

Proseminar
„Natürlichsprachliche Technologien“/
„Natural Language Technologies“

WiSe 06/06

Cristina Vertan

University of Hamburg • Informatics Department
Natural Language Systems Group (NATS)

WWW: <http://nats-www.informatik.uni-hamburg.de/~cri/>

E-Mail: [vertan@informatik, uni-hamburg.de](mailto:vertan@informatik.uni-hamburg.de)

Aims

- Learn to make a presentation
- Learn to write a short paper
- Learn to use a eLearning Platform
- Learn something about how can computer process natural language
- Learn about modern technologies

Typical Features of Natural Language

- Unclear focus of analysis, esp. with spoken input (the whole text only?)
 - Ambiguity on all levels
 - Self-reference, meta language capacity (I meant "lion")
 - Valencies, i.e. syntactic/semantic co-occurrences of categories
 - Multi-word lexemes and idioms with non-compositional meaning
 - Hierarchical syntax in non-linear order
 - Long distance dependencies
 - Discontinuous components
 - Ellipses
 - Paraphrases
 - Coherence
 - Understanding by word knowledge
- (Imagine these phenomena in programming languages ...)

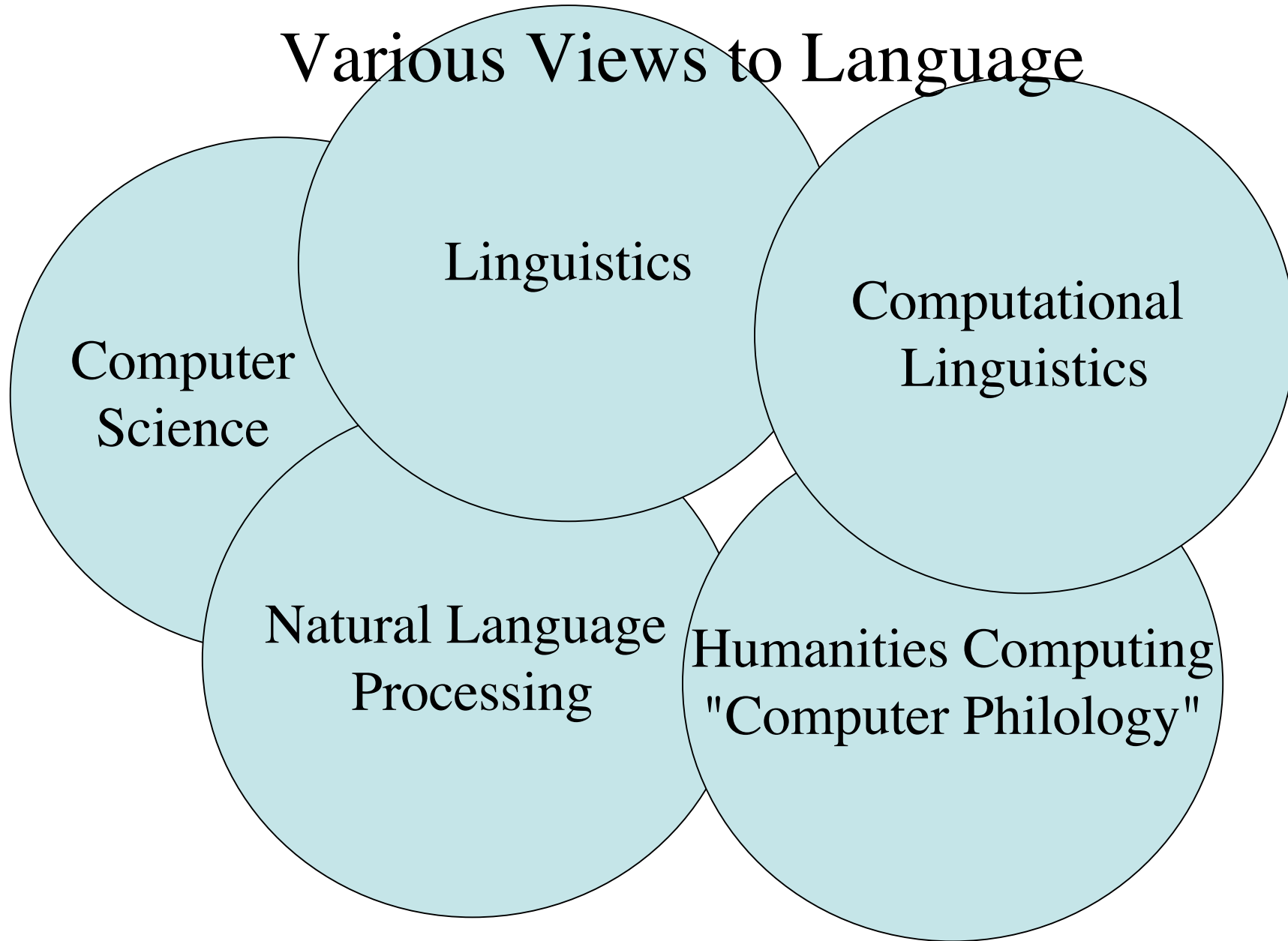
Two Parts of NLP

- Speech "From microphone to symbols"
- Language "From symbols to understanding"

+ their interplay, e.g.:

- prosody,
- expectation and focus,
- ambiguity resolution,
- mis-interpretations,
- irony, attitudes,
- hesitations and re-starts,

Various Views to Language



Intersections with other fields

- Cognitive Science
- Knowledge Management
- Man-Machine Interaction
- Multi-Modal Processing
- Psychology

Typical Tasks of Speech

- Speech recognition
- Speech generation
- Speaker recognition
- Speech distinction
- Language recognition

Typical Tasks of Language Processing

- Tokenizing
- Lemmatizing
- Parsing
- Semantic construction
- Discourse analysis
- Discourse construction
- Generation

Typical Fields of Language Processing

- Summarization
- Machine translation
- Question answering
- Semantic retrieval
- Presentation systems
- Tutoring systems
- E-Learning

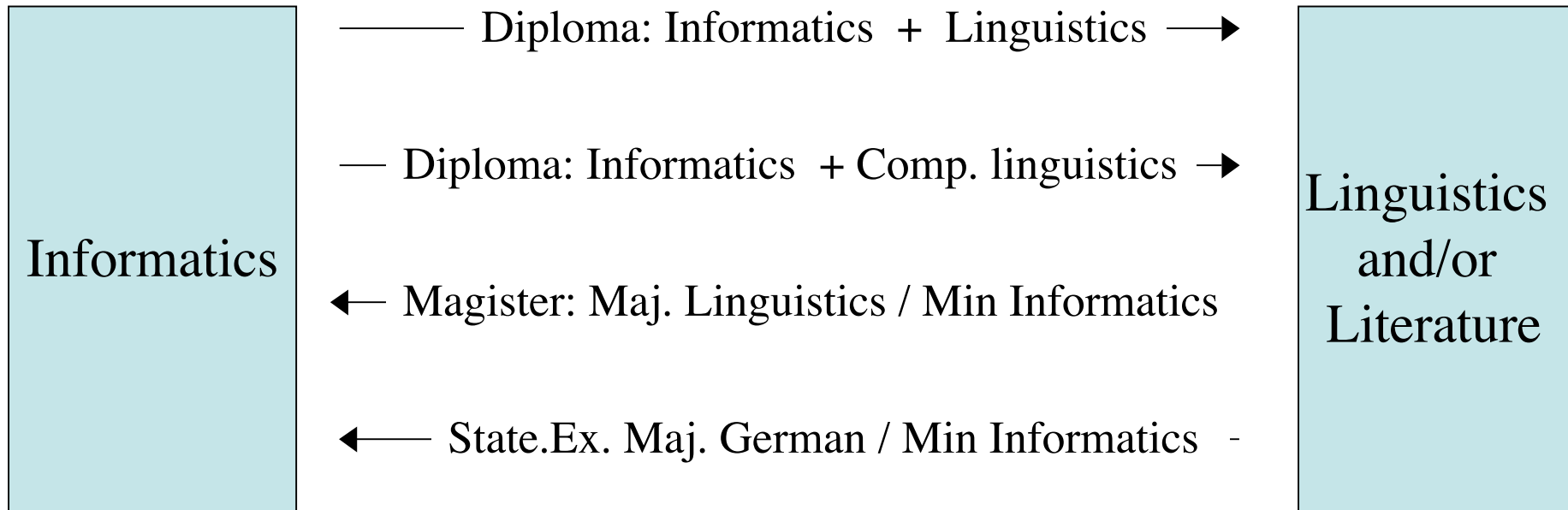
Limitations of Learning

- Rule-based methods have their limitations, e.g. in style, idiomatics and other exception handling
- stochastic methods have their limitations in ambiguity resolution, direct deixis, pronoun resolution, or cultural differences, therefore
- Future systems will have a hybrid architecture with finite state machines, rules, learning and probabilistic mechanisms.

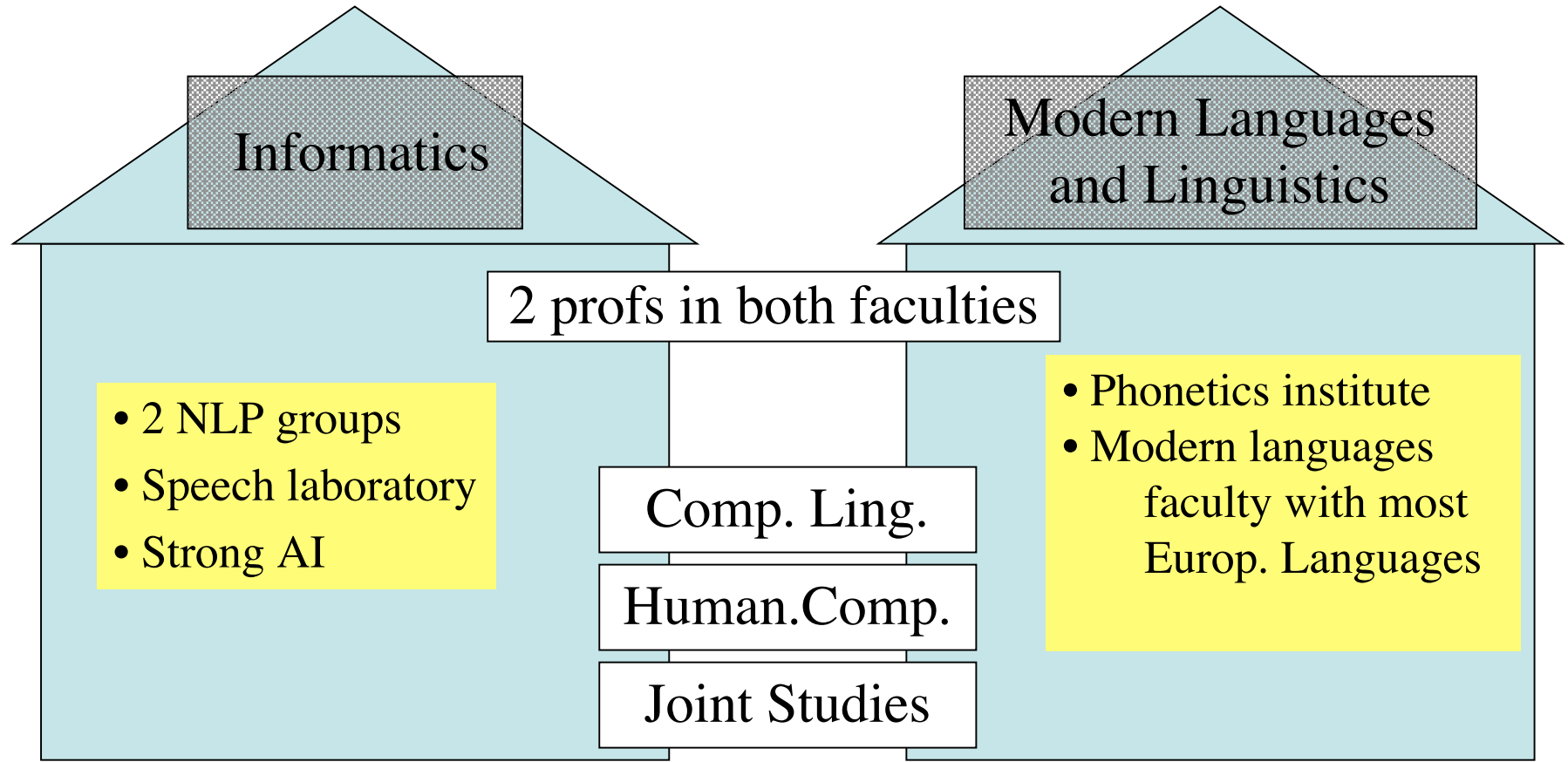
Why Linguistics? Why more than formal language analysis?

- For language understanding in general and other language specific phenomena like
- Ambiguities and vagueness,
- Interpretation of syntax and word order,
- Deictic expressions,
- Natural semantics like natural quantifiers
- Interpretation of prosody,
- Discourse coherency,
- Indirect speech acts,
- etc.

Studies between CS and Linguistics



NLP at Hamburg University



„Natural Language Systems Division“

Professors: Prof. Dr. Walther v.Hahn, Prof. Dr. Wolfgang Menzel

Senior Researcher: Dr. Cristina Vertan

Assistants: Kilian Foth, Michael Daum, Nguyen-Thanh Le, Monica Gavrilă, Christina Maack

Ph.D students: Solomon Teferra Abate, Natalia Elita

Other Members: Dr. Kerstin Fischer, Uwe Debacher

,

The Natural Language Systems Division (NATS) is dedicated to the design and realization of software systems for knowledge based natural language processing. Based on the assumption that natural language processing can contribute in a fundamental way

- * to ease human access to contemporary IT systems
- * to improve cross-lingual communication, and
- * to efficiently deal with the constantly growing problem of information overload

various application scenarios are considered.

Special focus is on

- * access to large information systems,
- * information extraction from documents,
- * information summarization for texts or complex information processing results,
- * machine translation and interpreting, and
- * computer-assisted language learning.

Themen für Proseminar

- 24. 10. Organisation meeting
- 31. 10. Introduction in NL technologies (CV)
- Proposal for further themes :
 - Text processing
 - Information extraction
 - Information retrieval
 - Question answering
 - Text summarization
 - Machine Translation
 - Natural Language Generation
 - Knowledge Aquisition
 - Speech recognition
 - Speech synthesis

Scheinkriterien

- Vortrag
- Referat
- Anwesenheit (max 2 motivierte Abwesenheiten)
- Benutzung der eLearning Plattform

Kontakt

- Webseite:

<http://nats-www.informatik.uni-hamburg.de/view/User/NatuerlichsprachlicheTechnologien>

- E-mail:

vertan@informatik.uni-hamburg.de

- Sprechstunde : Di 13-14 (F-211)und nach Vereinbarung

- Sekretariat:

Frau Jarck, F-205 (Vormittags), jarck@nats.informatik.uni-hamburg.de