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.Exzellenz initiative (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.Exzellenz-initiative (2010-12)

http://de.wikipedia.org/wiki/Exzellenzinitiative

87 deutsche Universitäten

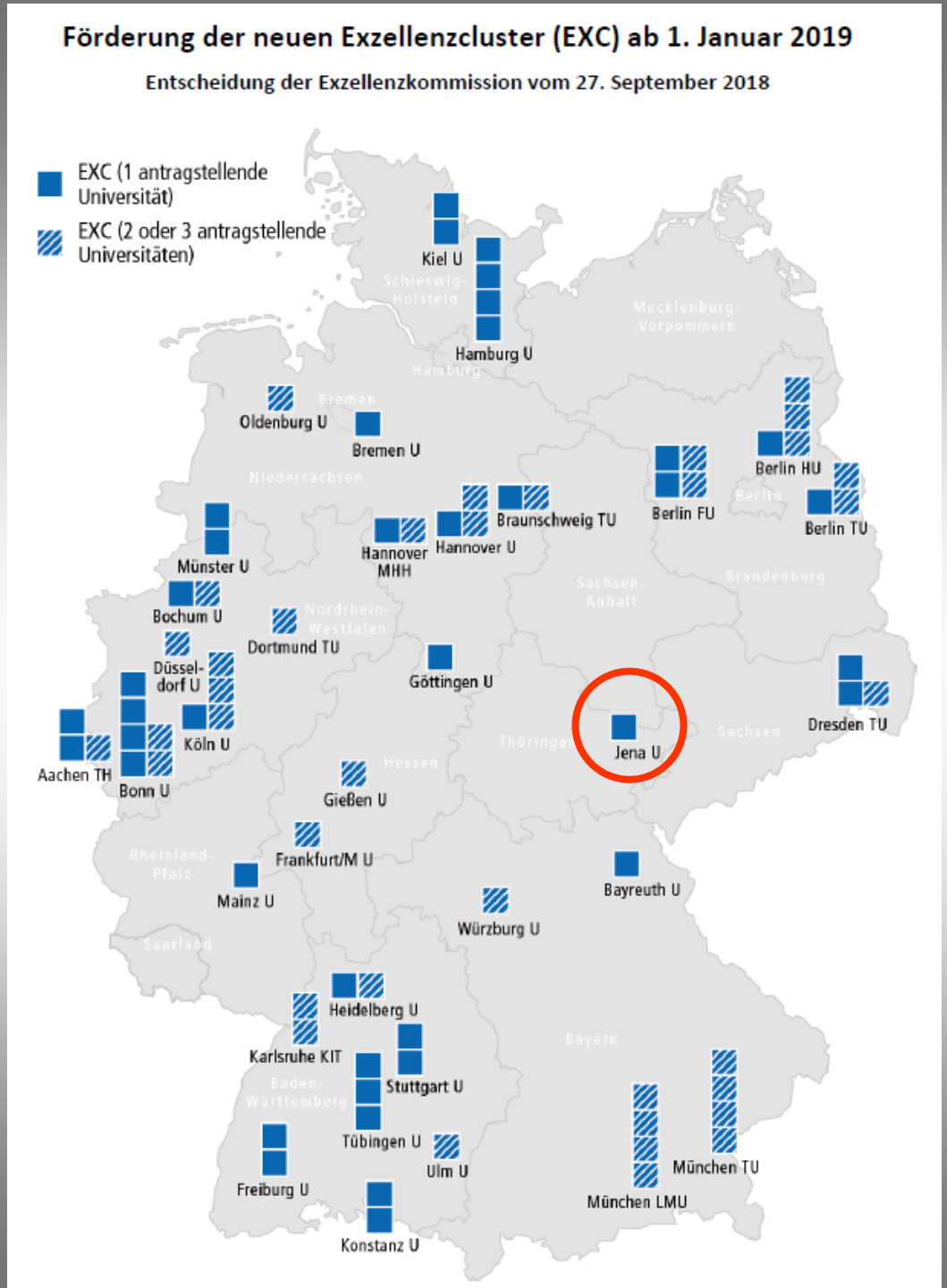
11+45+43 = 89 in Förderlinien

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.Exzellenz-initiative (2017-22)



Woher kommt Exzellenz ?

- (High-impact-)Publikationen
- Wissenschaftspreise
- Drittmitteleinwerbungen
  - SFBs, Graduiertenschulen …
- Zukunftsentwürfe

- Im internationalen Kontext weltweit sichtbar sein (visibility)

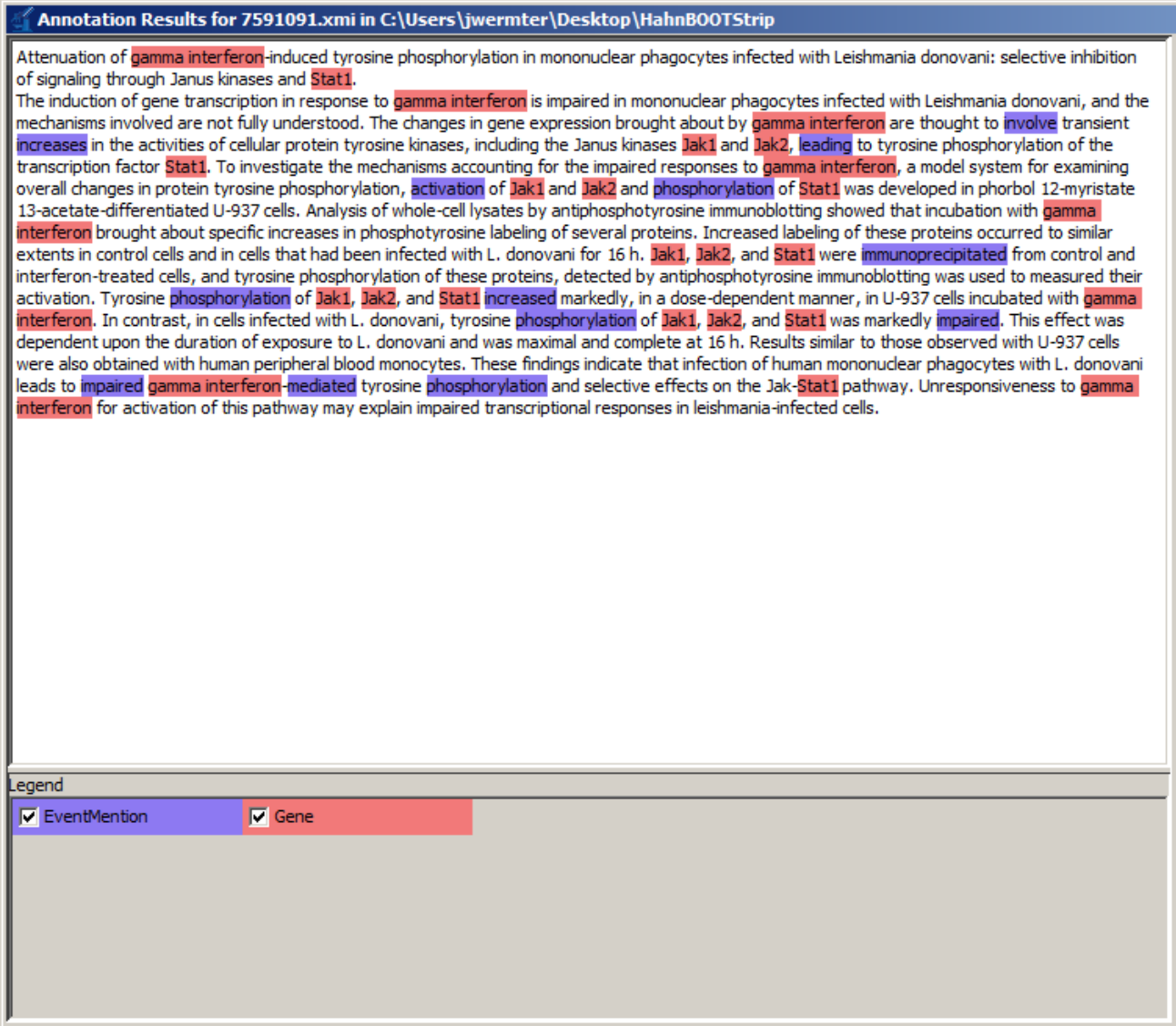
Ein Beispiel für den Nachweis wissenschaftlicher Exzellenz

- 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) ?

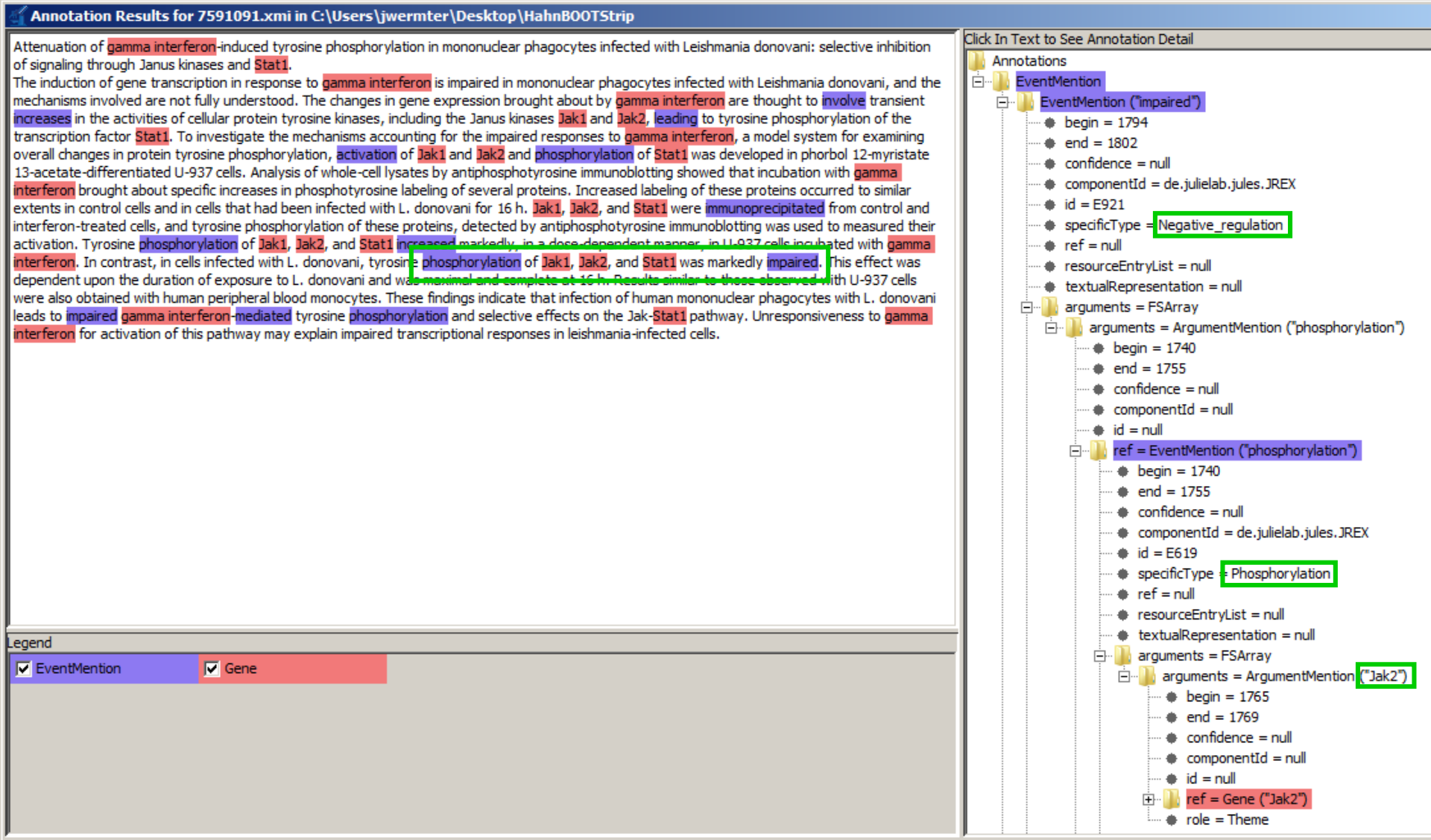
<Semantische Textanalytik>

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

Propositionale Semantik



Propositionale Semantik



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

http://www.duckipedia.de/images/3/39/Tick,_Trick_und_Track_.jpg

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

Team	gold (match)	answer (match)	recall	precision	fscore
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

Team	gold (match)	answer (match)	recall	precision	fscore
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

Team	gold (match)	answer (match)	recall	precision	fscore
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

… so erarbeitet man sich Forschungsexzellenz !

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 Methodenkonservativismus‘
- Exemplarischer Fall empirischer Wissenschaft: Rationalität

</Exkurs>

Computerlinguistik I

- Linguistik: Gegenstandsbereich sind (überwiegend) natürliche Sprachen
  - Deutsch, Englisch, Französisch, ...
- Beispiele für formale Sprachen
  - $L = \{a^n b^n, n \in \mathbb{N}\}$

    $= \{ab, aabb, aaabbb, aaaabbbb, \ldots\}$
  - 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

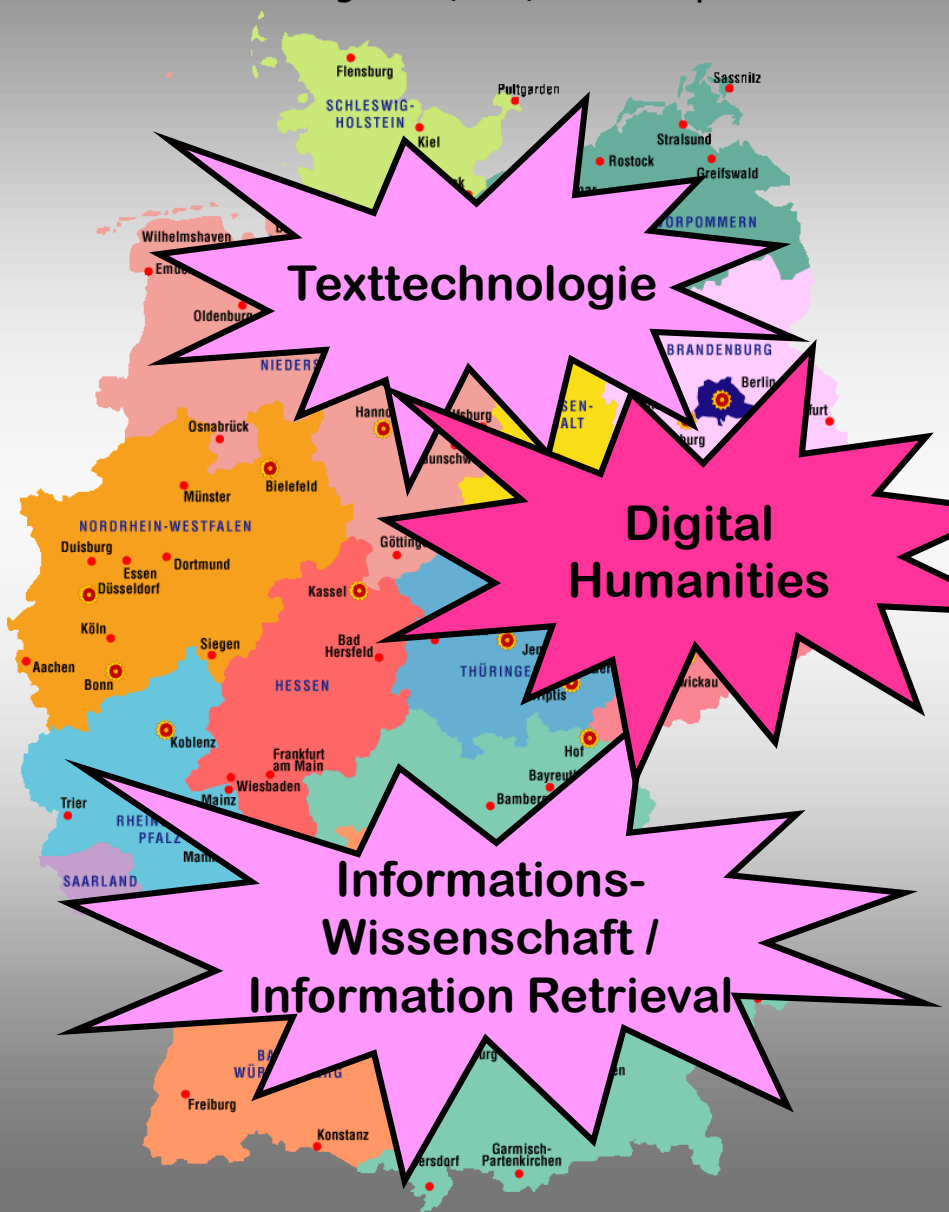
www.ims.uni-stuttgart.de/info/SitesEurope.html#Germany

- 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]

Computerlinguistik-Standorte

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]

TU Darmstadt (2)
U Frankfurt/M. (2)
U Leipzig

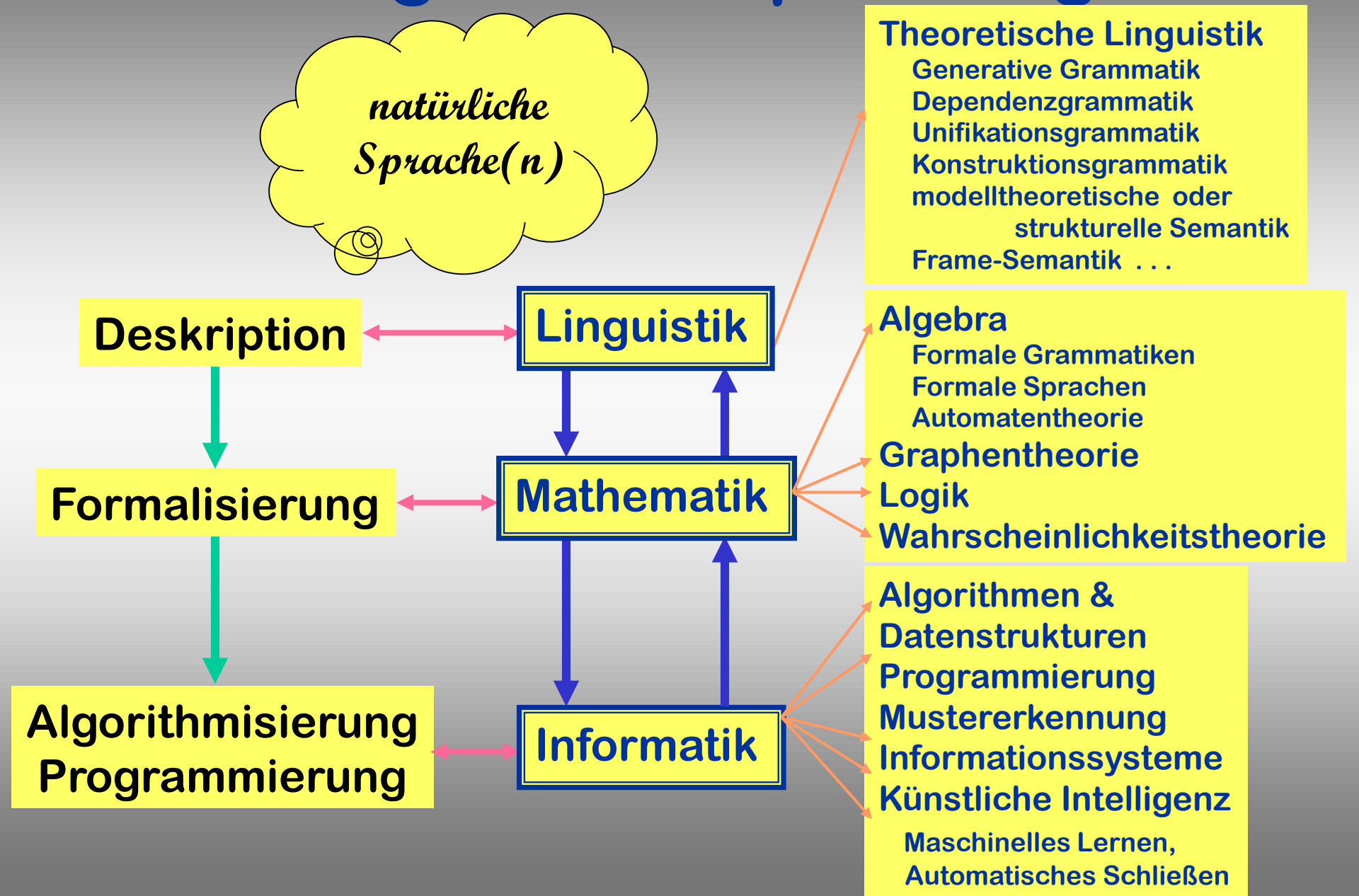
+23 [25]

U Bamberg
U Köln
U Passau
U Jena
HU Berlin
U Stuttgart
U Konstanz
U Dortmund
U Kassel
U Würzburg
U Göttingen
U Münster

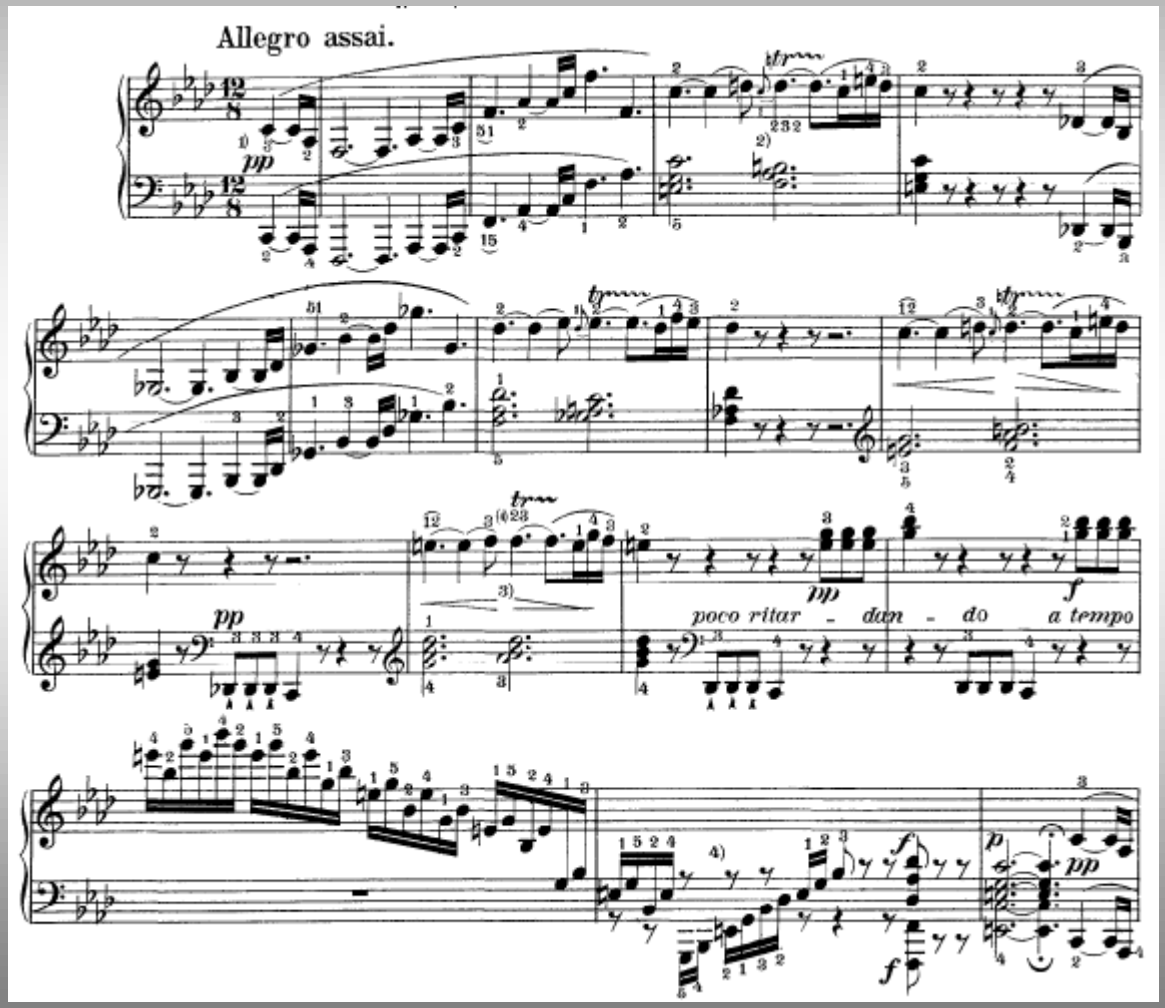
http://www.dig-hum.de/

U Regensburg
U Hildesheim
U Düsseldorf
U Dortmund
BU Weimar
U Bamberg
U Kaiserslautern
TU Dresden

Verortung der Computerlinguistik

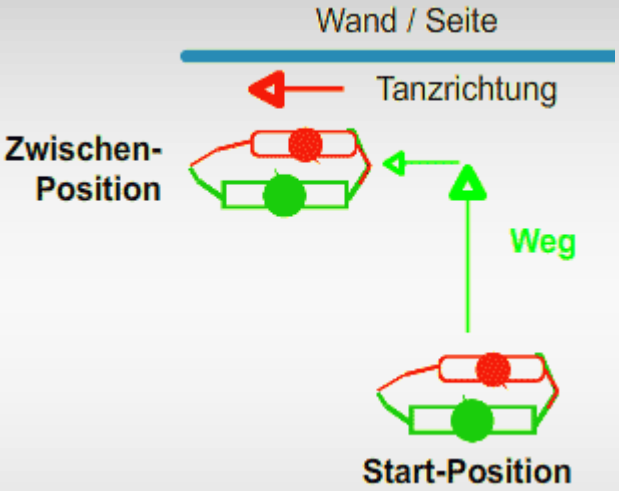


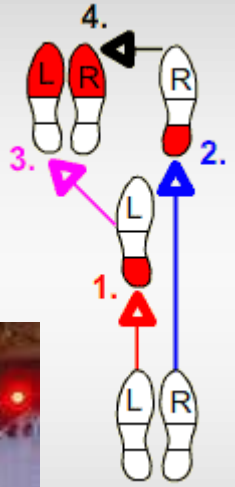
Keine natürlichen, aber doch auch Sprachen (1/6)



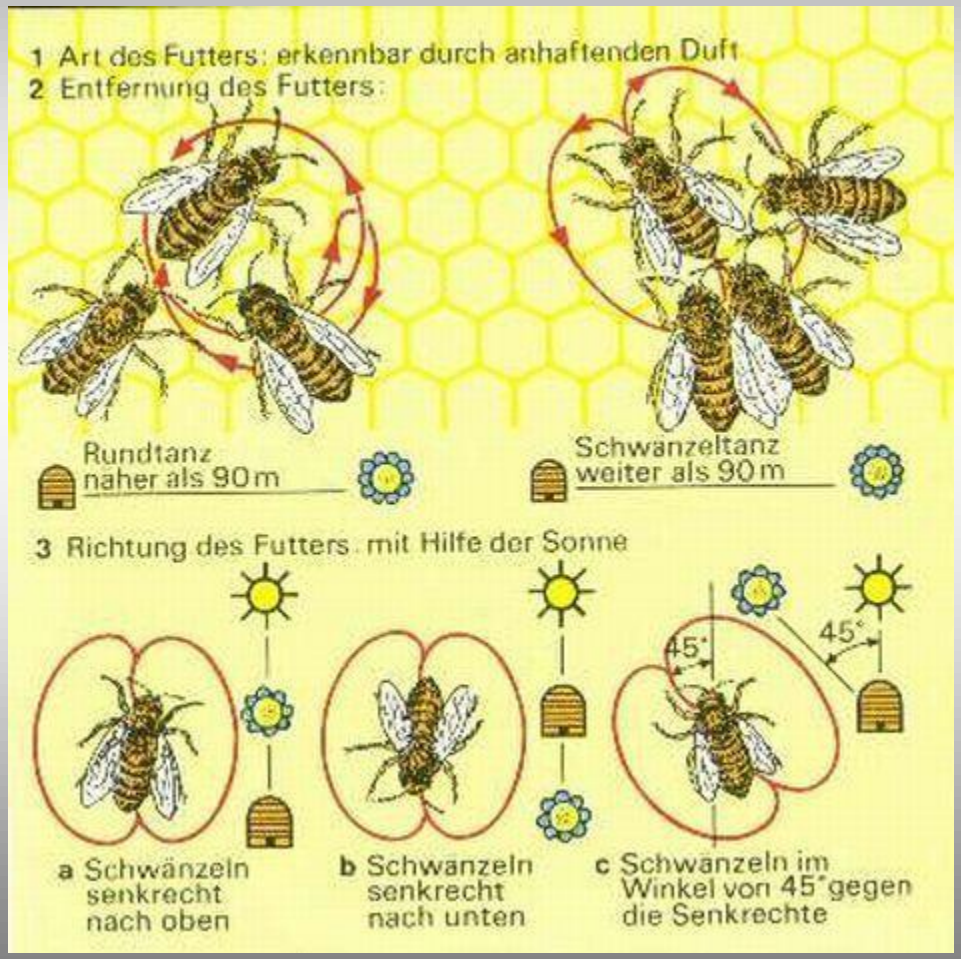
Takt 1-16 aus Beethovens op. 57 (Appassionata)

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

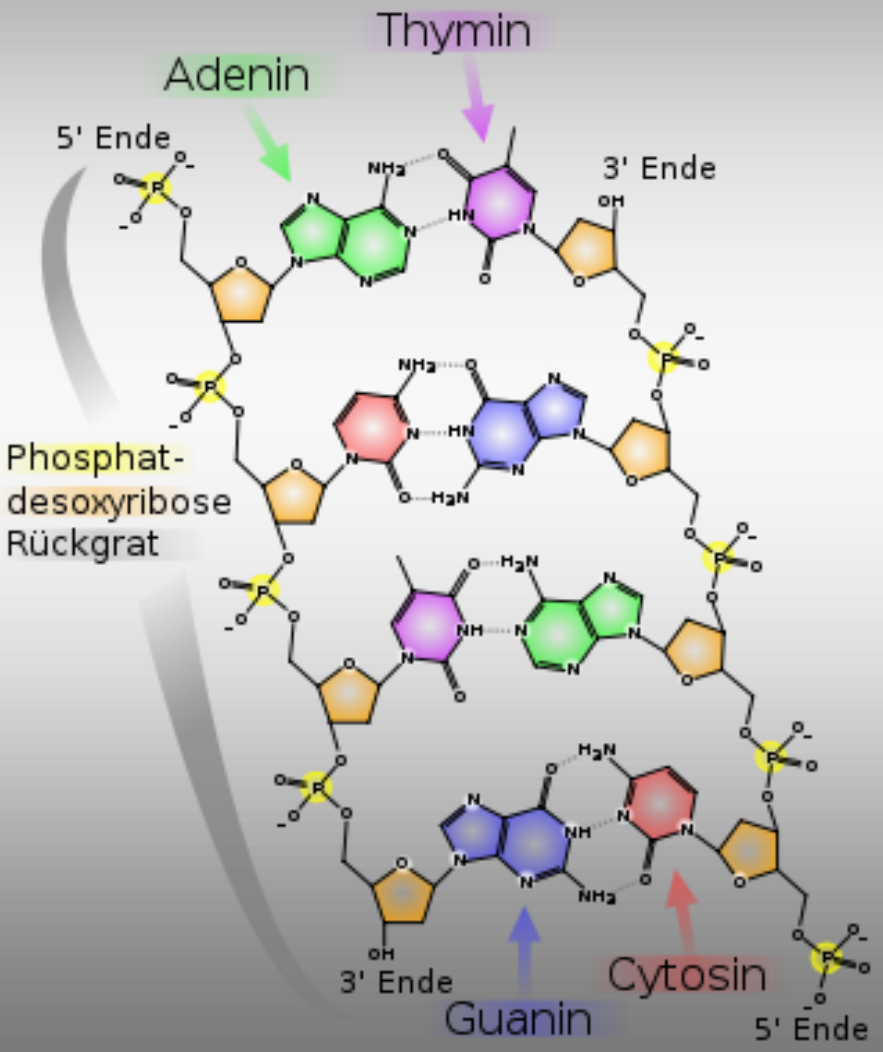


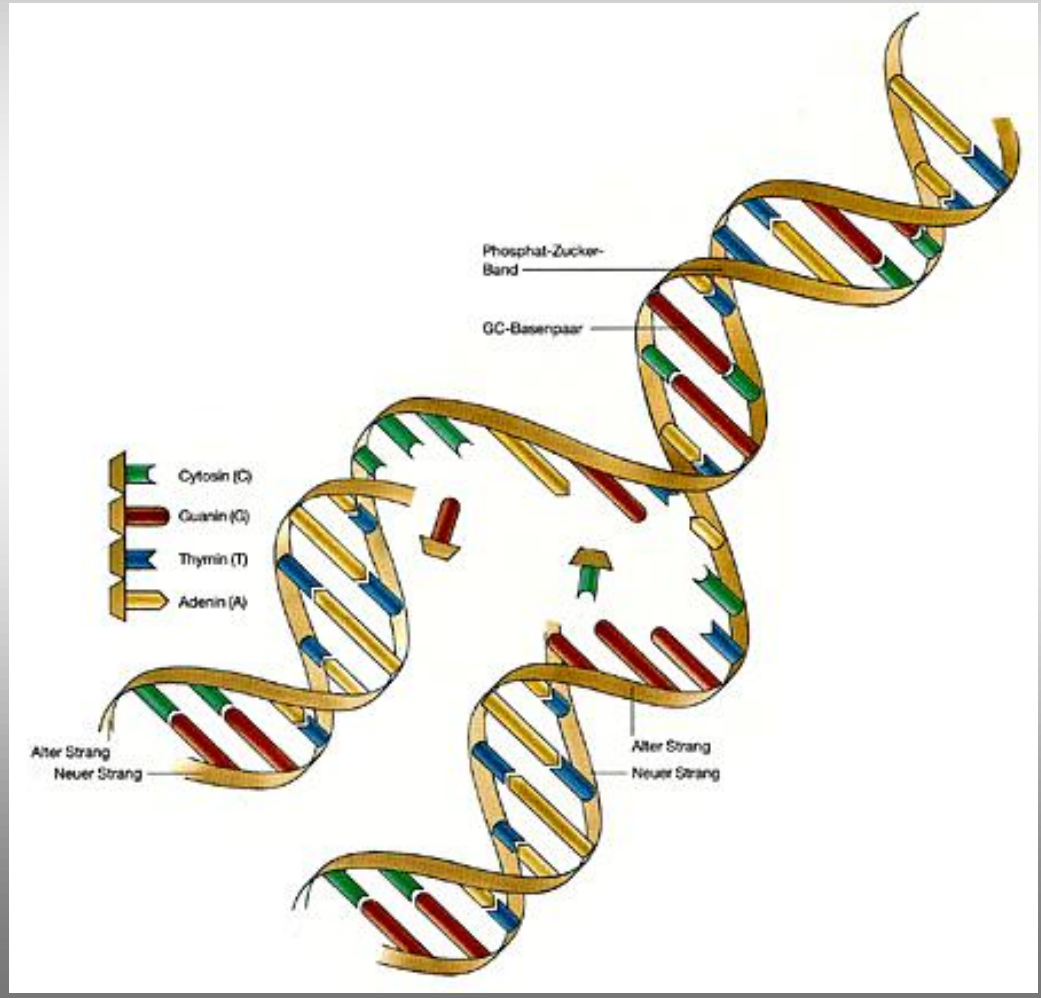


Keine natürlichen, aber doch auch Sprachen (3/6)



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

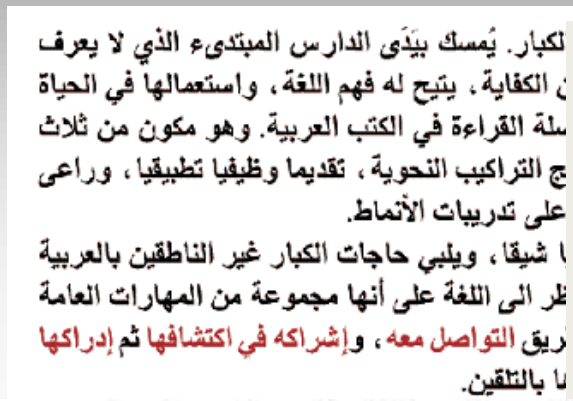
Natürliche Sprache
verschiedene Approximationsstufen

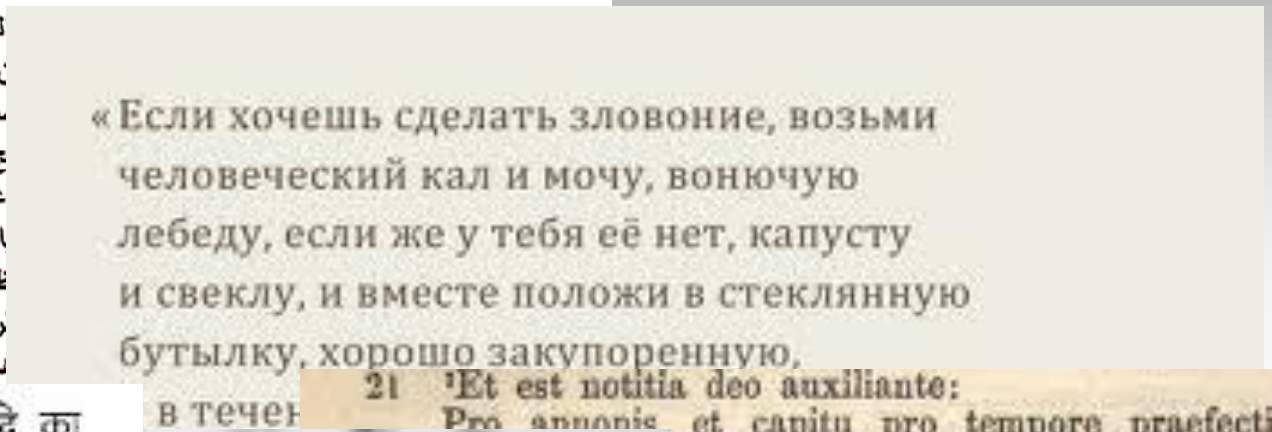
- hciltsinatsemdnEre!eSgnaf
- fan gSe! erEn dmest anis tlich
- fangSe!erEndmestanistlich
- fang Se ! er End mest an ist lich
- Endlich ist Semesteranfang!

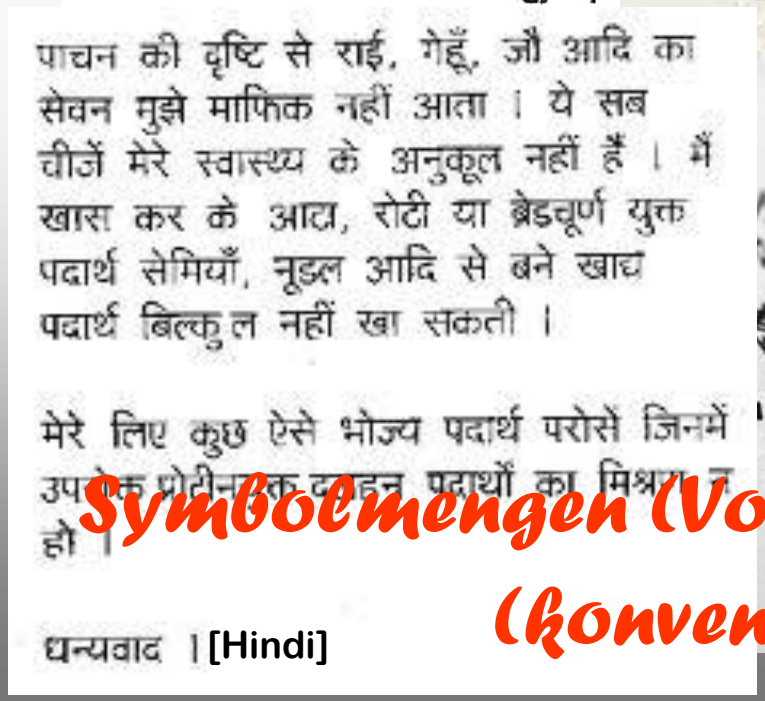
Symbolmengen (Vokabular)

(konventionelle) lineare Reihung

Natürliche Sprachen – verschiedene Schriftarten


لكبار. يُمسك بيَدَى الدارس المبتدىء الذي لا يعرف
ن الكفاية، يتيح له فهم اللغة، واستعمالها في الحياة
سلة القراءة في الكتب العربية. وهو مكون من ثلاث
ج التراكيب النحوية، تقديما وظيفيا تطبيقيا، وراعى
على تدريبات الأنماط.
ا شيقا، ويلبي حاجات الكبار غير الناطقين بالعربية
ظر الى اللغة على أنها مجموعة من المهارات العامة
ريق التواصل معه، وإشراكه في اكتشافها ثم إدراكها
ا بالتلقين.


«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,
в течен


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 प्रोटीन-युक्त दूसरे पदार्थों का मिश्रण न
हो ।

धन्यवाद । [Hindi]


是那個記得當時年
啊

21 ¹Et est notitia deo auxiliante:
Pro annonis et capitu pro tempore praefecti
m Africam auri libras centum.
consiliariorum auri libras viginti.
cancellariorum auri libras septem.
ius ita:
mo hominibus decem pro annonis
pitu XIIS, 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 capitum I capitus solido-
aliquis quattuor
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 – reden [V]
  - redest
  - reden
  - redet
  - Rede – Rede [N]
  - Reden
  - Redner – Redner [N]
  - Redners

Natürliche Sprache
Linguistische Ebenen: Lexikologie

- **Vollformen** **Grundformen** **Stämme**
  - rede – reden [V] – RED
  - redest
  - reden
  - redet
  - Rede – Rede [N]
  - Reden
  - Redner – Redner [N]
  - Redners

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]?
- * Schrieb er ein [es]?

Gruppierung (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)

- 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-semantisch

Wörter können nicht erschöpfend „aufgezählt" werden

- Derivation
  - 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
  - dieses interessante BuchØ
  - die neuen Bücher
- 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.

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-semantisch

- 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