

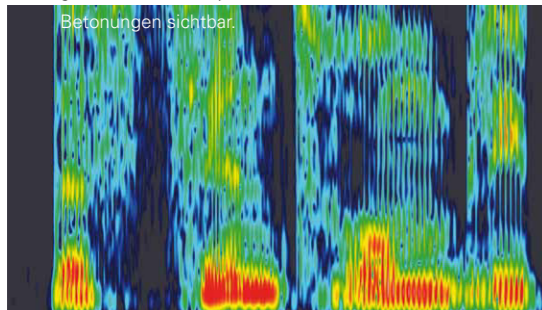
Maschinen als Helfer für den Geist

Digitalisierung schafft neue Zugangsmöglichkeiten zu Sprache und Literatur

Bücher, E-Mails, Blogs: Nie zuvor haben die Menschen so viel geschrieben und gelesen wie heute. Ein großer Teil der schriftlichen Kommunikation und Sprache läuft inzwischen aber über elektronische Kanäle. Darauf müssen sich Sprach- und Literaturwissenschaften einstellen. An mehreren Instituten der Universität Stuttgart leisten Geisteswissenschaftlerinnen und -wissenschaftler daher Pionierarbeit mit elektronischen Forschungsprojekten. Sie machen 700 Jahre alte Texte für jeden verfügbar, verhelfen zu neuen Erkenntnissen über Autoren und Quellen oder tragen dazu bei, Missverständnisse in der Kommunikation auszuräumen.

Wer auf seinem Mobiltelefon eine Kurznachricht schreibt, bekommt heute schon nach den ersten Buchstaben das komplette Wort vorgeschlagen. Damit ein Programm die richtigen Wörter vorschlägt, müssen Sprachwissenschaftler die Regeln der Grammatik erarbeiten. Gemeinsam mit Computerlinguisten gilt es dann, den Maschinen diese oft komplizierten Regeln beizubringen. Unter anderem mit diesen Regeln und mit statistischen

Experimentelle Phonetik: Spektrogramme machen Sprache und Betonungen sichtbar.



Worthäufigkeiten beschäftigen sich rund 50 Forscherinnen und Forscher im Sonderforschungsbereich 732 „Incremental Specification in Context“, dem derzeit größten Projekt der elektronischen Geisteswissenschaften, der „E-Humanities“, an der Uni Stuttgart. Unter der Federführung von Prof. Artemis Alexiadou vom Institut für Linguistik und mit einem Budget von jährlich zwei Millionen Euro von der Deutschen Forschungsgemeinschaft (DFG) gehen sie der Frage nach, wie man Mehrdeutigkeiten in der Sprache erkennen und auflösen kann. Dazu kooperieren die Sprachforscher mit den Computer-Fachleuten um Prof. Grzegorz Dogil und Prof. Jonas Kuhn vom Institut für Maschinelle Sprachverarbeitung (IMS), internationalen Forschungseinrichtungen sowie den Unternehmen Sony und Google.

Eine ihrer wichtigsten Fragen ist die Bedeutung, in welcher ein Autor ein Wort verwendet. „In jeder Sprache gibt es Wörter, die mehrere Bedeutungen haben“, erklärt Alexiadou. „Zuerst einmal erforschen wir, was diese mehreren Bedeutungen sind.“ In einem zweiten Schritt finden die Linguisten heraus, in welchen Kombinationen diese Wörter auftreten. „Dann wollen wir aus der Sicht der theoretischen Linguistik erklären, wieso das so ist, warum manche Wörter diese Möglichkeit haben und andere nicht.“ Gleichzeitig untersuchen die Computerlinguisten, ob man etwa mit Hilfe der Statistik vorhersagen kann, wie oft sich die Wörter in dieser Bedeutung in einem Text finden. Das Ziel sind korrekte maschinelle Übersetzungen. „Wenn Sie beispielsweise an das Europaparlament denken, dort müssen alle Vorlagen übersetzt werden. Dabei hat man das Problem, wie ein Wort exakt in der anderen Sprache wiedergegeben wird“, sagt Alexiadou. „Wenn die Maschine das wortwörtlich übernimmt, kommen manchmal total falsche Sätze heraus.“

Aufnahme eines Dialogs im schalltoten Raum im Rahmen des Sonderforschungsbereichs 732. Durch die Glasscheibe wird eine saubere Trennung der Sprachsignale gewährleistet. Das Projekt untersucht, wie sich Gesprächspartner im Dialog phonetisch aneinander anpassen.



IM ZIMMER, INS ZIMMER, ZUM BAHNHOF... Diese Schwächen will die Gruppe mit einer Kombination aus statistischen Verfahren und den Regeln der theoretischen Linguistik lösen. Vor welchen Schwierigkeiten die Zweiertteams des Graduiertenkollegs aus Sprachwissenschaftlern und Computerlinguisten stehen, macht die Professorin am Satzbau deutlich: „Bei den meisten Präpositionen ist es so: Wenn Sie den Dativ nehmen, ist es lokal, wenn Sie den Akkusativ nehmen, gibt er eine Richtung an“, erklärt Alexiadou. Also „ich tanze im Zimmer“ oder „ich tanze ins Zimmer“. Daraus könnte man eine Regel ableiten. „Aber dann haben Sie die Probleme mit einer Präposition wie *zu*. Diese nimmt den Dativ, gibt aber eine Richtung an“, verdeutlicht die Linguistin: „Ich gehe zum Bahnhof.“ Diese Vielfalt müssen die Maschinen erst noch lernen. Eine besondere Herausforderung stellt die Kommunikation dar, in der die eigentliche Information weglassen wird, wie Alexiadou an einem Alltagssatz verdeutlicht: „Wenn Sie zu mir sagen, dass es heute heiß ist, wollen Sie mir

damit vielleicht bedeuten, dass ich ein Fenster öffnen soll.“ Um Regeln für diese Art von Sprache zu finden, werten die Computerlinguisten auch Nachrichtentexte und Bundestagsreden aus. Langfristig sollen Maschinen in der Lage sein, selbst die Informationen korrekt zu verarbeiten, die der menschliche Geist verpackt und versteckt. Im dritten Teil des Sonderforschungsbereichs, der sich von 2014 an für vier weitere Jahre anschließen soll, wollen die Forscher daher die gesammelten Daten und Erkenntnisse zusammenführen und die Arbeit mit der Sprache weiter vereinfachen, wie Alexiadou hofft: „Teile davon werden sicher dazu beitragen, dass maschinelle Übersetzungen noch besser werden.“

ZÄHLEN STATT HERMENEUTIK

Auch in der Forschung über Literatur eröffnen Computer Möglichkeiten, die dem Geist bisher verwehrt blieben. Alte Texte und deren neue Nutzungsmöglichkeiten erschließen daher zwei Projekte am Institut für Literaturwissenschaft: Prof. Sandra Richter, Abteilung Neuere deutsche

Einleitung	Seite 60	Seite 60	Wehe ihm, wenn er sich einfallen lässt, auf seine Führerschaft zu pochen!
	Seite 61	Seite 61	Dann wird ihm erwidert, er habe nichts zu befehlen, man sei ihm freiwillig gefolgt.
	Seite 62	Seite 62	Eine Verpflichtung gehe es nicht.
	Seite 63	Seite 63	Jeder könne, sobald es ihm beliebt, wieder von dannen ziehen.
	Seite 64	Seite 64	In ähnlichem Verhältnis steht Zeus, der Göttervater, zu den Göttern.
	Seite 65	Seite 65	Gesanges prahlt er zwar in einer gewaltigen Rede, er sei in stände, das Meer und die Erde samt allen Göttern, die sich daran hängen wollten, in die Lüfte zu reißen.
	Seite 66	Seite 66	Übertreffe ja ich gewaltig Götter und Menschen.
	Seite 67	Seite 67	In diesen Versen scheint sich jedoch ein alterer Mythos erhalten zu haben.
	Seite 68	Seite 68	von der Homer sonst nichts mehr weiß.
	Seite 69	Seite 69	Im übrigen ist es mit der Macht des Zeus durchaus nicht so gut bestellt.
Lyrischer Stil: Erinnerung	Seite 70	Seite 70	Es wird zwar ständig versichert, dass alle Entscheidungen in seinen Händen ruhe.
	Seite 71	Seite 71	Hera, Ares, Athene, Poseidon jedoch sind anderer Meinung, murren, wenn Zeus Befehle erteilt, und erkühnen sich gar, mit List und Betrug den Willen des Höchsten zu umgehen.
	Seite 72	Seite 72	Dann muss sich Zeus gleichfalls mit Schlaueit oder mit Poltern und Drohen behelfen.
	Seite 73	Seite 73	Das Schauspiel ist peinlich für den Herrn.
	Seite 74	Seite 74	Doch eben deshalb treten sämtliche Götter und Helden so herrlich hervor.
	Seite 75	Seite 75	Sie sind nicht auf den Einen bezogen.
	Seite 76	Seite 76	Jeder hat seine besonderen Wünsche und Angelegenheiten.
	Seite 77	Seite 77	Jeder ist eine frei entfaltete Individualität.
	Seite 78	Seite 78	Ebenso bewahrt der Mensch gegenüber den Göttern Selbstständigkeit.
	Seite 79	Seite 79	Man hat Homer zwar schon im Altertum nachgesagt, seine Helden seien Marionetten in den Händen der Himmlichen.
Epischer Stil: Vorstellung	Seite 80	Seite 80	Wer aufmerksam liest, merkt jedoch bald, dass ein solcher Tadel nicht am Platz ist.
	Seite 81	Seite 81	Allerdings heisst es oft, ein Gott habe dies dem Menschen eingegeben;
	Seite 82	Seite 82	er habe seinen Verstand besorgt oder seinen Sinn zum Guten gelenkt.
	Seite 83	Seite 83	Doch das schliesst Freiheit des Handelns nicht aus.
	Seite 84	Seite 84	Der Mensch kann sich dem Willen der Götter fügen oder widersetzen.
	Seite 85	Seite 85	Er selbst trägt die Verantwortung und ist sich dessen durchaus bewusst.
	Seite 86	Seite 86	Und so geht es sogar noch weiter hinab.
	Seite 87	Seite 87	
	Seite 88	Seite 88	
	Seite 89	Seite 89	
Dramatischer Stil: Spannung	Seite 90	Seite 90	
	Seite 91	Seite 91	
	Seite 92	Seite 92	
	Seite 93	Seite 93	
	Seite 94	Seite 94	
	Seite 95	Seite 95	
	Seite 96	Seite 96	
	Seite 97	Seite 97	
	Seite 98	Seite 98	
	Seite 99	Seite 99	

Hierarchische Visualisierung von Emil Staigers ‚Grundbegriffe der Poetik‘ (Zürich 1946) in Kapitel, Seiten und Sätze.

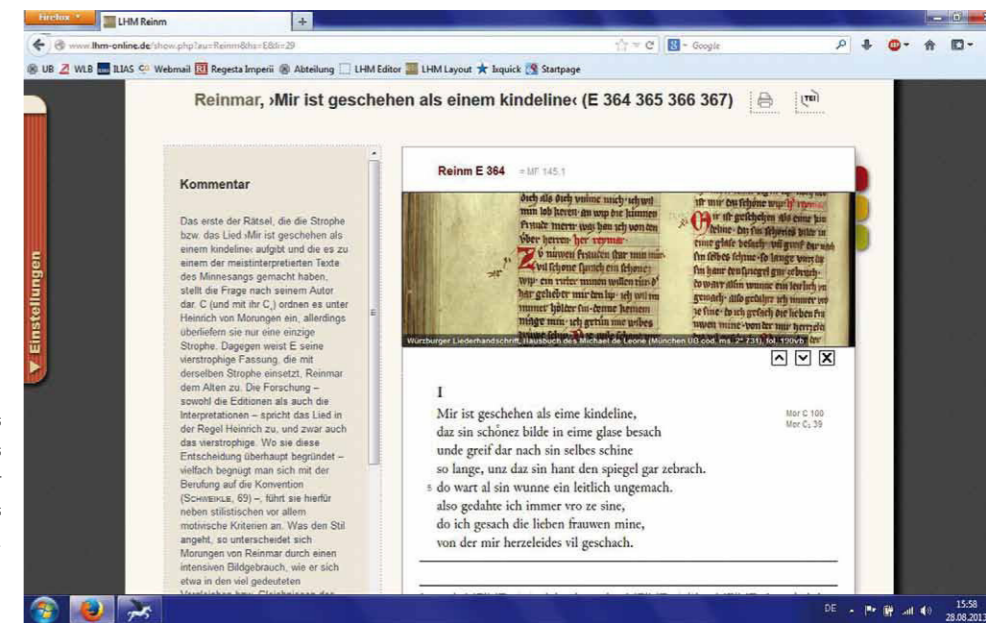
Literatur, hat sich gemeinsam mit dem IMS und den Informatikern vom Institut für Visualisierung und Interaktive Systeme der Universität Stuttgart vorgenommen, die Entwicklung der Wissenschaft über die Dichtkunst elektronisch zu analysieren. Dazu wollen die Forscher 20 repräsentative Texte aus der Literaturwissenschaft vom 18. bis zum 20. Jahrhundert digitalisieren und neue Werkzeuge daran testen und entwickeln. 650.000 Euro stellt das Bundesministerium für Bildung und Forschung für das Projekt „ePoetics“ zur Verfügung. Die Begründung: Es wird die wissenschaftlich bedeutsamen Texte digital verfügbar machen. Zudem will das Team, an dem sich auch die Technische Universität Darmstadt beteiligt, Programme entwickeln, mit denen man jede Art von argumentierendem Text digital analysieren lassen kann. Denn bei der Frage, wie man Texte so erfasst, dass man sie wissenschaftlich bearbeiten kann, steckt die Wissenschaft noch in den Kinderschuhen. „Vieles läuft über Quantitäten“, sagt Sandra Richter über die Vorteile der E-Humanities. „Und das ist ein Bereich, über den in den hermeneutischen Wissenschaften bisher überhaupt noch nicht ausreichend diskutiert worden ist. Da geht man bisher qualitativ vor und zählt nur selten.“ Um ein Werkzeug zu entwickeln, das solche quantitativen Verfahren und auch komplexe Funktionen beherrscht, hat sich die Literaturwissenschaftlerin die Poetiken seit Ende des 18. Jahrhunderts vorgenommen und die Frage, wie sich diese bis ins 20. Jahrhundert gewandelt haben – etwa, welche Schlüsselbegriffe aufkamen. Das Korpus der Poetiken und Ästhetiken von 1770 bis 1960 dokumentiert das Denken und Schreiben über Literatur und andere Künste für fast 200 Jahre.

Zunächst werden die Texte in Darmstadt erfasst. Anschließend entwickeln acht Informatiker, Fachleute für maschinelle Sprachverarbeitung und Literaturwissenschaftler Analyse-Tools. Entstehen soll eine Kombination aus algorithmischen und computergestützten mit traditionell hermeneutischen Methoden des Textverständnisses. Mit anfangs 20 ausgewählten Texten will Richter ihr Programm entwickeln und testen. Später soll es auf ein größeres Korpus ausgeweitet werden. Welche Erkenntnisse durch diese Kombination möglich werden, erklärt Richter am Beispiel der Erhabenheit in der Betrachtung der Literatur: „Wenn wir schauen, wie oft taucht der Begriff des Erhabenen in den Poetiken des ausgehenden 18. Jahrhunderts auf, dann ist das ziemlich häufig. Nun fragt man sich: Was heißt es, wenn in einem Text 100 Mal der Begriff des Erhabenen vorkommt? Da kann man hermeneutisch sagen, dieser Umstand ist begründet in der Rezeption von Edmund Burke und Immanuel Kant“, so Richter. „Damit lernen wir, dass diese Poetiken ganz zentral jene beiden Texte in den Mittelpunkt rücken. Das hat Konsequenzen für die Behandlung des gesamten Wissensbereiches, weil man sagen kann: Damals hat die ältere Poetik, die diese Konzepte ignorierte, sich umgestellt und ist zu einer Art moderner Wissenschaft geworden.“

DIGITALER MINNESANG

Die digitale Analyse von Texten zeigt auch den Weg auf für die Lösung eines Problems, das der Deutsche Wissenschaftsrat aufgeworfen hat: Was wird aus Bibliotheken, Archiven und Datenbanken im digitalen Zeitalter? Wie eine solche Lösung aussehen kann, führt Prof. Manuel Braun,

Die digitale Aufbereitung des Liedes „Mir ist geschehen als einem kindelîne“ hilft bei der Klärung der Autorenschaft des vielinterpretierten Minnesangs.



Abteilung für Germanistische Mediävistik, an seinem Computer vor – wenn auch erst in einer Demoversion. Mit dem Projekt „Lyrik des hohen Mittelalters. Eine exemplarische elektronische Edition“, hat sich Braun vorgenommen, den Minnesang in Teilen digital herauszugeben. Über den Strophen- oder Autorennamen wählt der Altgermanist in dem Programm zunächst ein Lied Walthers von Mezzese aus. Dieses kann er als Digitalisat aus der Handschrift Codex Manesse oder als Transkription aufrufen. Ein Kommentarfeld informiert über die Überlieferung des Liedes und die Sekundärliteratur. Die vier verschiedenen Abschriften, in denen das Lied weitergegeben wurde, lassen sich in beliebiger Reihenfolge nebeneinanderstellen und vergleichen.

HANDSCHRIFTEN WERDEN LESBAR

Mit wenigen Klicks macht Braun den handschriftlich überlieferten Text für heutige Augen lesbar, indem er Eigenheiten der mittelalterlichen Schreibpraxis an heutige Gewohnheiten anpasst, etwa die Darstellung des Graphems ö durch die Übereinanderschreibung von e und o, die fehlende Unterscheidung der Grapheme u und v oder den fehlenden Zeilenumbruch.

Für die Edition arbeitet Braun mit Prof. Florian Kragl und Dr. Sonja Glauch von der Universität Erlangen zusammen. Das von der DFG mit 200.000 Euro geförderte Projekt soll zeigen, welche enormen Vorteile es hat, wenn man die zwischen den Jahren 1150 und 1300 entstandenen Texte digital herausgibt. Zunächst planen die drei, neun Autoren des Minnesangs aus dem 13. Jahrhundert digital zu transkribieren. Danach werden 1.254 von 10.000 Strophen der mittelhochdeutschen

Lyrik im Internet zur Verfügung stehen. Die Autoren sind allesamt in verschiedenen Abschriften überliefert, weshalb die digitale Edition hier ihre Stärken ausspielen kann: Existieren von einem Lied mehrere Versionen, lässt sich das im gedruckten Buch nur im Paralleldruck darstellen. Dabei müssen die Texte in einer festen Reihenfolge angeordnet werden. „Bei einer elektronischen Ausgabe habe ich die Möglichkeit, das hin und her zu schalten“, sagt Braun. Zudem soll das Programm über ein Werkzeug verfügen, das jedes Wort gleich in seiner Stammform anzeigt. „Dies ist wichtig für die Suchfunktion“, erklärt Braun. „Denn wenn Sie beispielsweise wissen wollen, wie oft das Wort ezzen vorkommt, wollen Sie ja auch Formen wie az oder hât gezzen mit erfassen.“ So lässt sich ohne großen Aufwand herausfinden, was bisher fast unmöglich war: Was waren die meistgebrauchten Wörter eines Autors? Hatte er einen reichen oder eingeschränkten Wortschatz? Für Braun liegt der Vorteil der E-Humanities auf der Hand: „Früher hätte man die Faksimiles aus der Bibliothek holen, nebeneinander legen und vergleichen müssen.“ Heute geht das am Bildschirm mit wenigen Klicks.

Daniel Völpel

www.uni-stuttgart.de/linguistik/sfb732/
www.uni-stuttgart.de/literat/