus**, **schleuderte** einen Abhang **hinab** und **überschlug sich** anschließend mehrfach **auf einem angrenzenden Maisfeld**. Drei der vier Insassen wurden aus dem Wagen heraus geschleudert und wurden dabei schwer verletzt. Der Fahrer des Wagens, ein 19-jähriger Klempner aus Erding, konnte von der rasch eintreffenden Ortsfeuerwehr zwar noch lebend aus seinem Wagen mit der Rettungsschere geborgen werden, erlag aber seinen schweren Verletzungen auf dem Weg in die Klinik. Der Unfallwagen, ein getunter Golf, wurde vollständig zerstört. Der Schaden beläuft sich auf ca. 15.000 Euro.

F: Stürzte der Unfallwagen auf einen Acker?

A: Ja!

[ein Maisfeld ist (IS-A) ein Acker]

[ein Wagen SYN ein Auto & ist (IS-A) ad-hoc
ein in den Unfall verwickeltes Auto]

[aus Straßenverlauf ausbrechen & schleudern &
überschlagen Hat-Resultat/Impliziert stürzen]

F: In der Nähe welcher Großstadt geschah der Verkehrsunfall?

Textverstehen

Inferenzielle Fragebeantwortung

Am Freitagabend kam es auf der Bundesstraße 69 in der **Gemarkung Ampermoching** zu einem schwerwiegenden **Verkehrsunfall**. **Ein Auto brach** in Folge überhöhter Geschwindigkeit in einer lang gezogenen Rechtskurve nach links **aus**, **schleuderte** einen Abhang **hinab** und **überschlug sich** anschließend mehrfach **auf einem angrenzenden Maisfeld**. Drei der vier Insassen wurden aus dem Wagen heraus geschleudert und wurden dabei schwer verletzt. Der Fahrer des Wagens, ein 19-jähriger Klempner aus Erding, konnte von der rasch eintreffenden Ortsfeuerwehr zwar noch lebend aus seinem Wagen mit der Rettungsschere geborgen werden, erlag aber seinen schweren Verletzungen auf dem Weg in die Klinik. Der Unfallwagen, ein getunter Golf, wurde vollständig zerstört. Der Schaden beläuft sich auf ca. 15.000 Euro.

F: Stürzte der Unfallwagen auf einen Acker?

A: Ja!

[ein Maisfeld ist (IS-A) ein Acker]

[ein Wagen SYN ein Auto & ist (IS-A) ad-hoc
ein in den Unfall verwickeltes Auto]

[aus Straßenverlauf ausbrechen & schleudern &
überschlagen Hat-Resultat/Impliziert stürzen]

F: In der Nähe welcher Großstadt geschah der Verkehrsunfall?

A: München!

[Ampermoching ist Vorort von München]

Textverstehen

Textzusammenfassung

Am Freitagabend kam es auf der Bundesstraße 69 in der Gemarkung Ampermoching zu einem schwerwiegenden Verkehrsunfall. Ein Auto brach in Folge überhöhter Geschwindigkeit in einer lang gezogenen Rechtskurve nach links aus, schleuderte einen Abhang hinab und überschlug sich anschließend mehrfach auf einem angrenzenden Maisfeld. Drei der vier Insassen wurden aus dem Wagen heraus geschleudert und wurden dabei schwer verletzt. Der Fahrer des Wagens, ein 19-jähriger Klempner aus Erding, konnte von der rasch eintreffenden Ortsfeuerwehr zwar noch lebend aus seinem Wagen mit der Rettungsschere geborgen werden, erlag aber seinen schweren Verletzungen auf dem Weg in die Klinik. Der Unfallwagen, ein getunter Golf, wurde vollständig zerstört. Der Schaden beläuft sich auf ca. 15.000 Euro.

[WAS] Bei einem schweren Verkehrsunfall wurde
[WANN] am Wochenende
[WO] auf der Bundesstraße 69 nahe Ampermoching
[WER#1] ein 19-jähriger Klempner aus Erding
[TOPIC#1] getötet.
[WER#2] Die übrigen drei Insassen
[TOPIC#2] überlebten schwer verletzt.
[WER#3] Am Unfallfahrzeug
[TOPIC#3] entstand Totalschaden.

[TOPIC#1] BAD NEWS IS GOOD NEWS!

[TOPIC#2] BAD NEWS IS GOOD NEWS!

[TOPIC#3] BAD NEWS IS GOOD NEWS!

Dimensionen zur Beschreibung natürlicher Sprache

- Sprachspezifisches Wissen (I)
 - Phonologie/Graphematik
 - Laut- und Schriftzeichensystem *Auflistung*
 - Lexikologie *des Vokabulars*
 - Lexemsystem (Wörter, Kollokationen)
 - Morphologie
 - Flexion: Kombination des Lexemsystems mit grammatischen Informationen (Genus, Numerus, Kasus, Tempus, Modus) *Wort-*
 - Derivation, Komposition: lexikalisch- *grammatik* 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

*Satz-
grammatik*

*Syntax-Semantik-
Interface*

*Text-
grammatik
Diskurs-
logik*

Dimensionen zur Beschreibung natürlicher Sprache

- (Sprachunabhängiges) „Welt“wissen
 - Nicht-sprachliches Alltags- und Fachwissen
 - Der Kassierer **öffnete den Safe**. Er kannte den **Code**.
 - Die **Sopranistin** war bei der **Premiere** indisponiert. Keine Hand rührte sich **am Ende** zum **Applaus**.
 - Der Tumor hat ein **Grading von 3** und ein **Staging von 6**. Es wird empfohlen, eine **hochintensive Chemotherapie** einzuleiten.

*semantisches vs.
enzyklopädisches
Wissen*

Dimensionen zur Beschreibung natürlicher Sprache

- (Sprachunabhängiges) „Welt“wissen
 - Nicht-sprachliches Alltags- und Fachwissen
 - Der Kassierer **öffnete den Safe**. Er kannte den **Code**.
 - Die **Sopranistin** war bei der **Premiere** indisponiert. Keine Hand rührte sich **am Ende** zum **Applaus**.
 - Der Tumor hat ein **Grading von 3** und ein **Staging von 6**. Es wird empfohlen, eine **hochintensive Chemotherapie** einzuleiten.
 - Inferenzmuster (Folgern über Wissen)
 - deduktiv, induktiv, abduktiv

Inferenzmuster

- deduktives Schließen

$$\frac{\forall x : P(x) \rightarrow Q(x) \quad P(A)}{Q(A)}$$

Inferenzmuster

- deduktives Schließen (klassische Logik)

$$\frac{\forall x : P(x) \rightarrow Q(x) \quad P(A)}{Q(A)}$$

Inferenzmuster

- deduktives Schließen (klassische Logik)

$$\frac{\forall x : P(x) \rightarrow Q(x) \quad P(A)}{Q(A)}$$



- abduktives Schließen (Diagnostik ...)

$$\frac{\forall x : P(x) \rightarrow Q(x) \quad Q(A)}{P(A)}$$



Inferenzmuster

- induktives Schließen

$$P(A_1) \rightarrow Q(A_1)$$

$$P(A_2) \rightarrow Q(A_2)$$

...

$$P(A_n) \rightarrow Q(A_n)$$

$$\forall x : P(x) \rightarrow Q(x)$$



Natürlichsprachliche Bezüge auf nicht-klassische Rechensysteme

- **Unvollständigkeit**
 - Defaults
 - „Alle Vögel können fliegen – bis auf Strauße ...“
 - Inkonsistenz
 - „Hans studiert Informatik“ – „Nein, Hans studiert Physik“
- **Ungenauigkeit**
 - „Ich meine mich zu erinnern, dass Hans Physik studiert“
 - „Es scheint wirklich so, dass Hans eher Informatik studiert“
- **Vagheit**
 - „das ist aber ein großes Auto“
 - „Physik ist schwerer zu studieren als Informatik“

Dimensionen zur Beschreibung natürlicher Sprache

- (Sprachunabhängiges) „Welt“wissen
semantisches vs. enzyklopädisches Wissen
 - Nicht-sprachliches Alltags- und Fachwissen
Wissen
 - Der Kassierer **öffnete den Safe**. Er kannte den **Code**.
 - Die **Sopranistin** war bei der **Premiere** indisponiert. Keine Hand rührte sich **am Ende** zum **Applaus**.
 - Der Tumor hat ein **Grading von 3** und ein **Staging von 6**. Es wird empfohlen, eine **hochintensive Chemotherapie** einzuleiten.
 - Inferenzmuster (Folgern über Wissen)
„Rechnen“ mit unterschiedlichen Wissenstypen
 - deduktiv, induktiv, abduktiv
 - Unvollständigkeit: Defaults, Inkonsistenz
 - Ungenauigkeit, Vagheit

Dimensionen zur Beschreibung natürlicher Sprache

- **Abstrakte Problemlösungstechniken: sprachunabhängige Prozessmodelle (mit differenzierter Algorithmik in der Informatik!)**
 - **Planen**
 - Vortrag, Referat, Text, Diskussionsbeitrag
 - **Suchen**
 - Lexikalischer Zugriff
 - **Lernen**
 - Lexikon- und Grammatikerwerb (Muttersprache, L2, L3,...)
 - Neologismen in Erstsprache
 - Fachsprachen (der Linguistik, Medizin, ...)

Dimensionen zur Beschreibung natürlicher Sprache

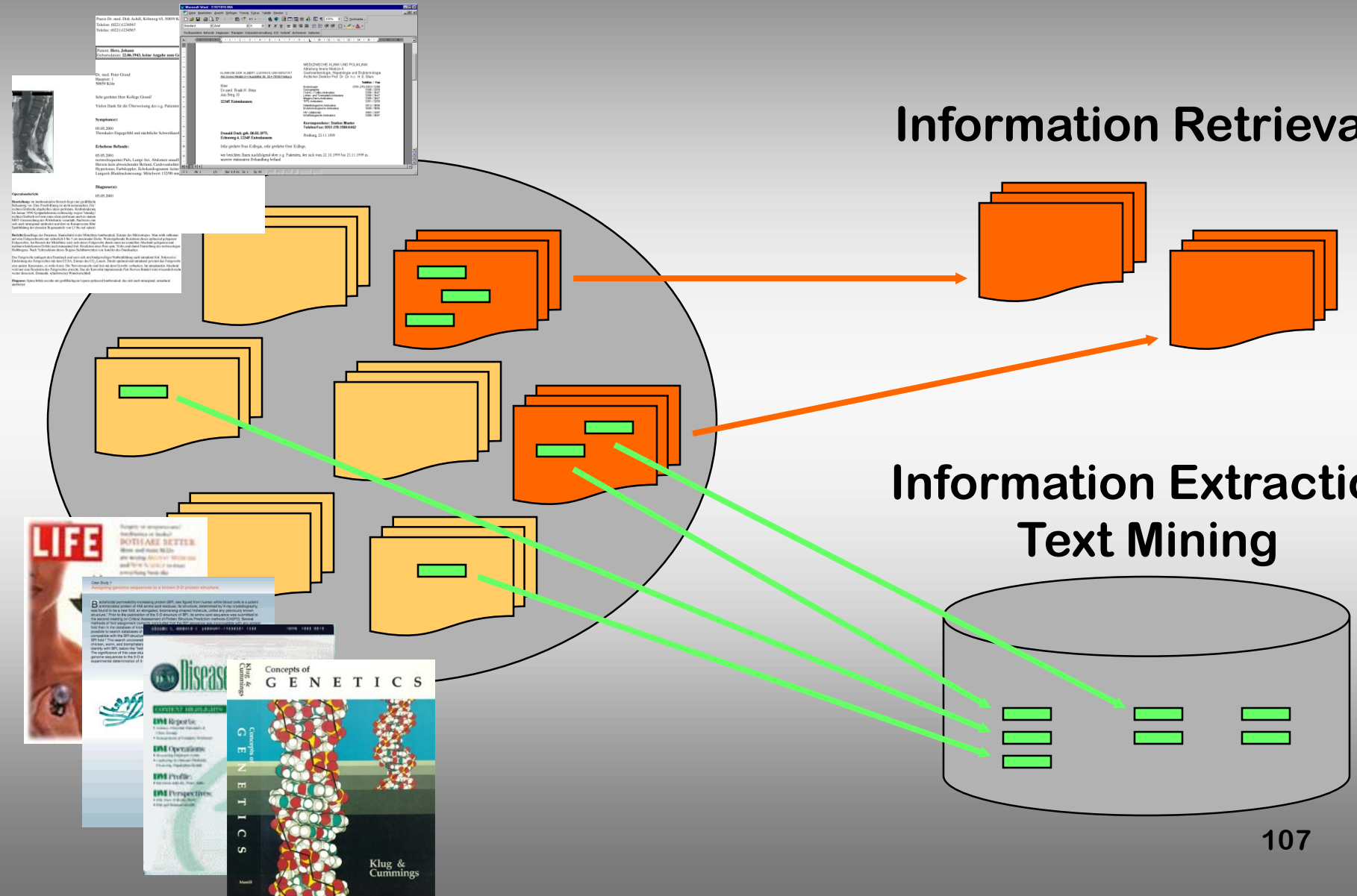
- Funktionaler Sprachgebrauch
 - Suche nach relevanten Texten
 - Faktenextraktion aus Texten
 - Textzusammenfassung
 - Übersetzung
 - Frage-Beantwortung

*Explizierung
menschlichen
Sprachverstehens*

Information Retrieval & Extraction

Information Retrieval

Information Extraction
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

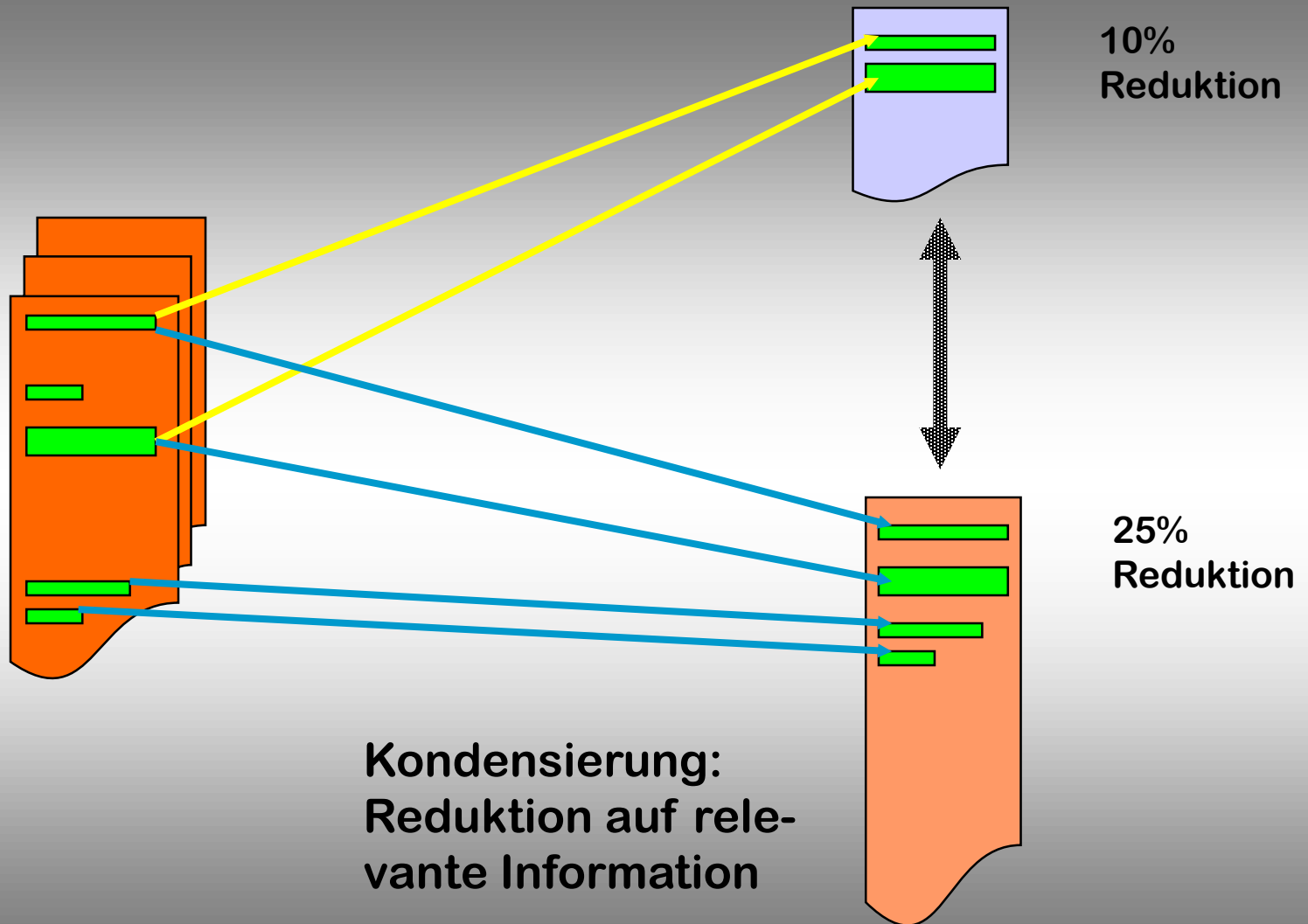
Effective-for: Thalidomide, cutaneous manifestations of leprosy

Disease: leprosy-associated fever

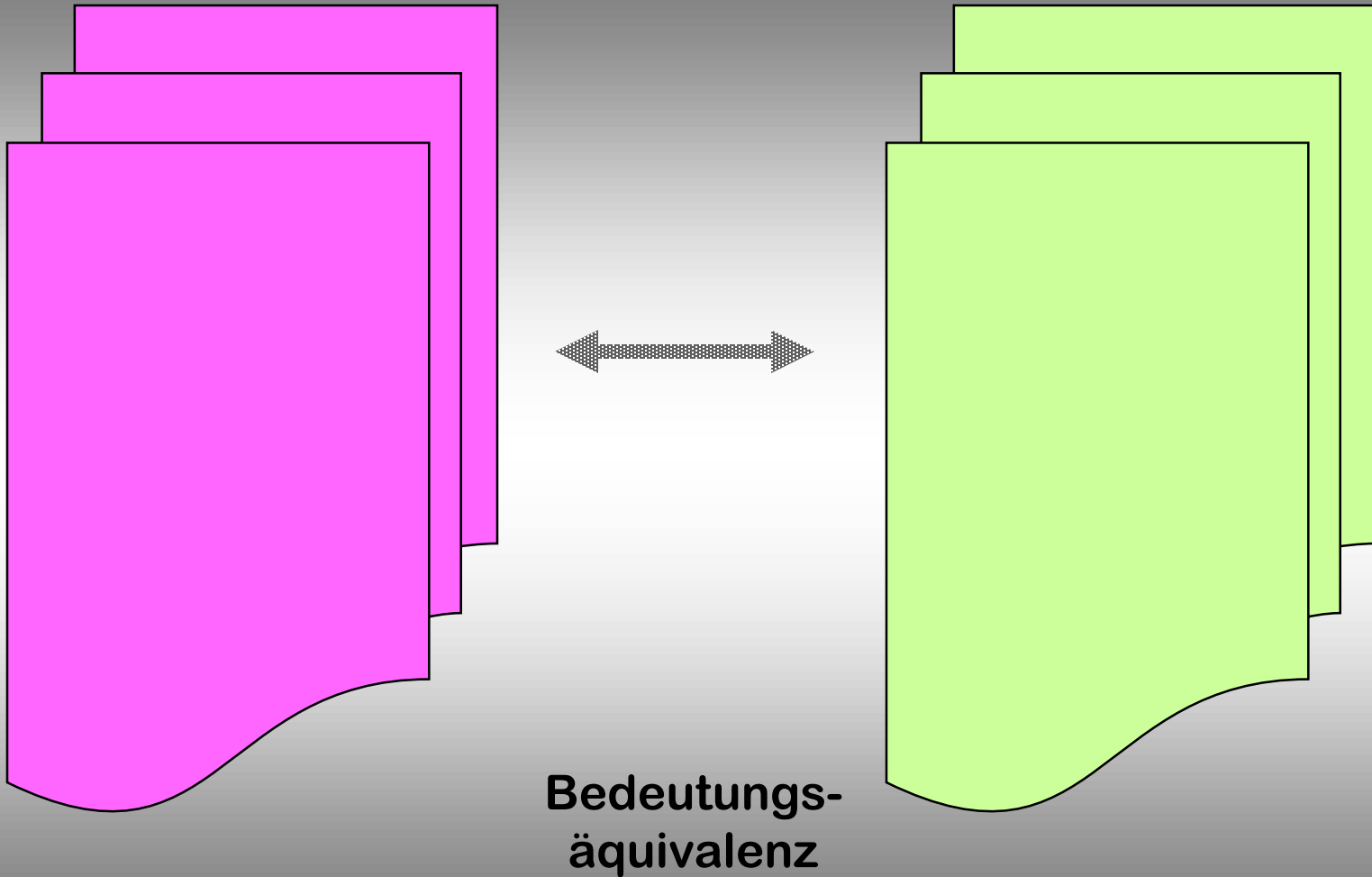
Drug: Thalidomide, Aspirin

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

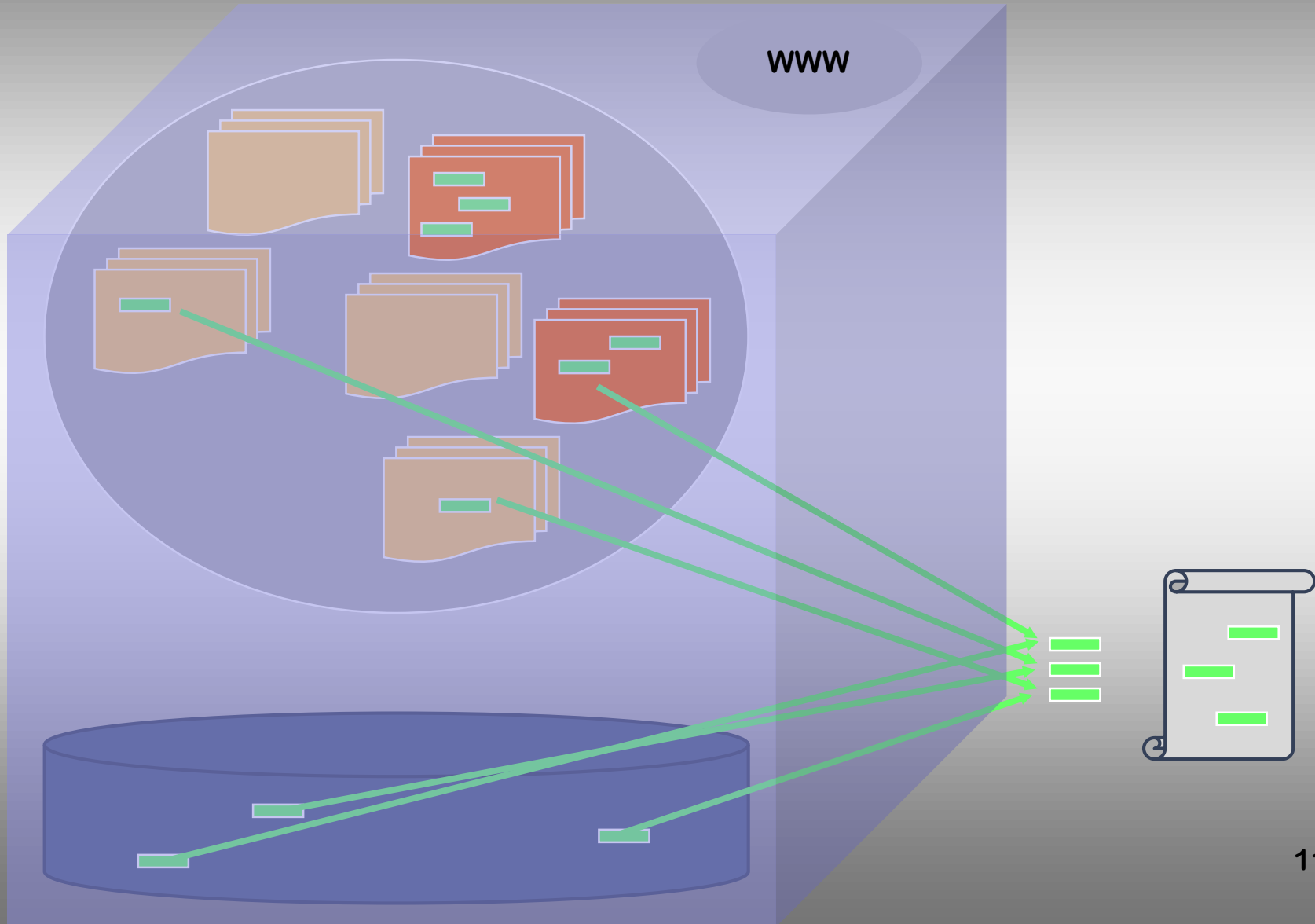
Textzusammenfassung



Maschinelle Übersetzung



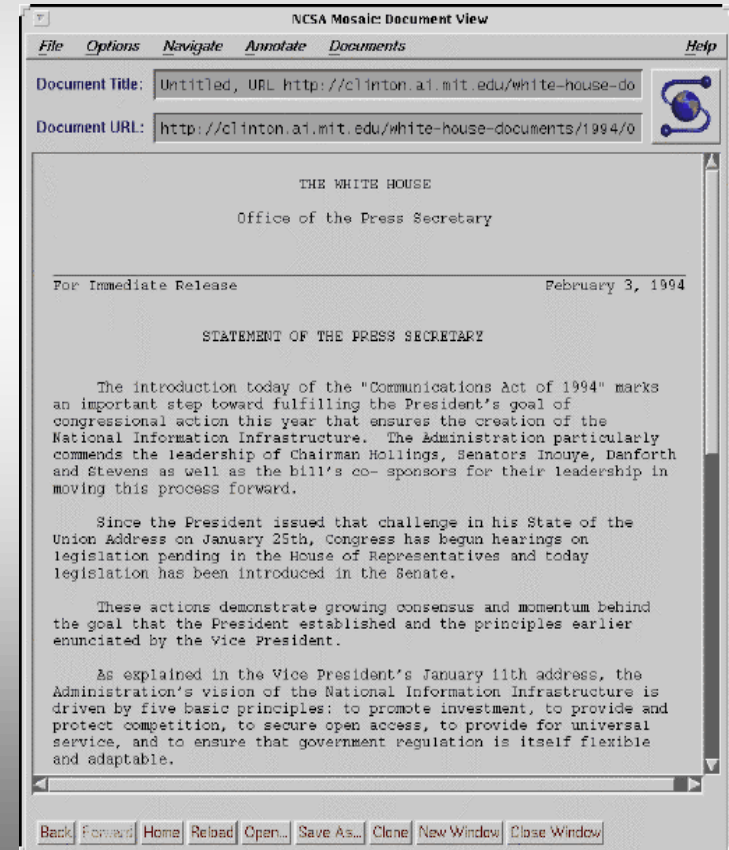
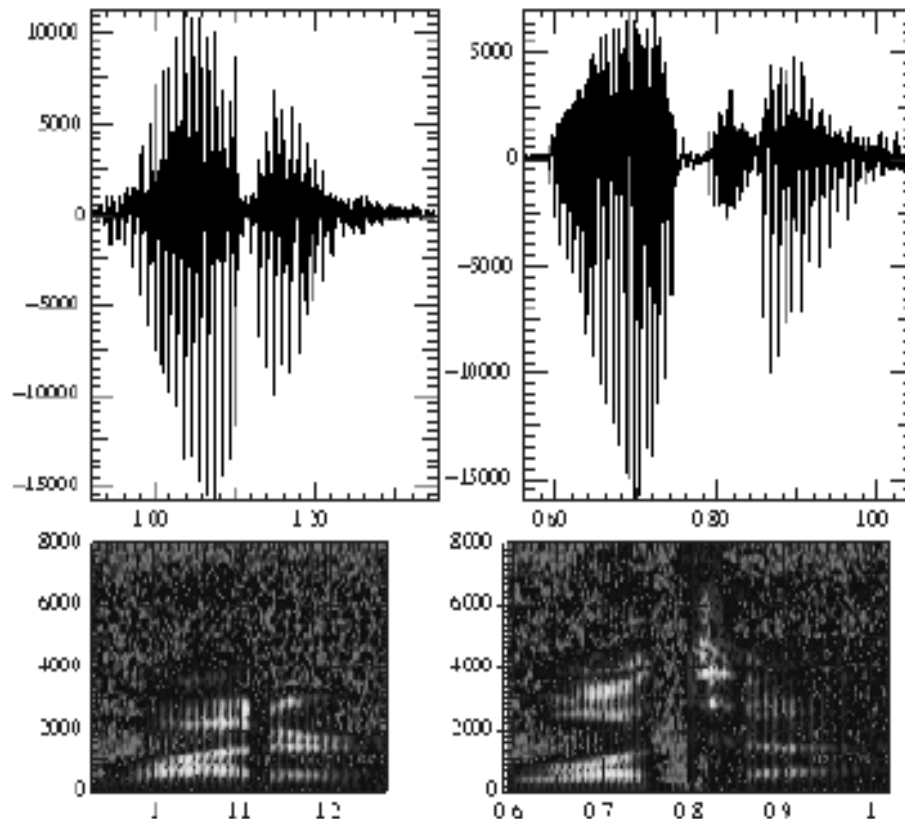
Frage-Beantwortung



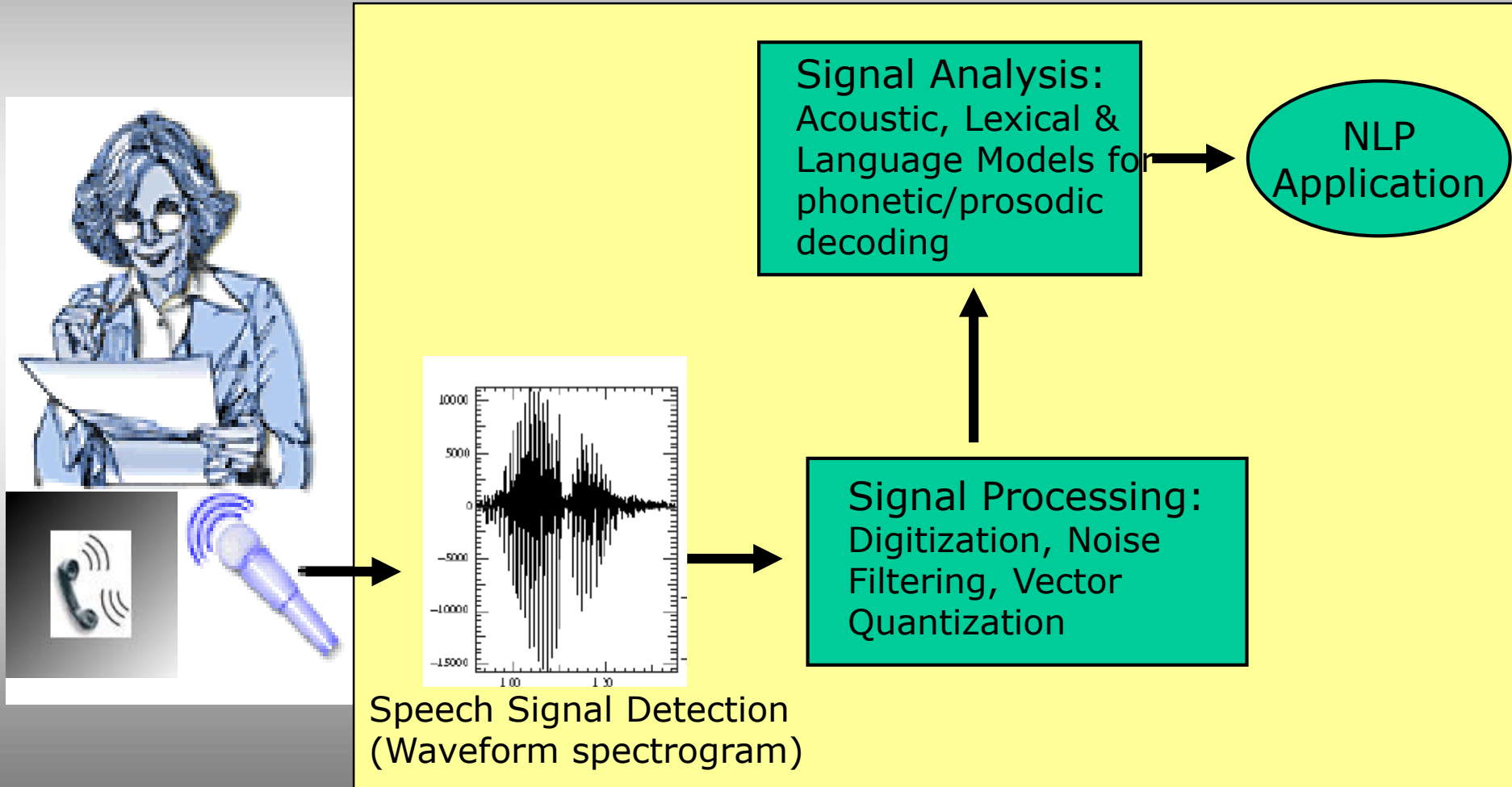
Die zwei Modi natürlicher Sprache

... in *gesprochener* Form
(Spontansprache)

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



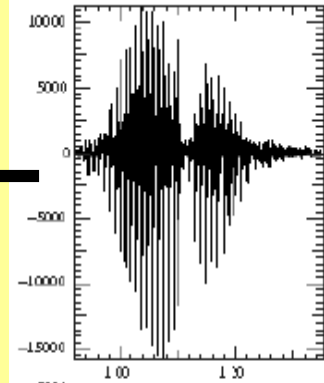
Speech2Text



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 Simulationsperspektive (Primaten)

KodE Alltag

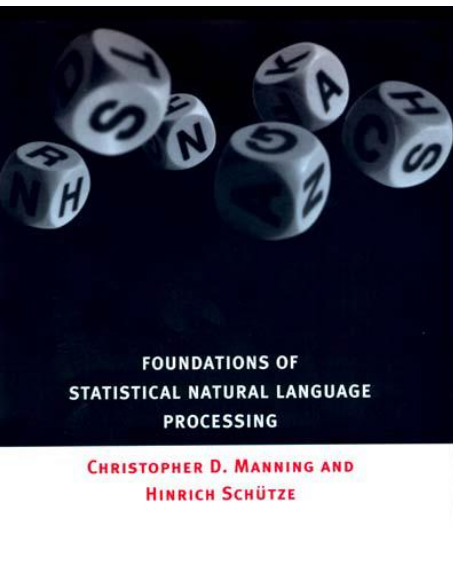
- Aufbau eines deutschen E-Mail-Korpus
- Spende einer persönlichen E-mail an
 - kodealltag@aau.at
- Anonymisierung aller persönlichen Daten
 - E-Mail-Adressen
 - Namen
 - Ortsangaben
- Explizite Zustimmung nötig – Fragebogen

Literatur

- D. Jurafsky & J.A. Martin (2000), *Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice Hall
- C.D. Manning & H. Schütze (1999), *Foundations of Statistical Natural Language Processing*. MIT Press.
- R. Mitkov (Ed.) (2003), *The Oxford Handbook of Computational Linguistics*. Oxford University Press,
- K.-U. Carstensen, Ch. Ebert, C. Endriss, S. Jekat, R. Klabunde & H. Langer (Eds.) (2004, 2nd ed.), *Computerlinguistik und Sprachtechnologie*. Elsevier – Spektrum Akademischer Verlag

Literatur

- D. Jurafsky & J.A. Martin (2000), *Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice Hall
- C.D. Manning & H. Schütze (1999), *Foundations of Statistical Natural Language Processing*. MIT Press.



... (Ed.) (2003), *The Oxford Handbook of Computational Linguistics*. Oxford University Press

...rstensen, Ch. Ebert, C. Endriss
...e & H. Langer (Eds.) (2004, 2nd ed.)
Interlinguistik und Sprachtechnologie
Akademischer Verlag

