Universität Hamburg • Institut für Germanistik I • Sommersemester 2006

Vorlesung
Computerphilologie

Themenfeld
Allgemeine und theoretische Grundlagen

„Was macht die CP aus theoretischer Sicht und wie steht sie zu den Text-Wissenschaften (Philologien)?“

Hat CP eine Theorie?

Nein

Arbeitsdefinition
(angelehnt an die Definition der Arbeitsstelle)

- Computerphilologie ist eine interdisziplinär ausgerichtete Hilfswissenschaft, die primär textbezogene philologische Fragestellungen mit methodologisch ausgewiesenen Verfahren einer rechnergestützten Repräsentation, Modellierung und Auswertung von Daten unterstützt.

- Strittig dabei ist die Zuschreibung "Hilfswissenschaft": Man kann der Meinung sein, dass die CP
  - Eine Hilfswissenschaft,
  - eine autonome Wissenschaft, oder aber
  - gar keine Wissenschaft, sondern eine Technologie
  - gar keine Wissenschaft, sondern eine Heuristik

  sei.

CP nach Jannidis

- „Unter dem Etikett „Computerphilologie“ soll also das Wissen um die Einsatzmöglichkeiten des Computers in der Literaturwissenschaft gesammelt werden. Insbesondre gehören dazu das Erstellen und Verwenden elektronischer Texte, einschließlich der Lektüre und des Information Retrievals, die Hypertexttheorie und -praxis mit Berücksichtigung von Hyperfiction und das Programmieren von Anwendungen für Literaturwissenschaftler.“
- „Analog dazu *(Algorithmen und Datenstrukturen in der Informatik, v.H.)* ist es Aufgabe einer Computerphilologie, solche dauerhaften Prinzipien zu ermitteln, zusammen zu stellen und zu tradieren.“

Jannidis‘ Position

- Literaturwissenschaftlich
- Stark editionstechnisch
- Abstraktionsorientiert / wissenschaftlich
- Wenig praktisches Anwendungswissen
- Kontrastiv oder komplementär zur Computerlinguistik
- Teilgebiet von Humanities Computing, aber ohne dessen Werkzeug-Sicht.

Wissenschaft oder Hilfswissenschaft?

- Eine Wissenschaft hat genuine Methoden der Klassen
  - Deskriptiv, Distributionsanalyse: Dies ist ein Phonem
  - Normativ und Jedes Wort muß mindestens ein Stammmorphem haben
  - Rekonstruktiv. Eine Sprache ist ein Quintupel {...}
- Sie hat einen (operationalen) Apparat zur Aufstellung von Hypothesen, zur Beschreibung von Sachverhalten und zur Bildung von Theorien.
- Die Computerphilologie ist eine typische Hilfswissenschaft, die
  - rationale Verfahren oder
  - Heuristiken

  für die o.g. Aufgaben einer (anderen) Wissenschaft zur Verfügung stellt.
- Die Verfahren der CP sind im Aufbau und bedürfen weiterer Erforschung.

Beispiele für Methoden der CP

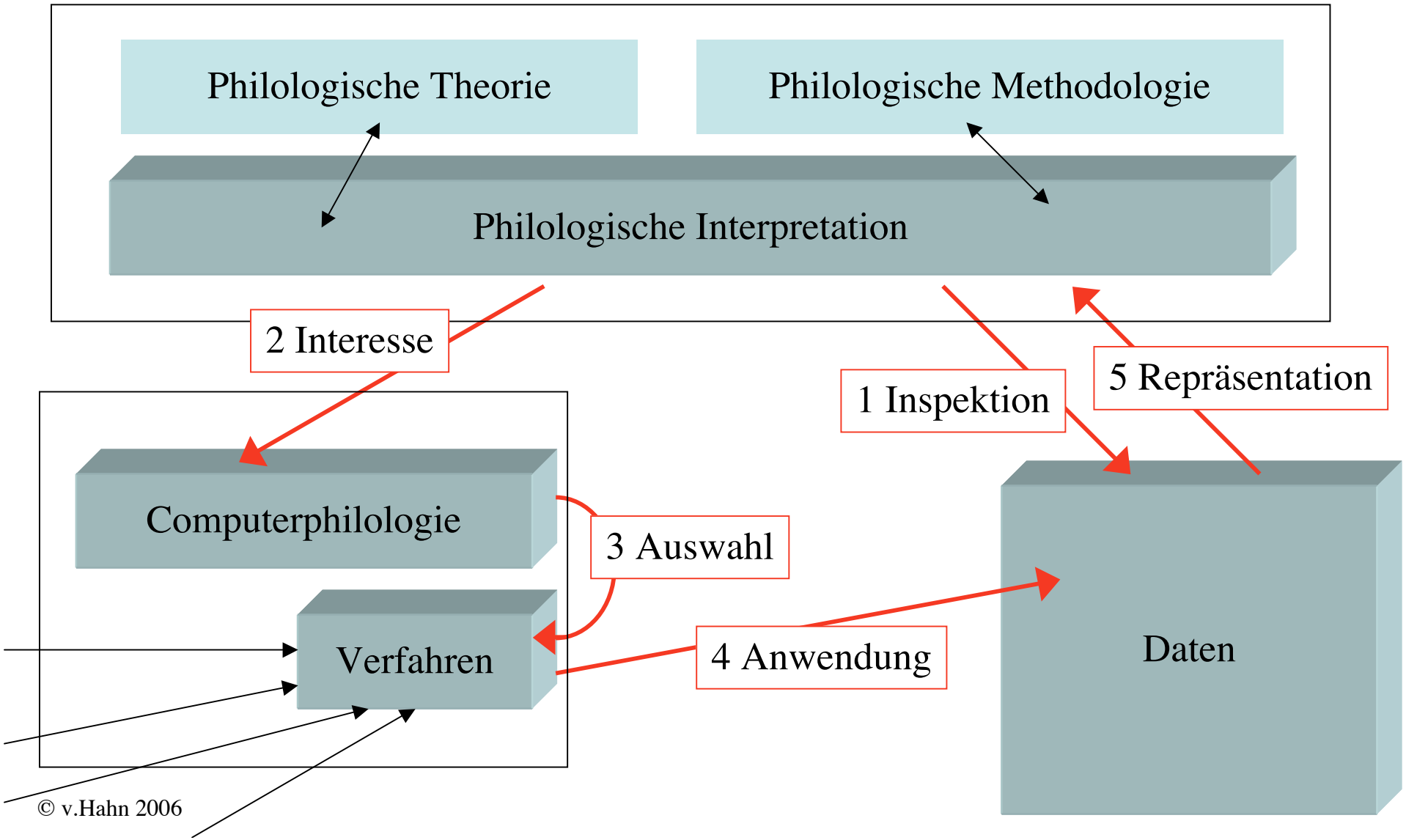
- Statistik
- Textuelle Präsentationstools
- Auszeichnungssprachen
- Analysetools
- Graphische Modelle
- Neuronale Modelle
- Regelbasierte Modelle

Bezug der Methoden der CP

Zum Gegenstandsbereich der CP: Was wird durch die Ergebnisse der CP erhellt?

- Sind z.B. Methoden der CP vorstellbar, die auf die Konstruktion der Computer oder der Philologie abzielen, in ähnlicher Weise, wie wie Methoden der Literaturwissenschaft auf wissenschaftliche Aussagen über Literatur abzielen?
- Sind es nicht vielmehr die Methoden der CP, die auf philologische (literaturwissenschaftliche, linguistische etc.,) Ergebnisse abzielen?

Abhängigkeit der CP



Klassische Themen I

1.Formale und theoretische Themen:

A. Textdarstellung und -auszeichnung

- *Informations- und Datenmodellierung,
- *Schreibsysteme und Codierungen,
- *Textmarkierungssysteme (TEX, SGML, XML) und Styles (TEX, CSS),
- Textrepräsentation auf allen linguistischen Ebenen
- Textuelle Repräsentation mit den Schwerpunkten Logik, Kohärenz und Dialog
- *Entwurf relationaler Datenbanken

* In der Liste der Europäischen „Working Group“

- B. Textanalyse
  - *Informationserhebung und -filterung mit den Schwerpunkten Corpora, Bibliographien und Internet,
  - Informationsdesign (Tabellen, Visualisierung, modale Transformationen)
  - *(Halb-)automatische Textmarkierung (Tagger)
  - *Automatische Textanalyse mit den Schwerpunkten: Lexikalische (z.B. LSA), syntaktische (ATN, Chart), semantische und textuelle Methoden,
  - *Statistische Methoden für textuelle Merkmalsverteilungen

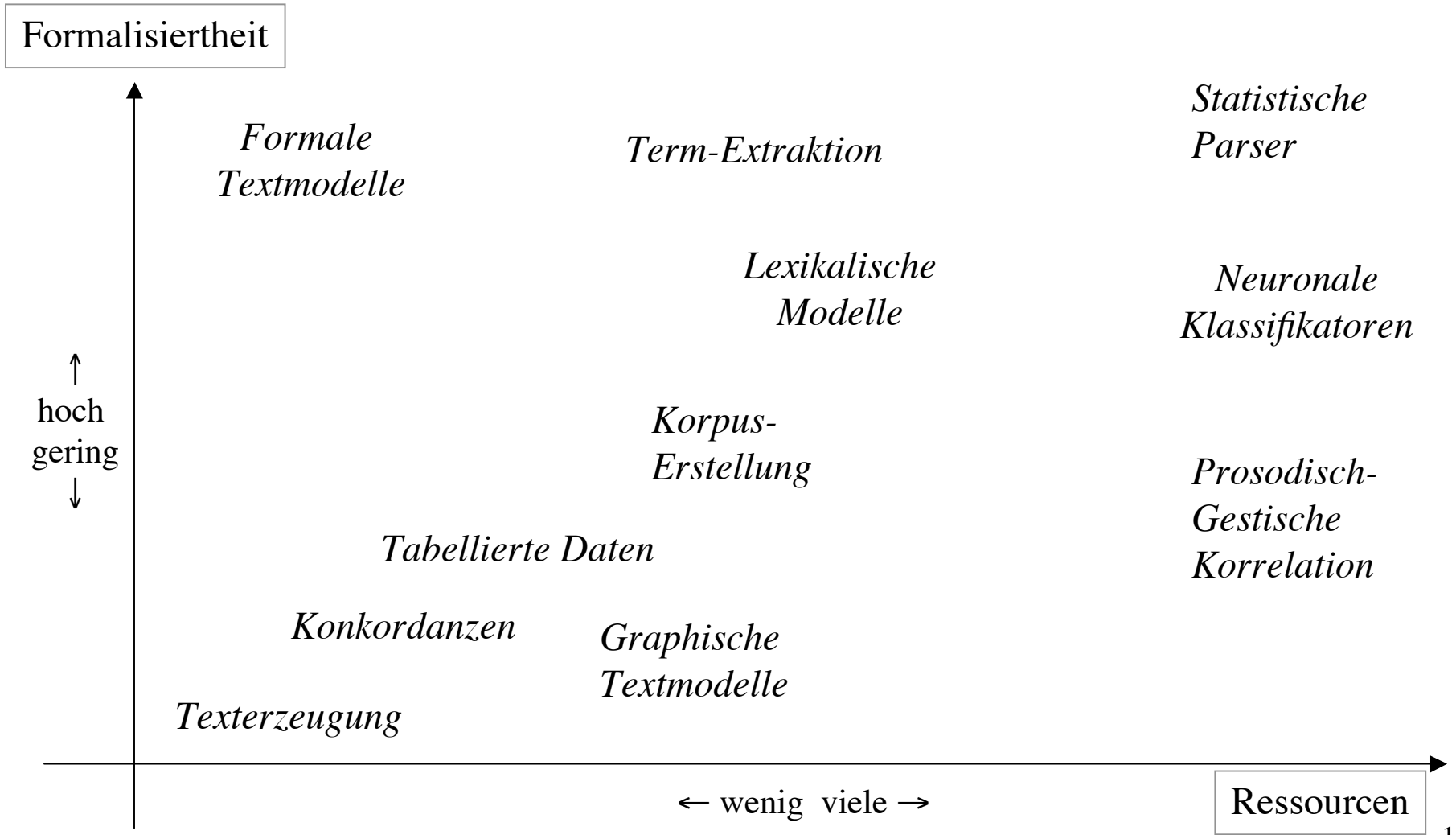
Klassische Themen II

2. Angewandte Themen:

- Hypertextsysteme (HTML, XHTML),
- Konkordanzen und Partituren,
- Multimodalität,
- Multilingualität,
- *Textdesign mit den Schwerpunkten Medientexte, Verlagswesen, E-Zines, Internet
- *Bildbearbeitung,
- *Erstellung, Verwaltung und Werterhaltung elektronischer Ressourcen,
- *Internationale und Europäische Standardisierungen,
- *Informationsgesellschaft und Neue Medienwirklichkeit.

* In der Liste der Europäischen „Working Group“

Komplexität (an frei gewählten Beispielen)



Annahmen der CP

- Sinn:
  - Veränderte Repräsentation von Daten verändert die Interpretation
- Qualität
  - Automatische Qualitäts- und Plausibilitätsüberprüfungen erhöhen die Qualität der Daten
  - Auf automatischen Datenanalysen oder -repräsentationen basierte Interpretationen sind zuverlässiger
- Ökonomie:
  - Es gibt Repräsentationen, die ohne Computer nicht in sinnvoller Zeit aufgebaut werden können
- Datenentstehung:
  - Durch statistische Verfahren kann neues philologisches Wissen entstehen (z.B. statistisches Übersetzen)
- Konsistenz:
  - Alle Computerverwendungen im Umfeld von Texten haben Gemeinsamkeiten
- Interdisziplinarität:
  - In allen Philologien spielen Texte eine ähnliche Rolle

Geschichtliches I

In den 70er Jahren

- setzte in der Linguistik eine sprachstatistische Bewegung ein,
- entstanden die ersten statistischen und formalen Literaturmodelle,
- erschien die Zeitschrift „Computers and the Humanities“ (CHUM).
- entstand die Arbeitsstelle für Nichtnumerische Datenverarbeitung, die erste Sprachverarbeitungsprogramme auf Zeichenebene in FORTRAN bereitstellte,
- setzte das „Bundesministerium für Forschung und Technologie“ ein Förderprogramm für „Information und Dokumentation“ auf, das viele Themen des textuellen Informationsmanagements förderte.
- begann die Simulation sprachlicher Fähigkeiten in Instituten in Darmstadt und Hamburg

Geschichtliches II

- In den Folgejahren wurde im angelsächsischen Raum die auf Literatur und andere geisteswissenschaftliche Themen gerichtete Computer-nutzung eingeführt und gelehrt.
- Dabei ging man von numerischen Methoden auf textuelle Methoden über.
- In den 90ern entstanden die ersten semantischen und konzeptuellen Verfahren vor dem Hintergrund der damals etablierten Forschung zur „Künstlichen Intelligenz“
- Nach der Jahrtausendwende hat das Forschungsthema „Semantisches Web“ auch die CP erfasst.

Caveat

- Hilfswissenschaften wie die CP haben keinen Selbstzweck; man kann sie zwar isoliert erforschen und entwickeln (z.B, spezielle Markup-Sprachen), aber sie bleiben immer **von den philologischen Frage-stellungen abhängig**. M.a.W., ihr Wert kann nur auf Grund ihrer philologischen Leistung beurteilt werden, nicht auf Grund inhärenter Methoden und Ziele
- Aufbereitungsverfahren sind oft sehr zeitaufwendig. Da die Arbeit mit dem Computer oft mehr Spaß macht als die mit Zetteln, wird der Zeitverlust einer Computerbearbeitung nicht realistisch eingeschätzt.
- Die interpretativen Ergebnisse stehen gelegentlich in keinem vernünftigen Verhältnis zum Zeitaufwand.
- Der Wert der Ergebnisse von CP-Verfahren wird oft überschätzt. Quantität überdeckt gelegentlich Qualität.
- Hinter allen statistischen oder formalen Verfahren muss vor allem ein sinnvolles Modell stehen.

Internet