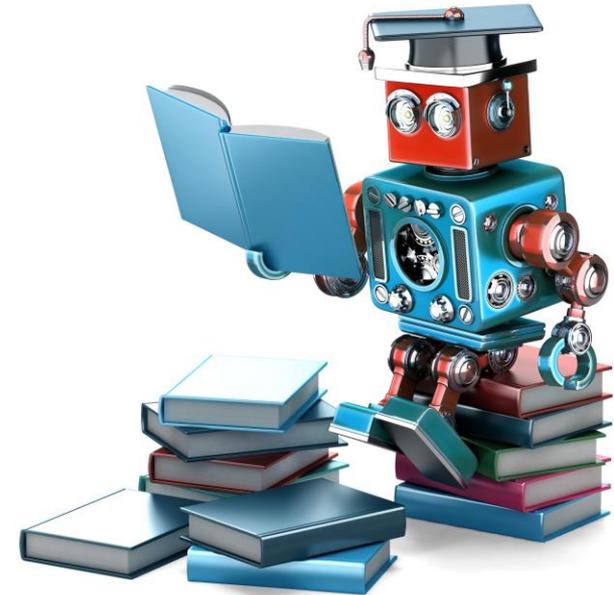


Interdisziplinärer Studiengang:

- Abteilung für Computerlinguistik
- Abteilung für allgemeine Sprachwissenschaft
- Institut für Informatik



Was ist Computerlinguistik?

Computerlinguistische Anwendungen begegnen uns überall:

- Menschen bedienen technische Geräte via **Sprachsteuerung**.
- Computer verwenden Sprachdaten, um daraus **strukturierte Informationen** zu gewinnen.
- Endgeräte verwenden **Sprachausgabe**, um Informationen an Menschen zu übermitteln.
- Wissenschaftliche Forschung ermittelt Korrelationen zwischen **Sprachnutzung und nicht-sprachlichen Phänomenen** (z.B. Anstieg von Suchanfragen nach Grippe-Symptomen während Erkältungswellen)

Computerlinguistik in Aktion



Write With Transformer `gpt2` ⓘ

↻ Shuffle initial text

⏴ Trigger autocomplete or `tab`

Select suggestion `↑` `↓` and `enter`

Cancel suggestion `esc`

Save & Publish 

My favorite thing about studying computational linguistics is

that you get to see how the languages you

that I am not just going to learn a

that I am not only a linguist,

Quelle: <https://transformer.huggingface.co/>, am 13.01.2020

Computerlinguistik in Aktion



Write With Transformer `gpt2` ⓘ

 Shuffle initial text

 Trigger autocomplete or `tab`

Select suggestion `↑` `↓` and `enter`

Cancel suggestion `esc`

Save & Publish 

My favorite thing about studying computational linguistics is that you get to see how the languages you use are different from the languages you study. For example, the language of "Coffee " can be quite different from the language of "English." In English, you could always say that you wanted to have a cup of coffee or you wanted to have a coffee, and that there was an easy way to make it so. However, a number of languages are more complex than English. "Coff ee" is a real language , but it doesn't say much about coffee . What you need to do in order to have a cup of coffee is: 1. Use a word like " cup " 2 . Add an inflection al construction : 3 . Add the noun (coffee) to the verb (want). This tells you that coffee is something you want .

Share screenshot 

Quelle: <https://transformer.huggingface.co/>, am 13.01.2020

Funktioniert das auch auf Deutsch?

Talk to Transformer

See how a modern neural network completes your text. Type a custom snippet or try one of the examples. This is a limited demo of [InferKit](#).

Custom prompt

Diese Woche sprechen wir über

Generate Another

Completion

Diese Woche sprechen wir über die Verbindungen der Verlage mit dem Computer und haben Verleger ihre Meinung im Gespräch. In der Mittagspause haben wir unter dem Motto "Was Verlage noch machen wollen" einige Rekorde erreicht. Unsere Gäste haben viele Worte gefunden - dass sie keinen Bock mehr auf die Zeitungsbranche haben.

Quelle: <https://app.inferkit.com/demo>,
am 18.01.2021

Was ist Computerlinguistik?

Computerlinguistik = Linguistik + Informatik

- Schnittstellenfach: Methoden der Informatik werden auf Fragestellungen der Sprachwissenschaft angewendet.
- Formalisierung linguistischer Modelle
- Implementierung von Softwarewerkzeugen zur maschinellen Sprachverarbeitung:
 - Bedeutungserkennung, Wörterbucherstellung, maschinelle Übersetzung, Dialogsysteme, Marktanalysen, Plagiatserkennung, ...

Studienziele

Eine breite computerlinguistische Ausbildung:

- Fähigkeit zur fundierten Analyse von Sprachdaten
- Verständnis mathematisch-informatischer Grundlagen
- Programmierkenntnisse und Kenntnisse in strategischer Softwareentwicklung
- Erfahrung in der Anwendung computerlinguistischer Modelle auf Sprachdaten (Hypothesenbildung, Implementierung, Evaluation)

Berufsfelder

Vor allem in der IT/Softwareentwicklung:

- Abfrage, Aufbereitung, Verwaltung von Sprachdaten
 - Erstellung von Textdatenbanken, Wörterbucharstellung, ...
- Analyse von Sprachdaten, Spracherkennung
 - Suchmaschinen, Information Retrieval, automatische Textzusammenfassung, Sentimentanalyse, ...
- Generierung von Sprachdaten
 - Textgenerierung, Text-to-Speech, maschinelles Übersetzen, ...

Voraussetzungen

- Interesse an Sprache(n)
- Aufgeschlossenheit gegenüber Mathematik und Informatik

Programmierkenntnisse sind keine Voraussetzung! Sie werden in den Grundlagenmodulen praxisbezogen vermittelt.

Studieren in Coronazeiten

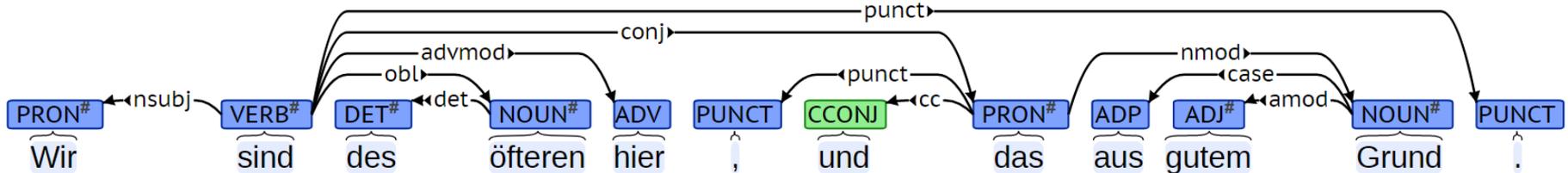
- Aktuell finden keine Lehrveranstaltungen auf dem Campus statt.
- Die meisten Vorlesungen haben eine der folgenden Formen:
 - Vorlesungsstream auf Twitch
 - Regelmäßige Videosessions (wie dieser Termin hier!)
 - Videos in der Mediathek
 - Schriftliche Übungen/Tests/Hausaufgaben
 - Virtuelle Sprechstunden
 - ...
- Es gibt einen Discord-Server für Studierende der (Computer)linguistik.

Studienverlaufsplan

Semester							
1	Propädeutikum Programmierung, Mathematik, Linguistik, Logik			Morphologie & Syntax		Wahlbereich	
2	Basismodul Einführung in die Computerlinguistik, Grammatikformalismen	Quantitative Methoden (1) Linguistische Ressourcen, Programmierung	Semantik & Pragmatik				
3		Mathematische Linguistik Automatentheorie, formale Sprachen Vertiefung	Computationelle Semantik (1) Programmierung, computerlinguistische Semantik	Informatik			
4	Quantitative Methoden (2) Vertiefung		Morphologische & syntaktische Sprachverarbeitung Parsing, Vertiefung				
5			Vertiefungsmodul Linguistik, CL oder Informatik	Computationelle Semantik (2) Vertiefung			
6	Berufsfeldpraktikum oder Teamprojekt		Abschlussarbeit				
Gesamt 180 CP							

Grammatikformalismen

- Sprache ist **hierarchisch** aufgebaut. Manche Elemente hängen enger zusammen als andere:



Quelle: Universal Dependencies treebank
Joakim Nivre and others
<http://universaldependencies.org/>

- Mit Grammatikformalismen können wir diese hierarchische Struktur beschreiben und analysieren.
- Ein Formalismus kann für manche Phänomene gut und für andere schlecht geeignet sein.

Quantitative Methoden

- Um sprachliche Phänomene zu erforschen, können wir große Datenmengen **statistisch auswerten**. So ermitteln wir z.B., wie oft bestimmte Konstruktionen in einem bestimmten Zeitraum beobachtet wurden.
- Basierend auf großen Datenmengen können **neuronale Netze** trainiert werden, die bestimmte sprachliche Aufgaben bearbeiten können.

Automatentheorie und formale Sprachen

- Wie **komplex** können natürliche Sprachen sein?
 1. Micky ist eine Katze.
 2. Micky ist eine flauschige Katze.
 3. Micky ist eine flauschige, dreijährige Katze.
 4. Micky ist eine flauschige, dreijährige, liebenswerte Katze.
 5. ...

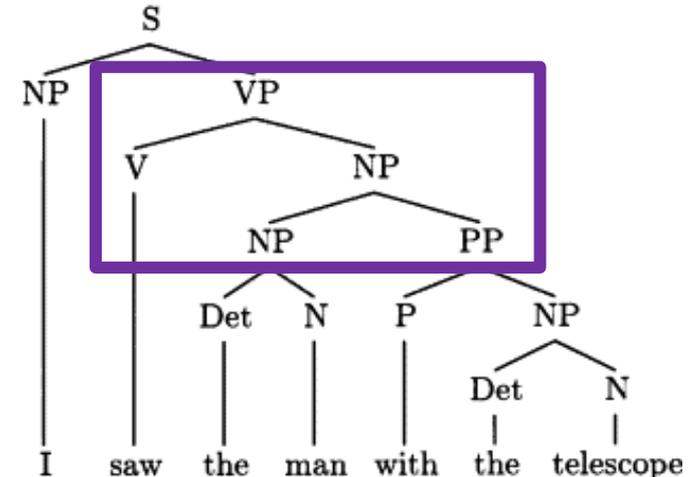
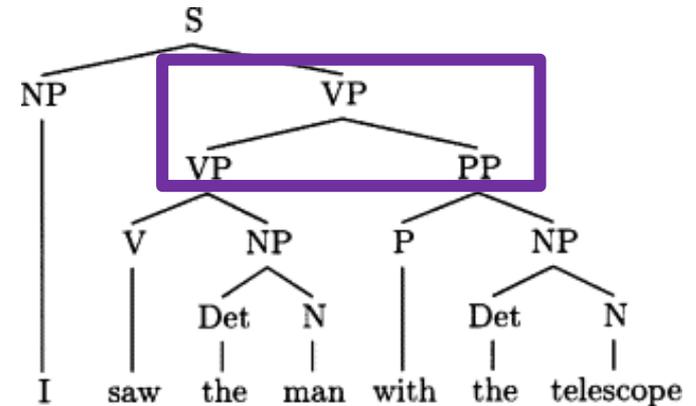
- Mit formalen und logischen Methoden können wir versuchen, die Regeln zu definieren, die alle möglichen sprachlichen Ausdrücke **erlauben** und gleichzeitig alle unmöglichen sprachlichen Ausdrücke **verbieten**.
 1. Micky ist eine flauschige, dreijährige Hund.
 2. Micky ist eine flauschige Katze dreijährige.

Computationelle Semantik

- Wörter sind nicht nur Sequenzen von Zeichen oder Sprachlauten, sondern haben auch Bedeutungen. Komplexere sprachliche Ausdrücke haben auch komplexere Bedeutungen:
 - Glaskiste (Kiste aus Glas)
 - Glaskiste (Kiste für Trinkgläser)
 - Glaskiste (Kiste, in der Glasscherben aufbewahrt werden)
 - Glaskiste (Kiste in Form eines Trinkglases)
 - ...
- Um mit Computern Bedeutungen abzubilden, kann man z.B. lexikalische oder statistische Modelle (Wörterbücher oder Häufigkeitsstatistiken) verwenden.

Morphologische und syntaktische Sprachverarbeitung

- Wenn wir die hierarchische Struktur eines Satzes ermitteln, können wir einzelne Phänomene analysieren und erforschen:
 - Welches Element in einem Satz ist das Adjektiv?
 - Wann dürfen Verben am Satzende stehen und wann nicht?
- Manche Sätze können **mehr als eine syntaktische Analyse** haben, wie im Beispiel rechts. Wer hat das Teleskop, ich oder der Mann?



Studienverlaufsplan

Semester							
1	Propädeutikum Programmierung, Mathematik, Linguistik, Logik			Morphologie & Syntax		Wahlbereich	
2	Basismodul Einführung in die Computerlinguistik, Grammatikformalismen	Quantitative Methoden (1) Linguistische Ressourcen, Programmierung	Semantik & Pragmatik				
3		Mathematische Linguistik Automatentheorie, formale Sprachen Vertiefung	Computationelle Semantik (1) Programmierung, computerlinguistische Semantik	Informatik			
4	Quantitative Methoden (2) Vertiefung		Morphologische & syntaktische Sprachverarbeitung Parsing, Vertiefung				
5			Vertiefungsmodul Linguistik, CL oder Informatik	Computationelle Semantik (2) Vertiefung			
6	Berufsfeldpraktikum oder Teamprojekt		Abschlussarbeit				
Gesamt 180 CP							

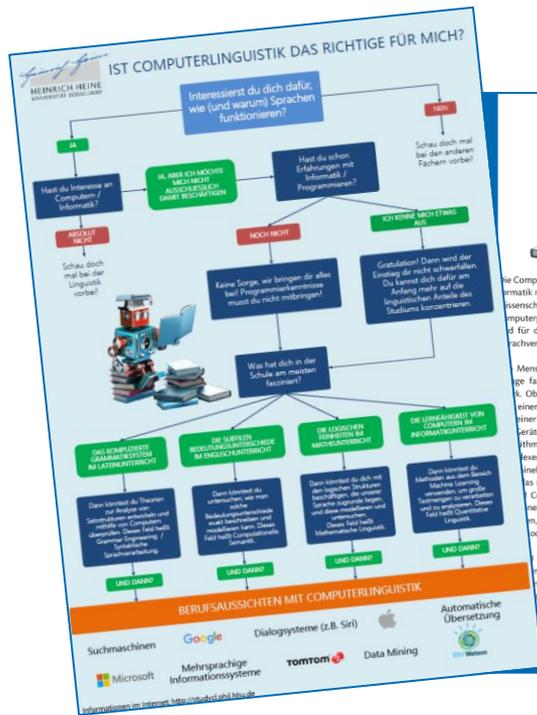
Lehrformen

- Vorlesung:
 - Wissensvermittlung Professor*in → Studierende
 - Vorwiegend Frontalunterricht
 - Prüfungsform: meist Klausur
- Seminar:
 - Interaktive Wissensvermittlung
 - Teils Frontalunterricht, teils Diskussion, oft mit freiwilligen Hausaufgaben
 - Prüfungsform: Klausur, Hausarbeit oder mündliche Prüfung
- Übung:
 - Interaktive Wiederholung oder praktische Anwendung von Stoff aus anderen Lehrveranstaltungen
 - Oft von Studierenden unterrichtet
 - Keine Prüfung

Ist Computerlinguistik das Richtige für mich?

- Für Nachfragen und individuelle Beratung steht der Fachschaftsrat Linguistik zur Verfügung!
- Bei weiterführenden Fragen können Sie sich per Mail an unser Team wenden:
 - Studiengangwechsel: Laura Kallmeyer (kallmeyer@phil.hhu.de), Wiebke Petersen (petersen@phil.hhu.de)
 - Organisatorisches oder Tipps zum Studienbeginn: Fachschaftsrat Linguistik (fsasw@phil.hhu.de)
 - Inhaltliche Fragen zum Studiengang: Wiebke Petersen, Laura Kallmeyer

Hilfe bei der Studienwahl: <http://studycl.phil.hhu.de/>



Die Computerlinguistik kombiniert Methoden aus der Informatik mit Fragestellungen aus der Linguistik (Sprachwissenschaft). Sie ist also eine wichtige Grundlage für computergestützte sprachwissenschaftliche Forschung für die Entwicklung von Softwarewerkzeugen zur Textverarbeitung.

Menschen mit Maschinen interagieren, sind heute fast immer computerlinguistische Prozesse am Werk. Ob wir unser Smartphone nach dem Wetter fragen, einen Navigator benutzen, um unser Ziel zu finden, oder eine Suchmaschine benutzen, um unsere Recherchen zu erleichtern, „verstehen“ wir, weil computerlinguistische Algorithmen unsere menschliche Sprache mit all ihren feinen Eigenschaften in eine Form übersetzen, die schnell verarbeitet werden kann. Allerdings funktioniert das nicht immer perfekt – es gibt noch viel zu tun.

Computerlinguistik ist vielseitig und spannend: hier im Studium und danach eigene Schwerpunkte setzen. In der Computerlinguistik einsteigen. Wir unterstützen dich bei deiner Studienwahl und deinem Studium mit einem sehr breiten computerlinguistischen Spektrum. Darüber hinaus kooperieren wir mit anderen Experten mit starkem Forschungsprofil in der Linguistik und Informatik.

Studienverlauf

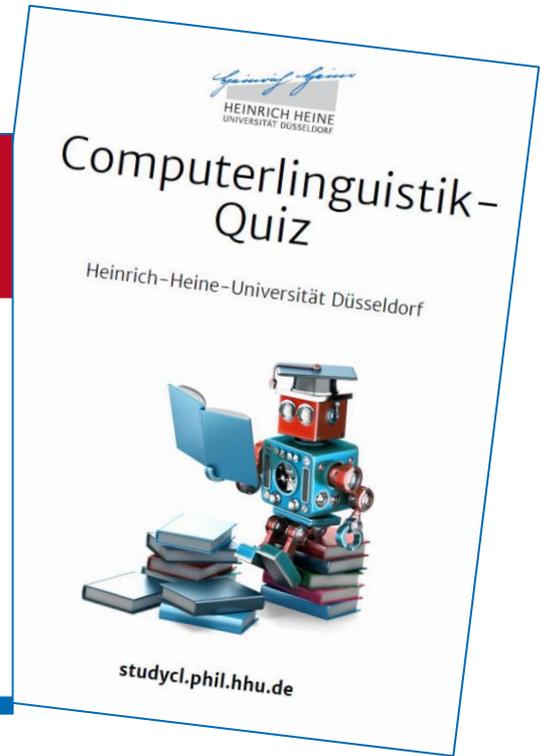
Semester	Propädeutikum		Morphologie & Syntax		Wahlbereich
1	Propädeutikum	Programmierung, Mathematik, Linguistik, Logik			
2	Basismodul	Einführung in die Computerlinguistik, Grammatikformalismen	Quantitative Methoden (1)	Linguistische Ressourcen, Programmierung	
3			Semantik & Pragmatik		
4	Quantitative Methoden (2)	Vertiefung	Mathematische Linguistik	Automatentheorie, formale Sprachen	Computationale Semantik (1)
5			Vertiefungsmodul	Linguistik, CL oder Informatik	Computationale Semantik (2)
6	Berufsfeldpraktikum oder Teamprojekt		Vertiefungsmodul	Linguistik, CL oder Informatik	Abschlussarbeit

Gesamt 180 CP

Im Studium erwerben Sie Kenntnisse in den Bereichen Linguistik, Logik und Informatik. Sie erhalten einen Überblick über den Forschungsstand und lernen, Phänomene und Problemstellungen der Computerlinguistik zu erfassen und zu beschreiben. Sie setzen Programmierprojekte um und entwickeln Lösungsansätze für spezifische computerlinguistische Aufgaben. Sie schließen das Studium mit einer Bachelorarbeit ab, in der Sie eine selbstgewählte Fragestellung bearbeiten und zeigen, dass Sie die Theorie und Praxis computerlinguistischer Forschung beherrschen.

Der computerlinguistische Arbeitsmarkt ist in den letzten Jahren stark gewachsen. Die fundierte und vielseitige computerlinguistische Ausbildung bereitet Sie auf Berufe in Wissenschaft oder Industrie bestens vor.

Einige Beispiele für typische computerlinguistische Aufgaben: Maschinelle Übersetzung, Textgenerierung, Sprachdatenanalyse, Lexikonerstellung, Entwicklung/Optimierung von Suchmaschinen, automatische Autorenerkennung oder Stilanalyse, Entwicklung von Dialogsystemen wie Siri oder Alexa ...



HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Computerlinguistik-Quiz

Heinrich-Heine-Universität Düsseldorf

studycl.phil.hhu.de

... oder einfach einen Termin zur Beratung vereinbaren!