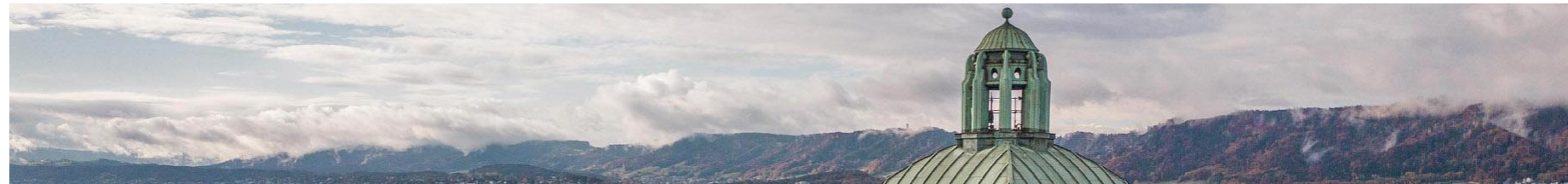




Universität
Zürich

Institut für Computerlinguistik



Studieninformationstage 2024 Computerlinguistik und Sprachtechnologie

Simon Clematide
September 2024

Mit Materialien von A. Säuberli und weiteren Studierenden



Themenübersicht

1. Was ist Computerlinguistik und Sprachtechnologie?
2. Wer forscht und lehrt am Institut für Computerlinguistik?
3. Was lernt man im Bachelorstudium "Computerlinguistik und Sprachtechnologie"?
4. Was kann man nach dem Bachelorstudium machen?
5. Fragen?

1. Was ist Computerlinguistik (CL)? Was ist Sprachtechnologie?

Computerlinguistik (*Computational Linguistics*)

Theorie:

Natürliche, d.h. menschliche Sprache mit dem Computer analysieren und modellieren

Sprachtechnologie (*Language Technology, Natural Language Processing, NLP*)

Praxis:

Systeme zur Verarbeitung und Generierung von Sprache erstellen

Gegenfragen:

Wo verwenden Sie Sprachtechnologie im Alltag?

Wie kann man Sprache analysieren?

Themenübersicht

1. Was ist Computerlinguistik und Sprachtechnologie?
- 2. Wer forscht und lehrt am Institut für Computerlinguistik?**
3. Was lernt man im Bachelorstudium "Computerlinguistik und Sprachtechnologie"?
4. Was kann man nach dem Bachelorstudium machen?
5. Fragen?

Institut für Computerlinguistik UZH am Campus Zürich Nord




ca. 70 Mitarbeitende,
darunter:


- 8 Professor:innen
- 30 Doktorierende
- studentische Mitarbeitende



Vielfältige interdisziplinäre Forschung und Lehre

 *und mehr ...*

 Maschinelles Lernen (AI)

 Sprachwissenschaft
(*Linguistics*)

 Mathematik

 Informatik

 Datenwissenschaft
(*Data Sciences*)

 Statistik

 Übersetzungswissenschaft

 Barrierefreiheit
(*Accessibility Studies*)

 Phonetik
(*Speech Sciences*)

 Neurowissenschaften

 Psychologie

 Geistes- und
Sozialwissenschaften

 Medizin

Die thematischen Forschungsgruppen unseres Instituts

- **Texttechnologie**
- Sprache, Technologie und Barrierefreiheit
- Digitale Linguistik
- Phonetik & Sprachsignalverarbeitung
- Computerbasierte Neurowissenschaften der Sprache & des Hörens

Maschinelle Übersetzung für (ressourcenarme) Sprachen

Our CL Start-up
Company

textshuttle.

Swiss German Beta



English

Language form

De Papscht het z Spiez s Späckbschteck
zschpoot bschtellt.

The Pope ordered the bacon cutlery too
late in Spiez.

Swiss German Beta



Romansh

Language form

De Papscht het z Spiez s Späckbschteck
zschpoot bschtellt.

Il Papa ha pustà a Spiez memia tard la
pusada da charnpiertg.

Digital Humanities: Eigennamen erkennen



BullingerDigital

Suche Listen Karte Hilfe

« ‹ Zurück Zeitleiste einblenden Namen markieren › »

[Johannes Haller d.J.] an Heinrich Bullinger



📅 18. Oktober 1546 📍 [Augsburg] Brief

Metadaten

Datum	Montag, 18. Oktober 1546
Absender	[Johannes Haller d.J.], [Augsburg]
Empfänger	Heinrich Bullinger
Autograph	Zürich StA, E II 370,24 f.
Sprachen	Deutsch, Latein
Gedruckt	HBBW 18, Nr. 2631, S. 150

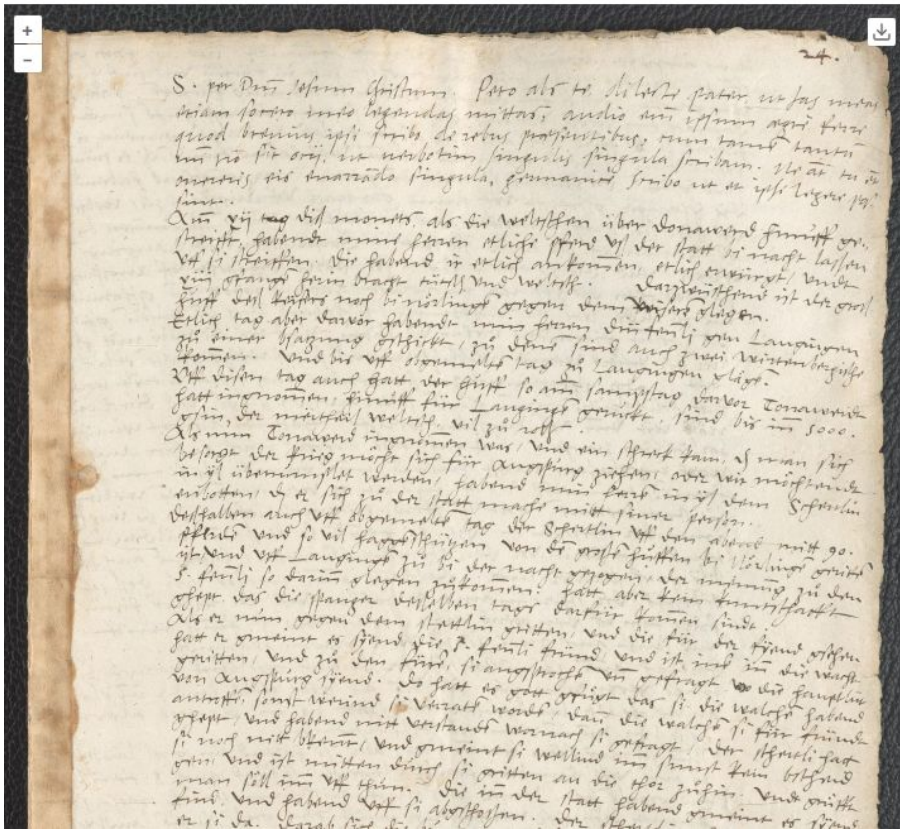
Regest ℹ

[1] Bullinger soll den vorliegenden Brief an **Hallers** Schwiegervater [**Ulrich Kambli**] schicken, der sich über **Hallers** zu kurze Nachrichten beklagt. Der Brief ist auf Deutsch verfasst, damit der Schwiegervater ihn lesen kann und Bullinger ihm nicht alle Einzelheiten erzählen muss. [2] Als die [kaiserlichen] Italiener am 7. Oktober Streifzüge in die Gegend von **Donauwörth** unternahmen, ließ die **Augsburger** Obrigkeit ihnen in der Nacht zu Pferd nachsetzen. Es wurden etliche Kaiserliche getötet und acht gefangen nach **Augsburg** gebracht. Das Haupttheer des Kaisers [**Karl V.**] lag damals noch bei **Nördlingen**, gegenüber den

Digital Humanities: Automatische Handschrifterkennung

Faksimile

Autograph: Zürich StA, E II 370,24 f.



Transkription ⓘ 🗨 Übersetzen 🗑 Sprachen markieren

Farbliegende: Deutsch (82.4%) Latein (17.6%)

S. per dominum Iesum Christum. Peto abs te, dilecte pater, ut has meas etiam socero meo¹ legendas mittas. Audio enim ipsum aegre ferre, quod brevius ipsi scribo de rebus praesentibus, cum tamen tantum mihi non sit ocii, ut verbotim² singulis singula scribam. Ne autem tu etiam oneraris eis enarrando singula, Germanice scribo, ut et ipsi³ legere possint.

Amm 7. tag diß monets, als die Weltschen⁴ über Donawerd hinuff gestreiff⁵, habendt mine herren etliche pferd uß der statt bi nacht lassen uff si streiffen.⁶ Die habend ir⁷ etlich ankommen⁸, etlich erwürgt undt 8 gfangen herin⁹ bracht, tütsch undt welsch. Darzüwüschend ist der groß huff¹⁰ deß keisers¹¹ noch¹² bi Nörlingen¹³ gegen dem unseren glegen.

Etlich tag aber darvor habendt min herren drü fennli gen Laugingen zü einer bsatzung gschickt. Zü denen sind auch zwei wirtenbergische kommen und bis uff obgmelten tag zü Laugingen glägen. Uff disen tag¹⁴ auch hatt der huff, so am sampstag darvor¹⁵ Tonawerd hatt ingnommen, hinuff für¹⁶ Laugingen geruckt. Sind bis inn 5000 gsin¹⁷, der mertheil welsch, vil zü roß.

Als nun Tonawerd ingnommen was und ein schreck kam, das man sich besorget, der krieg möcht sich für Augspurg ziehen oder wir möchtendt in yl überrumplet werden, habend min herren in yl dem Schertlin¹⁸ enbotten¹⁹, das er sich zü der statt²⁰ mache mitt siner person. Deßhalben auch uff obgmelten tag²¹ der Schertlin uff den abend mitt 90 pferden und so vil haggenschützen²² von dem großen huffen bi Nörlingen geritten ist und uff Laugingen zü bi der nacht gezogen, der meinung, zü den 5 fennli,

Forschungsgruppen

- Texttechnologie
- **Sprache, Technologie und Barrierefreiheit**
- Digitale Linguistik
- Phonetik & Sprachsignalverarbeitung
- Computerbasierte Neurowissenschaften der Sprache & des Hörens

Automatische Textvereinfachung

capito

Text mit KI vereinfachen

Sprachstufe wählen

leicht (B1) ▾

Ursprungstext

Vereinfachter Text

In der heutigen Zeit werden Texte in der schriftlichen Organisationskommunikation häufig unter Einsatz komplizierter Formulierungen und Fachjargon verfasst.

Firmen, Behörden und andere Gruppen schreiben oft sehr komplizierte Texte.

Schließen Vereinfachten Text kopieren

Übersetzung von Texten

Corona-Tests in Schulen, Kitas und Pflegeheimen

Schüler, Lehrer, Kita-Kinder und ihre Erzieherinnen können bald mehr auf den Corona-Virus getestet werden.

Das hat das Bundes-Gesundheits-Ministerium gesagt. Dann können Corona-Tests auch in Krankenhäusern, Pflegeheimen und Lebensmittel-Firmen gemacht werden. Das gilt für alle Menschen, die Kontakt mit Corona hatten.

Die Corona-Tests sollen auch für Menschen ohne Symptome

Wer bezahlt den Corona-Test?

- Der Arzt oder die Ärztin.
- Die Person, die Corona hat.
- Die Krankenkassa.
- Steht nicht im Text.

Automatische
Textverständnisprüfung

Automatisches Feedback für Gebärdensprachlernende

EN | DE | FR | IT

Wie es funktioniert

Katalog

Zusammenfassung

Linkshänder

stornay

VIOLETT (3)

→ VON

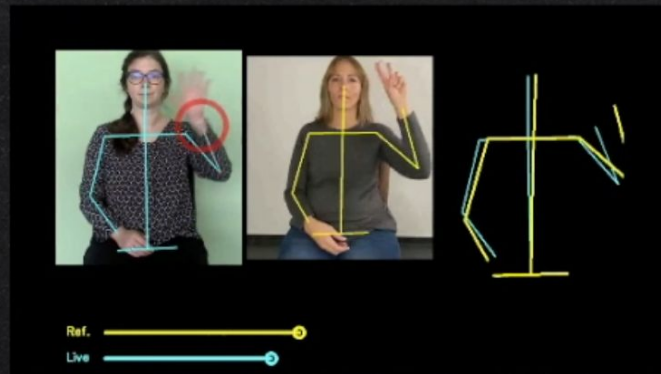
Live

Referenzvideo anzeigen

Ergebnisse ausblenden



Starte die Aufnahme



Gesamtpunktzahl



Nicht dominante Hand RICHTIG

Dominante Hand

Form

FALSCH



Stelle



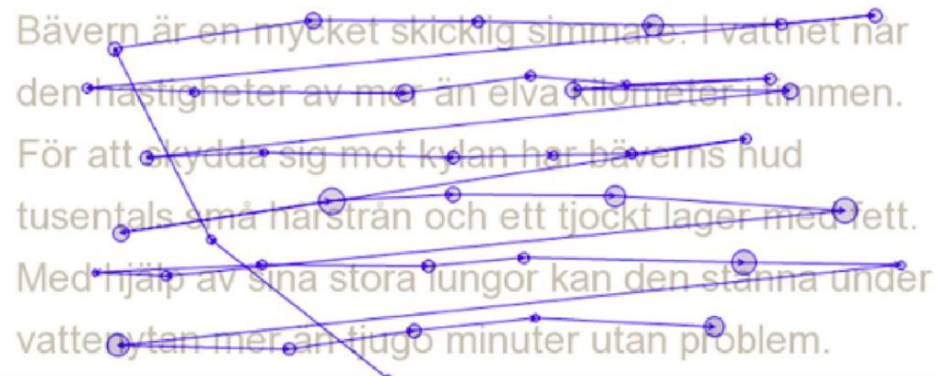
Forschungsgruppen

- Texttechnologie
- Sprache, Technologie und Barrierefreiheit
- **Digitale Linguistik**
- Phonetik & Sprachsignalverarbeitung
- Computerbasierte Neurowissenschaften der Sprache & des Hörens

Früherkennung von Legasthenie mit Eye-Tracking



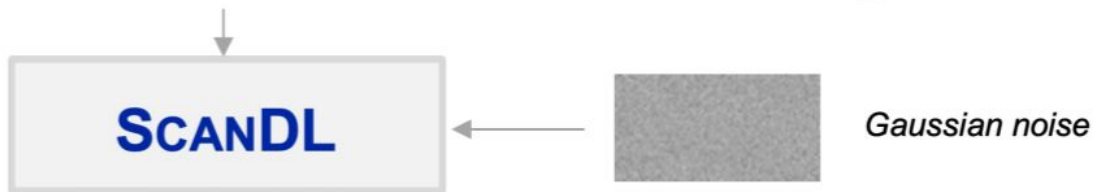
Früherkennung von Legasthenie mit Eye-Tracking



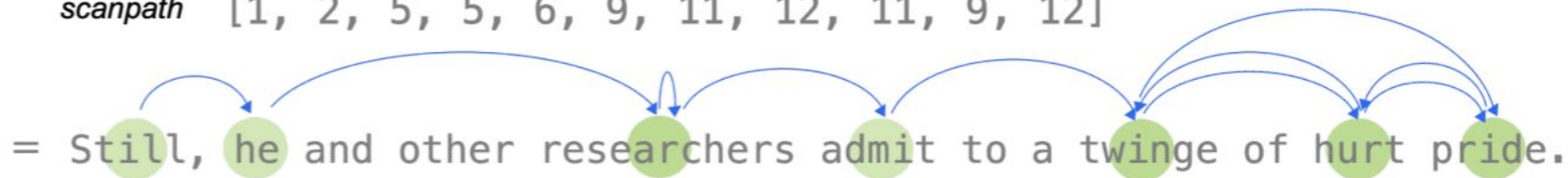
Je grösser der Kreis, umso länger blieb das Auge dort fixiert.

Automatische Vorhersage von Blickbewegungen

input text Still, he and other researchers admit to a twinge of hurt pride.



scanpath [1, 2, 5, 5, 6, 9, 11, 12, 11, 9, 12]

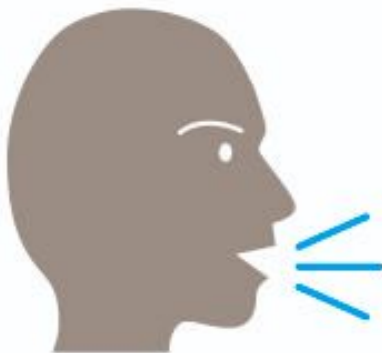


Forschungsgruppen

- Texttechnologie
- Sprache, Technologie und Barrierefreiheit
- Digitale Linguistik
- **Phonetik & Sprachsignalverarbeitung**
- Computerbasierte Neurowissenschaften der Sprache & des Hörens

Phonetik und Sprachsignalverarbeitung

Articulatory
phonetics



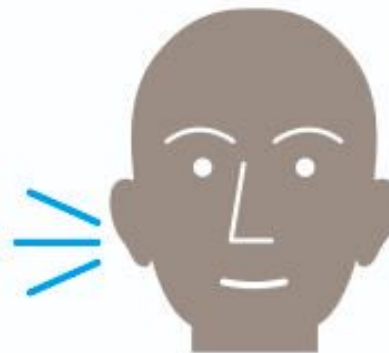
Wie produzieren wir Laute?

Acoustic
phonetics



**Welche Informationen sind im
Sprachsignal enthalten?**

Perceptual
phonetics



**Wie nehmen wir
gesprochene Sprache wahr?**

Phonetik



Phonetik: Anwendungen

Automatische Spracherkennung



Forensische Phonetik: Sprecherabgleich



Forschungsgruppen

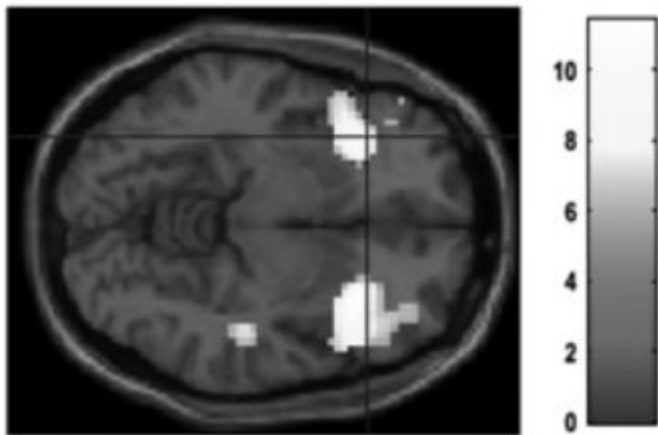
- Texttechnologie
- Sprache, Technologie und Barrierefreiheit
- Digitale Linguistik
- Phonetik & Sprachsignalverarbeitung
- **Computerbasierte Neurowissenschaften der Sprache & des Hörens**

Elektroenzephalografie (EEG): Hirnaktivität messen

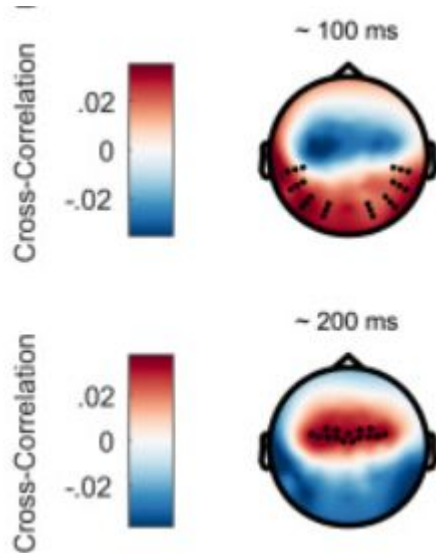


Neural Speech Tracking bei älteren Menschen mit Gehörverlust

Aktive Hirnareale bei der Sprachwahrnehmung mit fMRI:
Wo wird Wortbetonung verarbeitet?



Gehörverlust führt dazu, dass das Gehirn mehr “Speech Tracking” macht (Hirnaktivität schwingt im Rhythmus von Sprache mit)



Kompetenzzentren im Umfeld unseres Instituts

Centre for Forensic Phonetics and Acoustics (CFPA)

- Expertisen für Gerichte

Phonogrammarchiv

- Historische Dialektaufnahmen

LiRi: Linguistic Research Infrastructure

- Labore, Maschinen, Dokumentensammlungen, Sprachtechnologie

Hier sammeln Studierende als **Mitarbeitende** Erfahrungen.

Themenübersicht


1. Was ist Computerlinguistik und Sprachtechnologie?
2. Wer forscht und lehrt am Institut für Computerlinguistik?
3. **Was lernt man im Bachelorstudium "Computerlinguistik und Sprachtechnologie"?**
4. Was kann man nach dem Bachelorstudium machen?
5. Fragen?

Was muss ich mitbringen?

- **Faszination für Sprachen** und ihre Verarbeitung mit dem Computer
 - Kein Lateinobligatorium!
- **Keine Programmierkenntnisse, aber Freude an algorithmischem Denken**
 - Wir lehren Python!
- Interesse an der **Anwendung von maschinellem Lernen (AI)**
 - Keine Mathematik-Allergie!
- Bereitschaft, die **globale Wissenschaftssprache Englisch** in allen Vorlesungen und Fachtexten zu benutzen
- **Innovativer Spirit** für zukünftige Tätigkeitsfelder
 - Kein etablierter Beruf "Computerlinguist:in"!
Dafür *ready* für die **digitalen Transformation** unserer Gesellschaft.

Bachelor-Studium: "Computerlinguistik und Sprachtechnologie"

Ein interdisziplinärer Studiengang der Philosophischen Fakultät

- Abschluss "Bachelor of Arts UZH" mit 180 ECTS-Punkten (6 Semester Vollzeit) 
- Immer aufgeteilt auf Hauptfach (120 ECTS) und Nebenfach (60 ECTS)

Hauptfach "Computerlinguistik und Sprachtechnologie": 120 ECTS

- Vorlesungen aus allen Forschungsbereichen des Instituts
- Obligatorisch: Reine Informatikvorlesungen des Instituts für Informatik
 - *Software Engineering included!*

Nebenfach "Computerlinguistik und Sprachtechnologie": 60 ECTS:

- Vorlesungen aus allen Forschungsbereichen des Instituts
- Keine Informatikvorlesungen obligatorisch


Aufbau des Studiums: Module und Modulgruppen in "Computerlinguistik und Sprachtechnologie"

- Jede **Vorlesung** gehört in eine Modulgruppe.
- Pro **Modulgruppe** ist eine Mindestanzahl ECTS Pflicht.
- Einführende Vorlesungen sind **Pflicht**.
 - Programmieren
 - Maschinelles Lernen
- Der Rest ist **frei wählbar**.

Bachelorarbeit		15
Weitere Curriculare Module		0
Praxis der CL	0	6
Wissenschaftliche Vertiefung	0	6
Informatik	0	18
Kernbereich CL & SP	18	36
Einführung in die CL	30	30
Minimale Punktzahl pro Modulgruppe	Minor 60	Major 120

Nebenfach 30 ECTS: "Digital Humanities und Text Mining"

Wählst du ein **Hauptfach** aus der **Wirtschaftswissenschaftlichen (WWF)** oder der **Mathematisch-Naturwissenschaftlichen Fakultät (MNF)**?


- Im kleinen Nebenfach "Digital Humanities und Text Mining" kannst du Computerlinguistik-Vorlesungen belegen 
- Freie Modulwahl

Good To Know

Hauptfach oder Programmwechsel innerhalb der UZH sind jederzeit möglich!

Jedes bestandene Modul zählt automatisch für alle Studienprogramme, für die es zugelassen ist.

Wie findet man ein mögliches Nebenfach?

Schritt 1: Bestimme dein Hauptfach 



Studieren Lehren Forschen Entdecken Events News

Home / Studieren / Studienangebot / Studienprogramme Bachelor / Computerlinguistik und Sprachtechnologie

Computerlinguistik und Sprachtechnologie

Schritt 2: Wähle den "Major"

Schritt 3: Zeige alle "Minor-Programme"
– 62 Minors aus allen Fakultäten möglich!

Studienprogramme

Major ^

2

	Computerlinguistik und Sprachtechnologie Major 120
ECTS-Credits	120
Studiengang	Bachelor of Arts (RVO19)
Major-/Minor-Kombinationen	Der Bachelorstudiengang besteht aus einem Major-Studienprogramm von 120 ECTS Credits in Kombination mit einem Minor-Studienprogramm von 60 ECTS Credits. Das Minor-Studienprogramm kann an der Philosophischen Fakultät der UZH, an einer anderen Fakultät der UZH oder an einer anderen universitären Hochschule absolviert werden. Für allfällige Kombinationsverbote ist der programmspezifische Anhang zur Studienordnung zu beachten (s. Reglemente).

Minor-Programme verbergen ^

Minor 60

Über zulässige Major/Minor-Kombinationen gibt das Reglement Auskunft

- Allgemeine und Vergleichende Literaturwissenschaft
- Altertumswissenschaften
- Angewandte Wahrscheinlichkeit und Statistik
- Antike Religionsgeschichte

3

Wie lehren wir? Wie lernt ihr?

Formen

- Vorlesungen
- praktische Übungen in Teams
- Tutorate
- Studierendenpräsentationen
- Praktikum
- Sommerschulen

Gruppengrößen


- 40-60 Studierende in Einführungen
- 10-20 in Seminaren und Fachvorlesungen

Leistungsnachweise

- schriftliche Prüfungen
- Präsentationen
- schriftliche Arbeiten
- Programmierprojekte

Weiter Studieren? Alltag in der CL-Forschung

 Papers lesen, diskutieren, Forschungslücken finden

 Experimente planen

 Daten finden/erheben

 Experimente durchführen (Programmieren)

 Resultate auswerten (Statistik)

 Papers schreiben, publizieren

 An Konferenzen präsentieren

 Mit anderen Forschenden austauschen

How do we teach?

How do you learn?

- Lectures
- Practical exercises in teams
- Tutorials
- Student presentations
- Internships
- Summer schools
- Class sizes: 40-60 for introductory courses, 10-20 for seminars

Master-Studium: "Computerlinguistik und Sprachtechnologie"

Ein konsekutiver Studiengang

- Abschluss "Master of Arts UZH" mit 120 ECTS-Punkten (4 Semester Vollzeit) 
- CL als **Hauptfach** (90 ECTS) oder **Nebenfach** (30 ECTS)

Zulassung:

- Ein **Mausklick**: Wer mindestens CL BA "Computerlinguistik und Sprachtechnologie" als Nebenfach 60 studiert hat.
- Sonst: **Bewerbung und Zulassung mit Auflagen** (maximal 60 ECTS Punkte) aus Vorlesungen des CL BA "Computerlinguistik und Sprachtechnologie"

MA Hauptfach "Computerlinguistik und Sprachtechnologie": 90 ECTS

- Viel Wahlfreiheit

MA Nebenfach "Computerlinguistik und Sprachtechnologie": 30 ECTS:

Aufbau des MA Studiums: Module und Modulgruppen in "Computerlinguistik und Sprachtechnologie"

- Viel Gestaltungsraum im Major!
- Keine Pflichtmodule

Aber das ist jetzt Zukunftsmusik!

Master's Thesis		30
Other Curricular Modules		0
Computer Science	0	0
CL & LT in Practice	0	6
Scientific Specialization	6	6
Core Modules	18	24
Minimal ECTS Points per Module Group	Minor 30	Major 90

Themenübersicht

1. Was ist Computerlinguistik und Sprachtechnologie?
2. Wer forscht und lehrt am Institut für Computerlinguistik?
3. Was lernt man im Bachelorstudium "Computerlinguistik und Sprachtechnologie"?
- 4. Was kann man nach dem Bachelorstudium machen?**
5. Fragen?

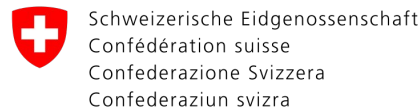
In den Medien, Industrie, Entwicklung oder Forschung arbeiten



Forschung & Entwicklung








Kanton Zürich

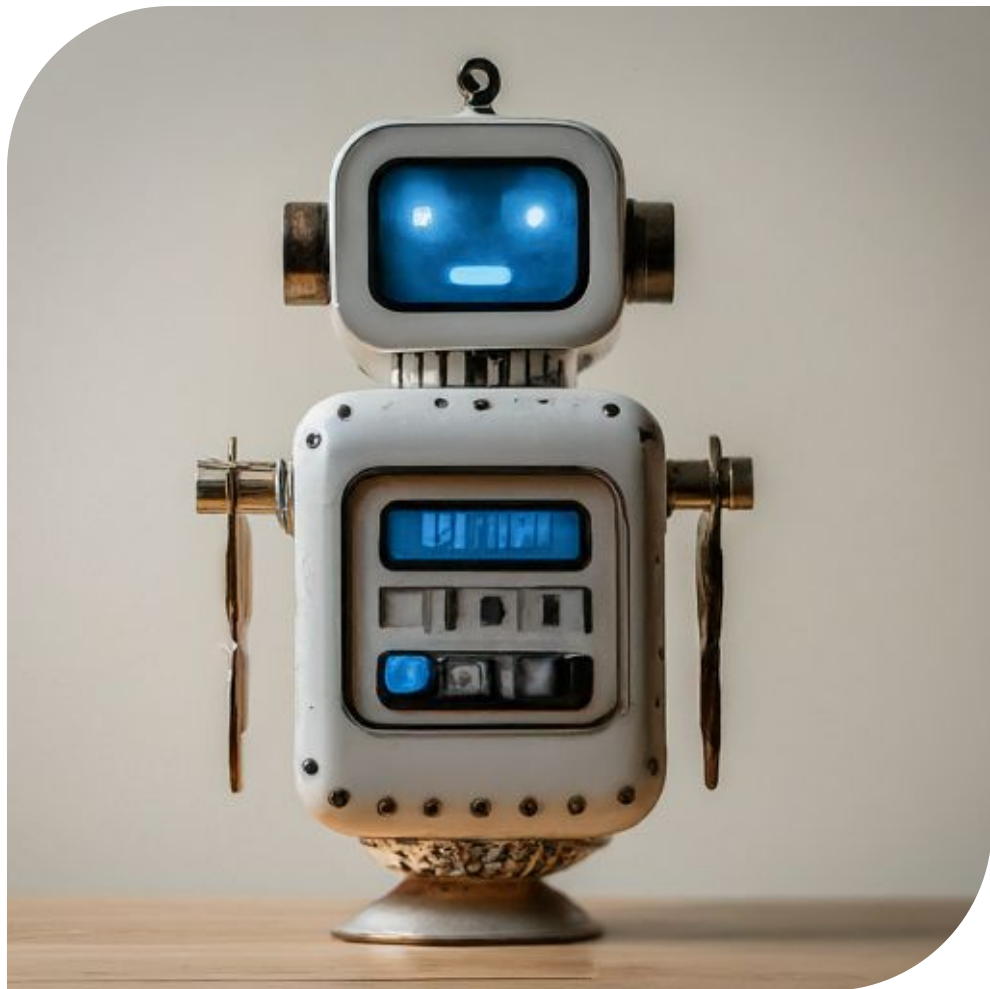


Archive & Medien

Generative AI und Chatbots

Neue Welle von Anwendungen

-  Kundendienste
-  Persönliche Assistenten
-  Bildung: Persönliche Tutoren
-  Gesundheitswesen: Berichte erzeugen, analysieren
-  Unterhaltung
- Deine innovative Idee...



Abheben mit der CL

Willst du zur wachsenden CL in Zürich gehören?

Das einzige BA-Programm für
Computerlinguistik in der Schweiz

- rund **80 Studierende** im Hauptfach
- und **20** im Nebenfach

Fragen?

<https://www.cl.uzh.ch>



Infos aus erster Hand

Sprechen Sie mit CL-Studierenden am
Stand 56!

Wir sind heute nur für euch da ...