

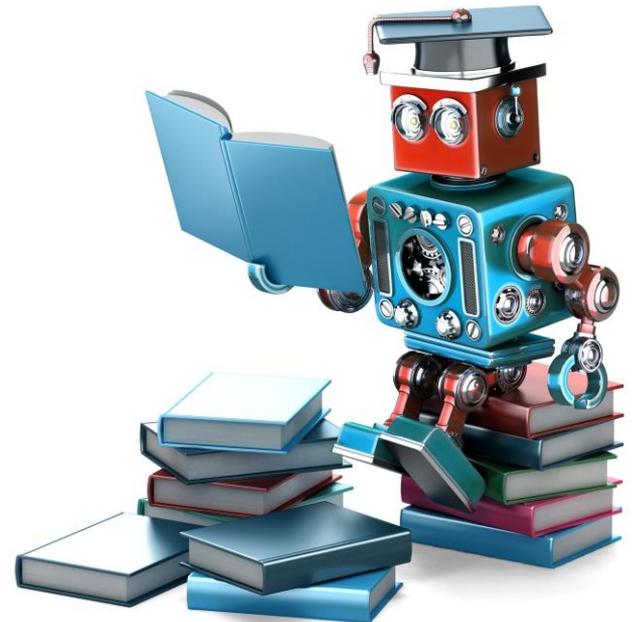


Informationsvortrag zum Studiengang Computerlinguistik

Heinrich-Heine-Universität Düsseldorf

Vortragende: Esther Seyffarth

1. Was bedeutet Computerlinguistik?
2. Welche Aufgaben hat ein_e Computerlinguist_in?
3. Welche Berufe kann man als Computerlinguist_in ausüben?
4. Wie läuft das Computerlinguistik-Studium an der Uni Düsseldorf?
5. Wo finde ich weitere Informationen?



Was bedeutet Computerlinguistik?

- Computerlinguistik ist ein Fach, das in den Überschneidungsbereichen der **Sprachwissenschaft** und der **Informatik** angesiedelt ist.
- Es geht darum,
 - ... Sprache mit informatischen Methoden zu erforschen.
 - ... Software zu entwickeln, die für die Verarbeitung von Sprache(n) geeignet ist.
 - ... zu erforschen, ob Computer oder andere Maschinen Sprache „verstehen“ können.
 - ... aus Daten in Textform statistisch verwertbare Informationen zu gewinnen.
- Oft wird über künstliche Intelligenz im Zusammenhang mit Computerlinguistik gesprochen. KI basiert hauptsächlich auf statistischen und mathematisch beschreibbaren Regelmäßigkeiten. Kann menschliche Sprache mit solchen Methoden überhaupt **ausreichend modelliert** werden?

Wo begegnet uns Computerlinguistik im Alltag?

Computerlinguistische Anwendungen sind heutzutage verbreitet:

- Menschen bedienen technische Geräte mit **Sprachsteuerung**.
- Computer verwenden Sprachdaten (zum Beispiel von sozialen Medien), um daraus **strukturierte Informationen** zu gewinnen (zum Beispiel, wo gerade ein Erdbeben stattfindet).
- Endgeräte verwenden **Sprachausgabe**, um Informationen für Menschen aufzubereiten (zum Beispiel Amazon Alexa oder ein Navi im Auto).
- Wir **recherchieren** im Internet nach Informationen, und die Suchmaschine ermittelt mit computerlinguistischen Methoden, welche Webseiten am besten zur Suchanfrage passen.

Ein computerlinguistisches Programm kann uns helfen, Texte zu schreiben:



Write With Transformer `gpt2` ⓘ

↻ Shuffle initial text

⬇️ Trigger autocomplete or `tab`

Select suggestion `↑` `↓` and `enter`

Cancel suggestion `esc`

Save & Publish 

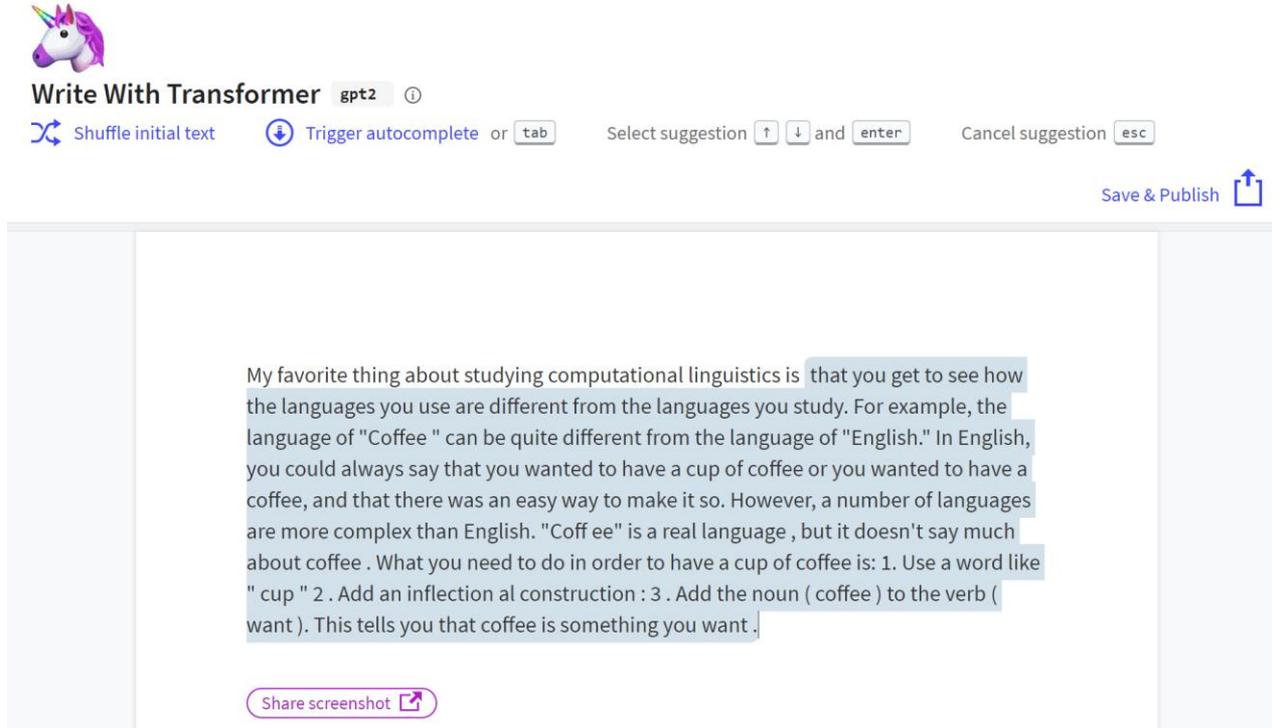
My favorite thing about studying computational linguistics is

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,

... und das kommt dabei heraus:



The screenshot shows the 'Write With Transformer' web interface. At the top left is a unicorn icon. The title 'Write With Transformer' is followed by 'gpt2' and an information icon. Below the title are several controls: a 'Shuffle initial text' button with a circular arrow icon, a 'Trigger autocomplete' button with a downward arrow icon and 'or tab', a 'Select suggestion' button with up and down arrow icons and 'and enter', and a 'Cancel suggestion' button with 'esc'. On the right side, there is a 'Save & Publish' button with an upward arrow icon. The main content area displays a paragraph of generated text: '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 .' At the bottom of the text area is a 'Share screenshot' button with a share icon.

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.

Textvervollständigung ist kein gelöstes Problem!

- Der englische Text sah ganz gut aus, aber im deutschen Text gab es noch einige **Fehler**.
- Bedeutet das, dass die deutsche Sprache für den Computer **schwieriger zu lernen** ist als die englische?
- Ihr könnt die Programme auch jederzeit selbst **ausprobieren**:
 - Englisch: <https://transformer.huggingface.co/doc/gpt2-large>
 - Deutsch: <https://app.inferkit.com/demo>

Was lernt man im Computerlinguistik-Studium?

- **Informatik-Grundlagen:** Programmiersprachen und -methoden, mit denen man Texte verarbeiten und Daten auswerten kann.
- **Sprachwissenschaftliche Grundlagen:** Begriffe und Konzepte der Linguistik.
- **Formalisierung linguistischer Modelle:** Sprachen logisch beschreiben und modellieren, um zuverlässige Vorhersagen über bestimmte Phänomene treffen zu können.
- Verständnis der Bereiche der **Mathematik**, die in der Computerlinguistik eine Rolle spielen: Statistik, Logik, Mengenlehre, Machine Learning.
- **Computerlinguistische Modelle auf Sprachdaten anwenden:** Hypothesen bilden, Experimente aufsetzen, Ergebnisse auswerten.

In der IT/Software-Entwicklung:

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

In der Wissenschaft/Forschung:

- Grundlagenforschung zu einzelnen Sprachen oder Sprachfamilien
 - Typologie, Korpusstudien, Vorhersage von sprachlichen Phänomenen, ...
- Entwicklung und Verbesserung von computerlinguistischen Methoden für bestimmte Herausforderungen
 - Machine Learning, automatische Übersetzung, automatische Vereinfachung von Texten, Erkennung von Plagiaten, ...
- Interdisziplinäre Arbeit mit Expert_innen aus anderen Gebieten
 - Computerlinguistische Analyse historischer Textquellen, „digital Humanities“, ...

Wie läuft das Studium der Computerlinguistik in Düsseldorf?

Die wichtigsten Voraussetzungen sind:

- Interesse an Sprache(n)
- Aufgeschlossenheit gegenüber Mathematik und Informatik
- **Programmierkenntnisse sind keine Voraussetzung! In den Grundlagenmodulen vermitteln wir praxisorientiert alle Kenntnisse, die ihr im Studium und danach braucht.**

Wie läuft das Studium der Computerlinguistik in Düsseldorf?

Studieren in Corona-Zeiten

- Während der Coronavirus-Pandemie wurden alle Lehrveranstaltungen auf ein Online-Format umgestellt.
- Vorlesungen konnten über Twitch, Videokonferenzen, synchrone und asynchrone Online-Sitzungen, virtuelle Sprechstunden und zusätzliche digitale Angebote verlustfrei fortgesetzt werden.
- Auch ohne regelmäßige Anwesenheit auf dem Campus könnt ihr ein Teil der Düsseldorfer Computerlinguistik-Community werden!

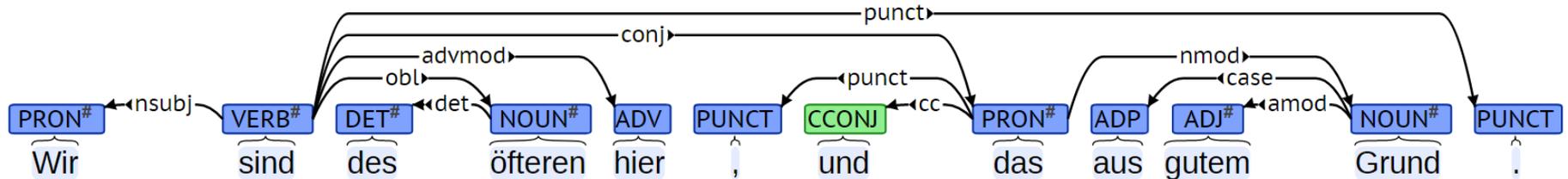
Wie läuft das Studium der Computerlinguistik in Düsseldorf?

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. Wir können das schematisch darstellen:



- Mit Grammatikformalismen können wir diese Struktur beschreiben und analysieren.
- Unterschiedliche Formalismen eignen sich besonders für verschiedene Phänomene und Sprachen.

Quantitative Methoden

- Um sprachliche Phänomene zu erforschen, können wir große Datenmengen **statistisch auswerten**. So ermitteln wir zum Beispiel, 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

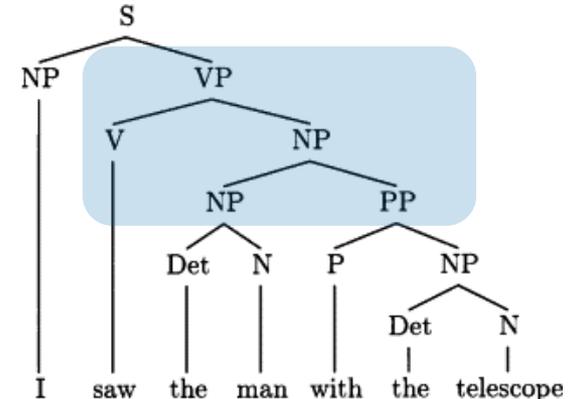
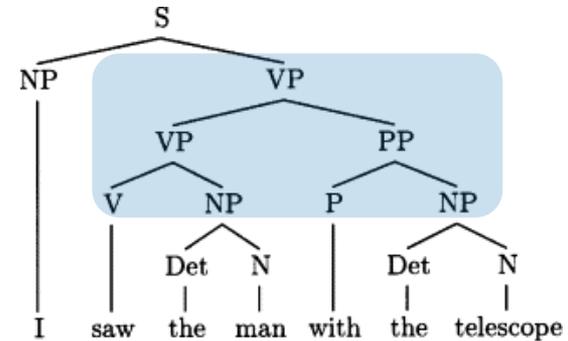
- Wie komplex können natürliche Sprachen sein?
 - Micky ist eine Katze.
 - Micky ist eine flauschige Katze.
 - Micky ist eine flauschige, dreijährige Katze.
 - ...
- Mit formalen und logischen Methoden versuchen wir, Regeln zu definieren, die alle **möglichen** sprachlichen Ausdrücke erlauben und gleichzeitig alle **unmöglichen** sprachlichen Ausdrücke verbieten.
 - Micky ist eine flauschige, dreijährige Hund.
 - 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)
 - ...
- Um mit Computern Bedeutungen abzubilden, kann man zum Beispiel 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 ein 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?



Wie läuft das Studium der Computerlinguistik in Düsseldorf?

Studienverlaufsplan

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

Unterschiedliche Lehrformen:

- **Vorlesung:** Wissensvermittlung, vorwiegend Frontalunterricht. Wird meist mit einer Klausur oder einer anderen Prüfung abgeschlossen.
- **Seminar:** Interaktive Wissensvermittlung. Oft mit Diskussionen über die aktuellen Themen, die mit Hausaufgaben vor- und nachbereitet werden. Wird oft mit einer Klausur, Hausarbeit oder mündlichen Prüfung abgeschlossen.
- **Übung/Tutorium:** Interaktive Wiederholung oder praktische Anwendung von Stoff aus anderen Lehrveranstaltungen. Wird nicht benotet und dient der Übung und Prüfungsvorbereitung.

Ist Computerlinguistik das Richtige für mich?

- Für Nachfragen und individuelle Beratung steht der Fachschaftsrat Linguistik zur Verfügung!
- Bei weiterführenden Fragen könnt ihr euch 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

