

Die Informatik als Geisteswissenschaft

John Nerbonne

Abstract dt./engl.

Der Titel des Beitrags bezieht sich indirekt auf das Thema der 2014 Konferenz der Digital Humanities im deutschsprachigen Raum (DHD): »Digital Humanities – methodischer Brückenschlag oder ›feindliche Übernahme?‹«. Ich gehe dabei von der Unterstellung aus, dass man sich die Informatik als Urheber einer solchen gedachten Übernahme vorstellt. Im ersten Abschnitt des Beitrags wird zunächst dargelegt, dass auch die Informatik, als Hauptbastion der ›Feinde‹, die unsere Disziplinen zu übernehmen drohen, eine unsichere Reifungsperiode gehabt hat, die sie jedoch glänzend überstanden hat. Man könnte demnach Inspiration oder wenigstens Mut aus der Geschichte dieser Disziplin schöpfen. Ein weiterer Abschnitt geht – getreu dem Titel des Beitrags – der Frage nach, wieso man die Informatik als Geisteswissenschaft betrachten kann, einschließlich einer historischen Betrachtung der Entstehungsperiode der zwei Disziplinen oder Disziplinengebiete. Die historische Dimension, aber auch die Unsicherheit, die sich im Konferenzthema widerspiegelt, nehme ich zum Anlass, um in einem sich daran anschließenden Teil zehn Thesen zu präsentieren, die erläutern, worauf es ankommen wird, wenn wir mit den Digital Humanities Erfolg haben wollen.

The title of this paper indirectly refers to the theme of the 2014 Conference on Digital Humanities in German-Speaking Areas: »Digital Humanities – methodischer Brückenschlag oder ›feindliche Übernahme?‹«. I assume from this title that computer science is viewed as the originator of such a deliberate conquest. The first section of this paper explores how computer science, the headquarters of the “enemy” threatening to conquer our disciplines, also experienced and overcame an uncertain period of early development. The history of this discipline could serve as a source of inspiration or, at the very least, courage. The following section explores – as indicated by the title of this paper – how computer science could be viewed as a humanities discipline, including a historical perspective on the foundation of both disciplines or disciplinary fields. The historical dimension, but also the uncertainty, that is reflected in the conference theme serves as my starting point to present ten theses in the following section that explain the necessary factors for success if we wish to succeed with the Digital Humanities.

1. Motivation

Die Unsicherheit, die im Titel der Konferenz –»Digital Humanities – methodischer Brückenschlag oder ›feindliche Übernahme?‹« – mitschwingt, ist im Zuge einer großen (informationstechnischen) Veränderung durchaus üblich: Traditionelle Fachleute verlieren an Bedeutung und neuartige Spezialisten werden wichtiger, manchmal sogar wichtiger als etablierte Autoritäten. Elisabeth Eisenstein hat die bis dato folgenreichste Änderung in der Informationstechnologie, das Entstehen der Buchdruckkunst, 1979 *en detail*

dargelegt.¹ Der Einzug der Informatik in viele andere Wissenschaftsgebiete, wie z.B. die Biologie, hat ähnlich große Verschiebungen in den dazu nötigen Expertisen verursacht, wobei manch kompetente Vertreter älterer Richtungen von den neueren Entwicklungen überrannt worden sind (so die Feldbiologie angesichts der Entstehung der Molekularbiologie und der Bioinformatik).

Zweck dieses Beitrags ist es nicht, die durch solche Entwicklungen und Umschwünge vermittelten Unsicherheiten kleinzureden oder deren Problematik gar zu leugnen, denn im aktiven Wissenschaftsdiskurs wird sich letztlich auch in diesem Bereich die Spreu vom Weizen trennen. Wenn wir – oder eben andere, jene titulierten ›Feinde‹, – wissenschaftlichen Fortschritt in den Geisteswissenschaften mittels Informationstechnologie erzielen können, dann werden sich die digitalen Geisteswissenschaften durchsetzen. Ob gerade eine feindliche Übernahme von statten geht, möchte ich bezweifeln, aber eine groß angelegte digitale Wende in den Geisteswissenschaften würde sicher einiges ändern, und nicht jeder Geisteswissenschaftler oder jede Geisteswissenschaftlerin würde gleichermaßen hiervon profitieren. Gerhard Wolf kritisiert schon jetzt, dass »Digitalprojekte Ressourcen [verschlingen], die wir dringend für unser Kerngeschäft nötig haben.«²

2. Die Klassifizierung der Wissenschaften – wozu gehört die Informatik?

1806/07 hat Hegel in seiner *Phänomenologie des Geistes* bekanntlich eine strenge Unterscheidung zwischen Natur und Geist gezogen und darüber hinaus auch abstrakteren Entitäten, wie einer Kultur oder einem Volk, einen objektiven Geist zugesprochen, den man demzufolge auch studieren kann.³ Dem nachfolgend verdanken wir Dilthey (1883) die Unterscheidung zwischen Naturwissenschaften und Geisteswissenschaften⁴: Naturwissenschaften wie Physik oder Mineralogie haben das Studium der Natur zum Ziel, d.h. alles, was ohne Einfluss des Menschen Bestand hat. Dilthey übernimmt von Hegel aber auch die Idee des kollektiven Geistes, wonach die Geisteswissenschaften nicht nur den individuellen Geist zu studieren haben, sondern auch alles, was Menschen zusammen machen, worunter beispielsweise Kulturgeschichte oder Architektur fallen. Es sei die Aufgabe der Geisteswissenschaften, so Dilthey weiter, die Erzeugnisse des menschlichen Geistes zu verstehen. Für Dilthey gehören somit auch die Sozialwissenschaften zu den Geisteswissenschaften, und er entwickelt weitere Ideen über geeignete Methoden innerhalb dieser zwei Gruppen von Disziplinen: Die Entwicklung von erklärenden Gesetzen in den Naturwissenschaften und den Gebrauch einer auf das Verstehen abzielenden Hermeneutik in den Geisteswissenschaften.⁵

Diese Unterscheidung lebt heute in der französischen Einteilung von *sciences naturelles* und *sciences humaines* fort, die wiederum die Basis für die Unterteilung in *Naturwissenschaften* und *Letteren* (>Geisteswissenschaften<) der niederländischen

¹ Vgl. für die weitere Reflexion zu den Parallelen zwischen den wissenschaftlichen Konsequenzen der Erfindung des Buchdrucks im 15. Jahrhundert einerseits und der Einführung der digitalen Informationsverarbeitung im 20. Jahrhundert andererseits auch meine Ausführungen in *Elektronische Incunabeln* 1995

² Zit. nach Berg 2014, S. 17.

³ Hegel ¹1806, 1988, Kap. VI. Vgl. „Der Geist ist das *sittliche Leben* eines Volks“

⁴ Dilthey ¹1883, 1959, S.14ff.

⁵ Leezenberg / de Vries (2001) referieren diese Entwicklung mit weitaus größerer Kompetenz, als ich es an dieser Stelle leisten kann.

Akademie der Wissenschaften ([KNAW](#)) bildet.

Die Informatik beschäftigt sich mit digitaler Informationsverarbeitung, sowohl in einzelnen Prozessen (Algorithmen oder Rechnern) als auch in Organisationen. Digitale Informationsverarbeitung aber ist ein Erzeugnis des kollektiven menschlichen Geistes, genau wie der mittelalterliche Städtebau, das Blues-Schema in der Musik, der Handelsbund der Hanse oder das moderne Bankwesen. Daraus folgt, dass die Informatik als wissenschaftliches Studium der Informationsverarbeitung zu den Geisteswissenschaften gehört, und ergo eine Geisteswissenschaft ist.

Diese Zuordnung der Informatik zu den Geisteswissenschaften impliziert nicht, dass die Informatik keine exakte (mathematische) Wissenschaft ist; denn auch andere eindeutige Beispiele von nicht den Naturwissenschaften zugehörigen Fachdisziplinen machen ausführlichen Gebrauch von mathematischen Modellen. Man denke an die Ökonometrie, die Architektur oder die Phonetik. Hinzukommt, dass gewisse Subdisziplinen der Informatik wie die Softwareentwicklung⁶ oder die Software-Ergonomie, einschließlich dem Entwurf von Benutzerschnittstellen⁷, eher sozialwissenschaftliche Methoden gebrauchen.

Die Schlussfolgerung oben ist qua Struktur ein disjunktiver Syllogismus: Aus $A \vee B$, $\sim A$ folgt B . Aus A oder B und *nicht* A , folgt B . Natürlich kann man einwenden, dass die Disjunktion, die Informatik sei entweder eine Naturwissenschaft oder eine Geisteswissenschaft, gar nicht zwingend sei. Man könnte auf eine dritte Möglichkeit als fachlicher Heimat der Informatik verweisen, etwa auf die Ingenieurwissenschaft oder die Sozialwissenschaft, doch dies würde nur manchen, aber lange nicht allen, vermutlich nicht einmal den meisten Informatikern gefallen.

Der wichtigere Einwand ist jedoch, dass die Klassifizierung in Natur- versus Geisteswissenschaften obsolet geworden ist.⁸ Die methodologischen Folgen der Unterscheidung, die noch für Dilthey Relevanz hatten, gelten schon lange nicht mehr. In vielen geisteswissenschaftlichen Fachdisziplinen wurden mittlerweile Gesetzmäßigkeiten erforscht und angewendet, um Phänomene in dem jeweiligen Gebiet zu erklären. Man denke nur an das Gesetz des sinkenden Ertrags in der Ökonomie oder an die Lautgesetze der Junggrammatiker. Umgekehrt spielt Intuition vermehrt in den Naturwissenschaften eine Rolle, wenn Forscher nach den Zusammenhängen ihrer Ideen fragen: Ein Schritt, der auf das Verstehen gerichtet ist. Auch das Gedankenexperiment kann von Belang sein. Die Klassifizierung ist demnach in beiden Richtungen unzureichend. Wenn wir erklären, dass eine neue Disziplin als Naturwissenschaft und nicht als Geisteswissenschaft klassifiziert wird, dann wissen wir sehr wahrscheinlich nach eingehender Prüfung und Präzisierung des Fachgegenstands, seiner Fragestellungen und Methoden, dass die Disziplin den akzeptierten Naturwissenschaften – Physik und Chemie – ähnlicher ist als den akzeptierten Geisteswissenschaften – Geschichte und Literaturwissenschaft –, aber viel mehr als das wissen wir nicht.

⁶ Bourque / Fairley 2014.

⁷ Shneiderman / Plasant 2010.

⁸ Von dieser Grundannahme geht auch Simons in *The Sciences of the Artificial* (1969) aus, in dem er Informatik und Ökonomie zusammen in eine Klasse setzt, weil sie beide künstliche Systeme ins Auge fassen, deren Komplexität es unmöglich macht, sie nur aus ihren Bauprinzipien heraus zu verstehen.

3. Die Unsicherheit und ihre Überwindung

Auf Unterschiede zwischen den Disziplinen in Fragen von Prestige, Finanzierung und Ähnlichem möchte ich hier nicht eingehen, auch wenn klar ist, dass es manchen Kollegen vor allem darum geht.

Es ist zu bemerken, dass auch die Informatik eine Periode der Unsicherheit gekannt hat, die u.a. dadurch charakterisiert ist, dass man sich über die Klassifizierung stritt. Sollte Informatik bei der Mathematik angesiedelt werden – weil es schließlich um Rechner und das Rechnen (nicht?) geht – oder aber bei den Ingenieurwissenschaften? Als ich 1985(!) eine Stelle als Informatiker bei Hewlett-Packard annahm, einer vom Ursprung her elektrotechnischen Firma, habe ich von den Elektro-Ingenieuren gehört, dass das Fach Informatik für sie vielleicht ›in‹ wäre, aber noch nicht substantiell. 1985 waren Computer und Informatik schon gut etabliert und wirtschaftlich bereits ein Faktor. Der Ruf, eine anerkannte Wissenschaft zu sein, musste erst noch kommen.

Wir können aber auch Inspiration aus der Art und Weise ziehen, wie andere mit der Konstitution eines neuen Fachs umgehen. John Hopcroft erzählte von seinen ersten Jahren in Princeton. Sein Abteilungsleiter hatte ihn gebeten, einen Kurs in Informatik zu geben, ohne genauer mitzuteilen, was für Material der Kurs beinhalten sollte. Hopcroft fand es merkwürdig, dass Princeton einen neuen Kurs einführen wollte, ohne dessen Inhalt zuvor zu spezifizieren. Später resümierte er jedoch: »In retrospect, I realize that people who believe in the future of a subject and who sense its importance will invest in the subject long before they can delineate its boundaries.«⁹

Man könnte es sogar radikaler formulieren, denn Hopcroft spricht von einem Lehrfach (*subject*), während viele gar nicht bereit wären, die Digitalen Geisteswissenschaften als Fach anzuerkennen. Man bemängelt die Verschiedenheit der Methoden und Themen und auch die manchmal fehlende Beziehung zu nicht-digitalen Studien.¹⁰ Auch hier sehe ich keinen Grund zur Unruhe. Viele anerkannte Fächer, z.B. Informatik, aber auch Medizin, sind von innen betrachtet sehr unterschiedlich, und es sollte klar sein, dass man mittels digitaler Studien manchmal neuartige Fragestellungen erproben will und kann – auch dort, wo die Verbindung zu traditionellen Fächern nicht auf der Hand liegt.

Von Belang ist es, dass wir wissenschaftlich zu den verschiedenen geisteswissenschaftlichen Disziplinen beitragen, und dass diese Beiträge anerkannt werden. Wir müssen auf die positiven Fälle verweisen, wenn wir den digitalen Geisteswissenschaften die gebührende Anerkennung zukommen lassen wollen. Ich habe dokumentiert, wie dies im Falle der Dialektologie einigermaßen gelungen ist,¹¹ und vorgeschlagen, dass man, um überzeugende DH-Beiträge zu finden, Projekten vier Fragen stellt, nämlich

- ob eine wichtige geisteswissenschaftliche Frage gestellt wird,
- ob die Informationsverarbeitung essentiell ist,
- ob man die Resultate kritisch überprüft (validiert) hat und

⁹ Turing Lecture, 1987. Natürlich ist dies auch ein Grund, die Frage nach der genauen Definition von Digital Humanities recht gelassen zu stellen, vgl. Vanhoutte et al. 2013.

¹⁰ Vanhoutte et al. 2013.

¹¹ Nerbonne 2005.

- ob dieses Vorgehen den Weg zu größeren Themen und Analysen freimacht.

Nicht jedes Projekt muss jeder dieser Forderungen entsprechen, auch um Raum für nicht-traditionelle Fragestellungen zu geben. Aber wir sehen vermehrt Forschungsberichte, in denen diese Fragen wenigstens teilweise bejaht werden können. Auf einige werde ich im folgenden Abschnitt kurz eingehen.

4. Zehn Thesen zur aktuellen Lage in den Digital Humanities

Erstens: Es geht bergauf mit den Digital Humanities. Die erste DHd-Konferenz in Passau (mit 320 Teilnehmern!) belegt diese These auf eindrucksvolle Weise, aber auch neue Organisationen in Italien, Japan und Australien in der [Alliance for digital Humanities Organizations](#) (ADHO) unterstreichen dies. Darüber hinaus werden auch Konferenzen von neuen Gruppen in Spanien (z.B. [Humanidades Digitales Hispánicas](#)), in den Beneluxländern (z.B. [DHBenelux](#)) und in der französischsprachigen Schweiz (z.B. [Humanistica](#)) geplant und durchgeführt.

Zweitens ist der erschwingliche Zugang zu Forschungsdaten für die verbesserte aktuelle Lage essentiell gewesen. Denn die Digital Humanities profitieren von allgemeinen Entwicklungen in der Gesellschaft, die dahin gehen, dass immer mehr in digitale Daten investiert wird. Die digitalen Geisteswissenschaften haben neue Wege eingeschlagen, und es wurde erst sukzessive der Bedarf an einer solchen Disziplin deutlich, wozu einige vorausschauende Kollegen, Institute, Bibliotheken und Firmen maßgeblich beigetragen haben. Dass so viel Aktivität erst in letzter Zeit aufgekommen ist, kurz nachdem Daten viel zugänglicher geworden sind, deutet natürlich an, dass Interesse und Potential schon vorhanden waren.

Drittens geht es um viel mehr als um den bloßen Zugang zu Daten. Echte Informationsverarbeitung ist von Anfang an unentbehrlich. Um ein Beispiel zu gebrauchen: Ohne angepasste Algorithmen kann man in alten Texten von der Zeit vor der Normalisierung der Schrift gar nicht suchen, geschweige denn Worthäufigkeitszählungen ausführen.¹² In einem rezenten Projekt in Groningen, wo Zollerhebungen am Öresund (zwischen Schweden und Dänemark) aus einer Periode von über drei Jahrhunderten digitalisiert wurden, fanden sich beispielsweise über 300 Schreibweisen für den Stadtnamen ›Kopenhagen‹. Ich werde darauf im folgenden Abschnitt über latente Strukturen noch weiter eingehen.

Viertens ist es zunehmend möglich, nicht nur relativ oberflächliche Elemente, wie z.B. Wörter (und deren Frequenz), zu analysieren, wie man dies lange Zeit bei der Autorschaftsbestimmung gemacht hat,¹³ sondern auch die *latenten Strukturen*, wofür wir uns als Geisteswissenschaftler eher interessieren. Van Halteren et al. bestimmen Autorschaft aufgrund des Wortschatzes und aufgrund syntaktischer Merkmale, die sie mittels *Tagging* zuweisen, ein vollkommen automatisches Verfahren, das jedem Wort im Text eine syntaktische Kategorie zuordnet. Sie gebrauchen die Frequenz kurzer Sequenzen von Tags, um Autoren zu erkennen, und ziehen daraus den Schluss, dass die syntaktischen Merkmale viel distinktiver sind, als sprachwissenschaftliche Theorien dies würden erwarten lassen.¹⁴ Hirst / Feiguina verwenden partielles Parsing, wobei jeder Satz

¹² Gotscharek et al. 2011, S. 159-161

¹³ Nerbonne 2007, S. xvii.

¹⁴ Van Halteren et al. 2005, S. 71.

im Text (vielleicht nur partiell) zergliedert wird.¹⁵ Die Frequenz von Kategorien (und Paaren davon), die beim Parsing zugeordnet werden, wird gebraucht, um Autoren zu erkennen. Hirst / Feiguina zeigen, dass Ihre Methode imstande ist, auch sehr ähnliche Autoren zu unterscheiden, nämlich die Brontë Schwestern, Anne und Charlotte. Wiersma / Nerbonne / Lattamus entwickeln ähnliche Techniken, um Unterschiede in der Sprache von Immigranten und Muttersprachlern zu detektieren,¹⁶ und Lučić / Blake benutzen ein anderes automatisches Verfahren, bei dem sie ihre Aufmerksamkeit insbesondere auf die Syntax in der Nähe von Nominalphrasen lenken.¹⁷

Fünftens dürfte der größte Vorteil für die digitalen Techniken dort entstehen, wo man viel Material analysieren will. Das Harvard Projekt *Culturomics*¹⁸ protzt damit, dass ihm »1,5x10⁷« Bücher zur Verfügung standen. Und auch wenn die anfängliche Begeisterung angesichts mancher Details etwas nachgelassen hat, illustriert das Projekt doch eine Art von Analyse, die den herkömmlichen Geisteswissenschaften vorenthalten ist: die Analyse auf großer Skala. Man erinnert sich gern an das Gedankenspiel, dass selbst jemand, der 80 Jahre lang 5 Bücher pro Tag läse, nur 1% des gesamten Materials kennenlernen würde. **Abbildung 1** gibt Messungen zur Frage der Resonanz einer Person wieder und sucht damit die Frage zu beantworten, wie lang auf eine Person verwiesen wird. Dass Ruhm vergänglich ist, ist sprichwörtlich und wird selbst in immer neuen Formen geäußert – so Warhols Bemerkung, dass jeder in der Zukunft 15 Minuten lang berühmt sein wird. Das Messergebnis spiegelt demnach keine neue Einsicht wider, aber es liefert zum ersten Mal eine empirisch belastbare Aussage, und dies in einer präzisen Form.

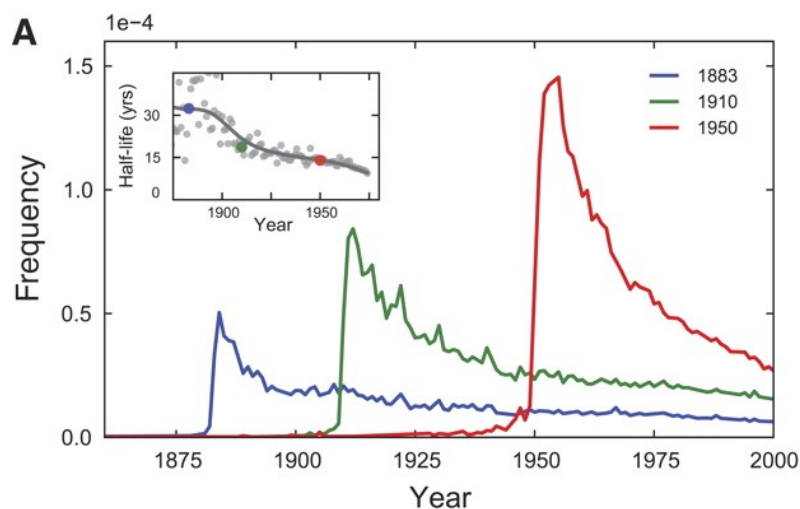


Abb. 1: »*O quam cito transit gloria mundi!*«, Thomas a Kempis quipped famously 500 years ago, but the Culturomics project not only confirmed this, but also measured how quickly fame flees, and showed that it is accelerating. The three curves show how quickly the frequency of reference to a name drops after an initial, nearly vertical climb, and the most recent curve show the steepest decline. Quelle: Michel et al. 2011.

Culturomics ist das bekannteste Projekt, in dem die Verwendung einer großen Skala die neueren Möglichkeiten zeigt, aber es gibt weitere Projekte, deren Datenbasis jenseits menschlichen Vermögens liegen, oder aber die wegen des Einsatzes digitaler Technik viel schneller ausgeführt werden können. So analysiert Jockers über 3.500 Romane aus dem 19. Jahrhundert, wobei er sich nicht auf oberflächliche Elemente konzentriert, sondern

¹⁵ Hirst / Feiguina 2007, SS. 414-415

¹⁶ Wiersma / Nerbonne / Lattamus 2011.

¹⁷ Lučić / Blake [im Erscheinen].

¹⁸ Michel et al. 2011.

auf die Themen der Romane, die er mittels maschineller Lernalgorithmen extrahiert.¹⁹ Leinonen untersucht die SVEDIA Daten, deren Datengrundlage die Tatsache bildet, dass 1.200 Sprecher des Schwedischen jeden der 20 Vokale mehrmals aussprachen.²⁰ Sie hat darauf aufbauend den gemittelten Abstand zwischen den Vokalen für alle Sprecher bestimmt und ein Verfahren entwickelt, diese mittels multidimensionaler Skalierung (MDS) in drei Dimensionen zu repräsentieren. Die so gewonnenen MDS-Koordinaten wurden durch verschiedene Farbtintensitäten abgebildet und anschließend die älteren (ab 65 Jahren) und jüngeren Sprecher (bis 27 Jahren) getrennt voneinander auf die Karte von Schweden (inklusive der Küstenregion von Finnland, wo auch Schwedisch gesprochen wird) projiziert. Die solchermaßen erzeugten Karten (Abbildung 2) geben ein eindrucksvolles Bild davon, wie die Dialektunterschiede sich ausglätten. Weil Leinonen ein automatisiertes Verfahren zur Bestimmung der Vokalqualität (der Frequenz der Vokalformanten) gebrauchte, wurde das Projekt als Doktorarbeit überhaupt erst realisierbar. Sie erhielt für diese Arbeit 2010 den königlichen Gustav-Adolph-Preis für schwedische Volkskultur.

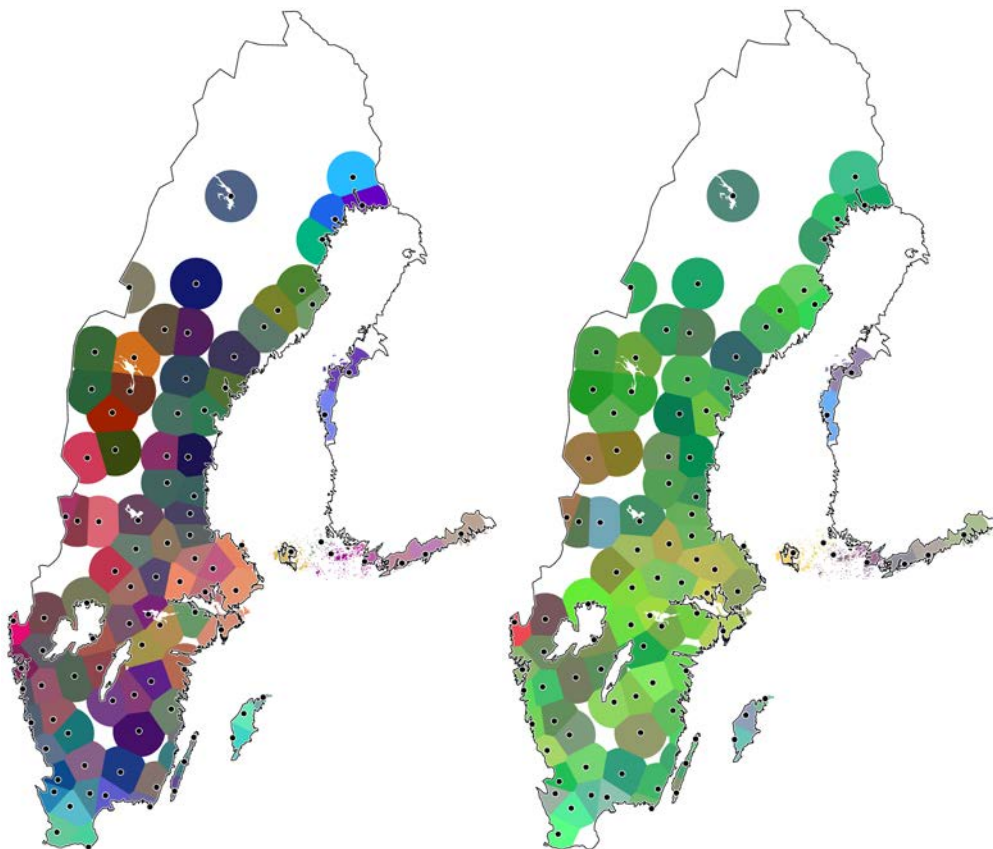


Abb. 2. Die Farben in den zwei Karten von schwedisch-sprechenden Skandinaviern spiegeln die Verschiedenheit der Aussprache wider, wobei links die Sprache von 65-jährigen steht und rechts die von 27-jährigen. Ohne automatisiertes Verfahren wäre die Fülle der Daten nicht zu analysieren gewesen. Quelle: Leinonen 2010.

Nach diesen ersten fünf Thesen will ich auch einen Blick in die Zukunft wagen. In welche Richtung werden sich die Digital Humanities bewegen und was dürfen wir in der Zukunft von ihnen erwarten?

¹⁹ Jockers 2013. Vgl. auch die Webseite [Macroanalysis](#), auf der Jockers sehr viel Material digital zur Verfügung stellt.

²⁰ Leinonen 2010.

Sechstens sage ich keinen Marsch durch (alle) (Sub-)Disziplinen voraus. Einige Disziplinen und Subdisziplinen sind sehr phänomenologisch eingestellt. Hier behält die tiefere Reflexion über die Begriffe Priorität und in solchen Fällen ist der Vorteil von digitalen Methoden minimal, im Prinzip dient er nur der Kontrolle von Konsistenz, wenn die Theorien komplexer sind. Aber selbst da kommen die Digital Humanities nur sinnvoll zum Einsatz, wenn diese Theorien formalisiert werden können. Weiterhin wird es Platz innerhalb digital angelegter Forschungsrichtungen für nicht-technisch versierte Teilnehmer geben (siehe auch unten, Punkt 9).

Siebtens haben die Geisteswissenschaften durch den Einsatz von digitalen Techniken keine »feindliche Übernahme« zu befürchten, wie der Titel der DHD-Konferenz provokant andeutet, auf die dieser Beitrag zurückgeht. Mit Forschungsfragen, etwa über die Träger von Kultur oder ob verschiedene Kulturbereiche (Alltagskultur, Hochkultur, Technologie, Mundart, Musik, Essgewohnheiten u.a.) sich entlang der gleichen Wege ausbreiten, wissen Nicht-Geisteswissenschaftler nur sehr grob etwas anzufangen. Etwas frech formuliert, behaupte ich, dass die Techniker keine Forschungsfragen innerhalb der Geisteswissenschaft formulieren können, aber auch, dass sie die Möglichkeiten der digitalen Technik innerhalb der Geisteswissenschaften unzureichend erkennen können, weil sie kein Teil dieser Forschungsgemeinschaft sind.

Achtens wird es von Vorteil sein, wenn unsere Methoden besser vergleichbar werden und wir gegenseitige Kontrolle und Kritik ausüben können. Bevor ich versuche, dies zu untermauern, will ich betonen, dass wir uns die Situation in anderen digitalen Wissenschaften vor Augen halten sollten. Der Gebrauch von digitaler Technik ist keine Panazee, nach deren Eintritt alles durchschaubar und alle Einzelstudien replizierbar werden. Die Erfahrung ist anders. Aber die »digitale Wende« bringt wohl einen Schritt in diese Richtung mit sich mit. Es ist jetzt schon so, dass digitale Geisteswissenschaftler manchmal ihre Daten austauschen und ihre Methoden am gleichen Material erproben: Van Halteren et al. stellten einen Korpus für Autorschaftserkennung zur Verfügung und andere haben ihn bereits gebraucht, um Alternativtechniken zu testen.²¹ Eder / Kestemont / Rybicki entwickelten Stylo, ein Softwarepaket für Stilometrie.²² Der wissenschaftliche Diskurs wird keinesfalls überflüssig, aber wir können viel direkter voneinander profitieren – nicht nur in Einsicht, sondern auch in Datenvorbereitung, Methoden und Analyseergebnissen.

Die Kehrseite von der Zunahme der Empirie könnte durchaus eine Minimierung der Autorität der »Weisen« bedeuten – ich spiele hierbei auf ältere, sehr respektierte Kollegen an, deren Urteilen vor allem bezüglich unserer Kultur vertraut wird. Hierüber ist schon spekuliert worden (siehe §1 oben, Diskussion zu Eisenstein 1979), und obwohl die spezifischen Beispiele sich durchaus plausibel anhören, muss man gleichzeitig zugeben, dass freilich alle Wissenschaftsgebiete ihre »Weisen« haben und auch feiern. Es wäre andererseits gut, wenn das bloße Urteil dieser »Weisen« weniger tragend oder kanonisch wirken würde, weil man sich eben gewöhnen wird, mehr nach empirischen Gründen fragen zu können.

Neuntens werden die Digital Humanities mehr Zusammenarbeit mit Informatik, Statistik und anderen Nicht-Geisteswissenschaften mit sich bringen. Dies hat natürlich schon angefangen und entwickelt sich wie die Situation in vielen anderen Wissenschaften, z.B. in der Biologie und der Astronomie. Auch dort handelt es sich übrigens um keine

²¹ Van Halteren et al. 2005, Juola und Baayen 2005

²² Eder / Kestemont / Rybicki (2013)

»feindliche Übernahme«. Der Gegenstand der Wissenschaft bleibt, während die Methoden sich opportunistisch entwickeln.

Zehntens und schließlich wird die Arbeit in den Digital Humanities innovativ, aufregend und stimulierend sein. Ich hoffe, dass mehrere Kollegen in diese Richtung schauen werden!

5. Schlussbemerkung

Den Titel – Informatik als Geisteswissenschaft – habe ich in gewissem Maße missbraucht, denn ich habe mit einer bewussten Ironie argumentiert, dass die Informatik eine Geisteswissenschaft sei, und sicher nicht versucht, diese These seriös zu erhärten. Ich habe vielmehr den Einwand erhoben, dass die Klassifizierung der modernen Wissenschaften anscheinend nicht mehr im gleichen Maße sinnstiftend ist, wie sie es zu Zeiten Diltheys war.

Mein Ziel war es dabei, zu mehr Mut bei der digitalen Wende der Geisteswissenschaften aufzufordern, denn auch die Informatik – heute sicher eine blühende Wissenschaft, sowohl theoretisch als auch angewandt – hatte Geburtswehen durchzustehen, wie vielleicht alle neuen wissenschaftlichen Richtungen.

Die ersten fünf meiner 10 Thesen zielten darauf ab, zu mehr geisteswissenschaftlicher Arbeit mittels digitaler Techniken zu ermuntern. Das Interesse an DH steigt rasant, denn die Voraussetzungen hierfür haben sich in den letzten zehn Jahren enorm verbessert. Es war mir ein Anliegen, einige Missverständnisse auszuräumen, u.a. dass die digital Humanities nur die Aufgabe hat, unsere Forschungsgegenstände digital anzubieten, und dass unsere Techniken verdammt seien, sehr oberflächlich zu bleiben. Zum Schluss habe ich ein paar erfolgreiche geisteswissenschaftliche Projekte vorgestellt, die mir wegen ihres Vermögens, viele Daten unter die Lupe zu nehmen, aufgefallen sind.

Die letzten fünf der Thesen wagten vorsichtige Vorhersagen über die Digital Humanities: Nicht jede Disziplin innerhalb der Geisteswissenschaften wird imstande und gewillt sein, von den neuen Techniken zu profitieren. Geisteswissenschaftler sind und bleiben diejenigen, die den Gegenstand (oder Gegenstände) ihres Fachs ausreichend kennen, um die Forschungsfragen zu identifizieren und die erreichten Resultate kritisch zu evaluieren. Wir werden mehr als bisher miteinander zusammenarbeiten – so die hier skizzierte Vision –, auch mit Kollegen aus anderen, nicht-geisteswissenschaftlichen Disziplinen. Abschließend bleibt mir dem nur die wichtigste Vorhersage hinzuzufügen, nämlich: dass es sich lohnen wird.

Bibliographische Angaben

Lilo Berg: Rotkäppchen 2.0 (Digital Humanities. Märchenhafte Chance oder Modetrend). In: Humboldt Kosmos 102 (2014), S. 12–23 [[online](#)].

Guide to the Software Engineering Body of Knowledge, Version 3.0. Hg. von Pierre Bourque / Richard E. Fairley. IEEE Computer Society, 2014 [[online](#)].

Wilhelm Dilthey: Einleitung in die Geisteswissenschaften. Versuch einer Grundlegung für das Studium der Gesellschaft und der Geschichte. In: Gesammelte Schriften 1. 1875–

1900. Stuttgart 1959 (zuerst 1883) [auch [online](#)].

Maciej Eder / Mike Kestemont / Jan Rybicki: Stylometry with R: A Suite of Tools. Digital Humanities 2013: Conference Abstracts. Lincoln 2013, S. 487–89 [[online](#)].

Elisabeth Eisenstein: The Printing Press as an Agent of Change. 2 Bde. Cambridge 1979.

Anders Eriksson: Swedia-projektet: dialektforskning i ett jämförande perspektiv. In: Folkmålsstudier 43 (2004), S. 11–32 [[online](#)].

Annette Gotscharek / Ulrich Reffle / Christoph Ringlstetter / Klaus Schulz / Andreas Neumann: Towards information retrieval on historical document collections: the role of matching procedures and special lexica. In: International Journal on Document Analysis and Recognition (IJ DAR) 14.2 (2011), S. 159–171.

Patrick Juola / R. Harald Baayen: A controlled-corpus experiment in authorship identification by cross-entropy. In: Literary and Linguistic Computing 20.1 (2005), S. 59–67.

Georg Wilhelm Friedrich Hegel: Phänomenologie des Geistes. Hamburg 1988 (zuerst 1806/1807).

Graeme Hirst / Olga Feiguina: Bigrams of syntactic labels for authorship discrimination of short texts. In: Literary and Linguistic Computing 22.4 (2007), S. 405–417.

John E. Hopcroft: Computer science: the emergence of a discipline (Turing Award Lecture). In: Communications of the ACM 30 (1987), S. 198–202.

Matthew L. Jockers: Macroanalysis: Digital methods and literary history. Champaign 2003.

Michiel Leezenberg / Gerard De Vries: Wetenschapsfilosofie voor geesteswetenschappen. Amsterdam 2001.

Therese N. Leinonen: An acoustic analysis of vowel pronunciation in Swedish dialects. Diss. Universität Groningen 2010.

Ana Lučić / Catherine L. Blake: A syntactic characterization of authorship style surrounding proper names. In: LLC: the Journal of Digital Scholarship in the Humanities. DOI: 10.1093/llc/fqt033 [im Erscheinen; online seit 29.06.2013].

Jean-Baptiste Michel / Yuan Kui Shen / Aviva Presser Aiden / Adrian Veres / Matthew K. Gray / The Google Books Team / Joseph P. Pickett / Dale Hoiberg / Dan Clancy / Peter Norvig / Jon Orwant / Steven Pinker / Martin A. Nowak / Erez Lieberman Aiden: Quantitative analysis of culture using millions of digitized books. In: Science 331 (2011), Nr. 6014, S. 176–182.

John Nerbonne: Elektronische Incunabeln. Antrittsvorlesung in Groningen. 1995.

John Nerbonne: Computational Contributions to the Humanities. In: Literary and linguistic computing 20.1 (2005), S. 25–40.

John Nerbonne: The exact analysis of text. Foreword. In: Inference and Disputed Authorship: The Federalist Papers. Hg. von Frederick Mosteller / David Wallace. 3. Auflage. Stanford 2007. S. XI–XX [[online](#)].

Ben Shneiderman / Catherine Plaisant: *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. 5. Auflage. Reading 2010 (zuerst 1987).

Hans Van Halteren / Harald Baayen / Fiona Tweedie / Marco Haverkort / Anneke Neijt: (2005) New machine learning methods demonstrate the existence of a human stylome. In: *Journal of Quantitative Linguistics* 12.1 (2005), S. 65–77.

Defining Digital Humanities: A Reader. Hg. von Edward Vanhoutte / Julianne Nyhan / Melissa Terras. London 2013.

Wybo Wiersma / John Nerbonne / Timo Lauttamus: Automatically extracting typical syntactic differences from corpora. In: *LLC: the Journal of Digital Scholarship in the Humanities* 26.1 (2011), S. 107–124. DOI: 10.1093/llc/fqq017.