

Fricke, Uwe

Digital Humanities und Auslandsgermanistik? Eine Erkundung der Möglichkeiten

formal überarbeitete Version der Originalveröffentlichung in:

formally revised edition of the original source in:

Vietnamesischer Deutschlehrerverband [Hrsg.]: DaF und Germanistik in Süd(ost)asien. Nachhaltige Entwicklung und Qualitätssicherung. 5. Internationale Deutschlehrertagung des Vietnamesischen Deutschlehrerverbandes. Ho-Chi-Minh-Stadt : VNU-HCM-Press 2019, S. 379-387



Bitte verwenden Sie beim Zitieren folgende URN /

Please use the following URN for citation:

urn:nbn:de:0111-pedocs-180561

<http://nbn-resolving.de/urn:nbn:de:0111-pedocs-180561>

Nutzungsbedingungen

Dieses Dokument steht unter folgender Creative Commons-Lizenz:
<http://creativecommons.org/licenses/by/4.0/deed.de> - Sie dürfen das Werk bzw. den Inhalt vervielfältigen, verbreiten und öffentlich zugänglich machen sowie Abwandlungen und Bearbeitungen des Werkes bzw. Inhaltes anfertigen, solange Sie den Namen des Autors/Rechteinhabers in der von ihm festgelegten Weise nennen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

This document is published under following Creative Commons-License:
<http://creativecommons.org/licenses/by/4.0/deed.en> - You may copy, distribute and render this document accessible, make adaptations of this work or its contents accessible to the public as long as you attribute the work in the manner specified by the author or licensor.

By using this particular document, you accept the above-stated conditions of use.



Kontakt / Contact:

peDOCS

DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Mitglied der


Leibniz-Gemeinschaft

Erschienen in: 5. Internationale Deutschlehrertagung des Vietnamesischen Deutschlehrerverbandes (2019): DaF und Germanistik in Süd(Ost)asien – Nachhaltige Entwicklung und Qualitätssicherung. Ho-Chi-Minh-Stadt, 11.-13. Oktober –VNU-HCM Press

ISBN: 978-604-73-7223-2

Digital Humanities und Auslandsgermanistik? -Eine Erkundung der Möglichkeiten

Uwe Fricke, Thammasat Universität, Bangkok

Abstract: Nach einer kurzen Einführung in den Begriff der *Digital Humanities* (zu deutsch "Digitale Geisteswissenschaften") soll eine Erkundung von Nutzungsoptionen für digitale *Tools* und Anwendungen in der Germanistik bzw. im Studium des Deutschen als Fremdsprache (DaF) erfolgen. Dabei wird fokussiert auf den sogenannten *low-end*-Bereich, technisch schon vorhandene bzw. ohne größere technische Infrastruktur und ohne langen zeitlichen Vorlauf von jedem einsetzbare *Tools* oder Techniken. Es geht dabei vorzugsweise um das in den *Digital Humanities* lange weniger beachtete Feld der Lehre und auch um von Studierenden nutzbare *Text-Tools* und vorhandene *Textkorpora*.

Schlüsselwörter: *Auslandsgermanistik, Digital Humanities, Lexik, Textanalyse, Tools*

A. Digital Humanities – eine Hinführung

Mit diesem Beitrag sollen nicht die technikaffinen Spezialisten unter den Germanisten angesprochen werden, die sich bereits im Feld der *Digital Humanities* (DH) bewegen, sondern jene, die vielleicht zu Unrecht vermuten, dass der zeitintensive Erwerb von Spezialkenntnissen und/oder besondere Computerausstattung eine unabdingbare Voraussetzung für den Zugang zum Feld darstellt. Der Aufsatz soll also neugierig machen, sich auf das einzulassen, was unter dem Label *DH* verhandelt wird. Daher werden hier einige *Tools* bzw. Anwendungen vorgestellt, die ohne technische Spezialkenntnisse in der Germanistik bzw. in der Lehre von Deutsch als Fremdsprache Verwendung finden können und vielleicht auch schon Verwendung finden, möglicherweise ohne dass NutzerInnen darin eine Verbindung zu den *DH* sehen. Damit ist schon angedeutet, dass hier eher von einer weiten Begriffsbestimmung ausgegangen wird.

Gleichwohl soll damit nicht die These vertreten werden, dass *Digital Humanities* als Einbahnstrasse in dem Sinne verstanden werden, dass allein Techniker und Informatiker (sowie ggf. auch einige Spezialisten in der Germanistik) Anwendungen entwickeln und so bereitstellen, dass technische Laien ihre *Tools* unreflektiert nutzen. Vielmehr ist es unabdingbar, dass existierende *Tools* in reflektierter Weise genutzt werden. Dazu muss man sich mit ihrer Logik, genauer der Modellierung (=Daten zu extrahieren und in eine Form zu bringen, die eine Analyse mit informationstechnischen Mitteln erst ermöglicht, Thaller 2017:16) und den Prozessen, aber nicht notwendigerweise mit jedem technischen Detail befassen. Andererseits ist es genauso wichtig, mit den disziplineigenen Fragestellungen die Entwicklung von Anwendungen anzustossen und sich an ihrer Entwicklung zu beteiligen.

Hier soll jedoch mit einer Erkundung schon existierender Anwendungen nur ein erster bescheidener Schritt ins Feld getan werden. Vorangestellt werden einige der Orientierung dienende Worte zum Begriff und der Entwicklung der *Digital Humanities*. Dass hier im Text

bisher und weiterhin zumeist mit dem Begriff des Feldes gearbeitet wurde bzw. wird, hat seinen Grund in einer Unbestimmtheit. So wird diskutiert, ob es sich bei *Digital Humanities* um eine Disziplin oder (eine) Methodologie bzw. ein *im Originaldruck S.379*

Bündel von Methoden oder nur eine Denkweise handelt.

Eine vorsichtigere und umsichtige Definition für *Digital Humanities* liefert der englischsprachige Wikipedia-Artikel zu *Digital Humanities*. Dort heisst es:

„Digital humanities (DH) is an area of **scholarly activity at the intersection of computing or digital technologies and the disciplines of the humanities**. It includes the **systematic use of digital resources** in the humanities, as well as the **reflection on their application**. DH can be defined as new ways of doing scholarship that involve collaborative, transdisciplinary, and **computationally engaged research, teaching, and publishing**.“ [Hervorhebungen, uf]

Im Feld oder Bereich der Digital Humanities ist die Definitionsfrage aber nicht geklärt bzw. das Ringen um eine Definition wird etwas selbstironisch als "Sport" (Patrick Sahle 2015: Kap. 1, Abs. 11) unter den Vertretern der *DH* charakterisiert. Andererseits scheint die Diskussion auch unter strategischen Gesichtspunkten geführt zu werden: Wenn in Selbstbeschreibungen das Feld als "*ever emerging*" (Svensson 2016:477) und "*constantly growing and changing*" (Warwick u.a. 2012, nach dem englischen Wikipedia-Artikel zu DH) charakterisiert wird, geht es wohl auch darum sich als besonders innovativ darzustellen. Doch gibt es auch einen Gründungs-Mythos als festen Bestandteil des Diskurses. Danach entstanden die *DH* schon 1949 mit dem Projekt des Jesuiten Roberto Busa, der ein Verzeichnis aller bei Thomas von Aquin zu findenden Worte (in ihrem Kontext), den sogenannten *Index Thomisticus* erstellen wollte. Um dieses mit zeitaufwendigen Routineaufgaben verbundene Unterfangen in überschaubarer Zeit erledigen zu können, wandte er sich zwecks Nutzung der Computertechnik an die *International Business Machines Corporation* (IBM) (Thaller 2017:3f.). Mehr zur Vorgeschichte der *Digital Humanities* ist im Virtuellen Museum *DH* [<http://dhmuseum.uni-trier.de/node/67>] online nachzulesen.

Im deutschsprachigen Zusammenhang wurde in der Computerlinguistik, einem Vorreiter in Sachen *DH* innerhalb der Germanistik, ab 1993 bis 1995 mit dem COSMAS-Projekt ein *Corpus Search, Management and Analysis System* entwickelt und initiiert und ist mittlerweile, vom *Instituts für Deutsche Sprache* getragen, als für jederman nach unentgeltlicher Registrierung zugängliches *Portal für die Korpusrecherche in den Textkorpora* [<https://www.ids-mannheim.de/cosmas2>] etabliert. Seit 2011 gibt es zudem für die Geistes- und Kulturwissenschaften mit Dariah (*Digital Research Infrastructure for the Arts and Humanities*) eine im Netz vertretene Initiative zur Schaffung einer digitalen Forschungsinfrastruktur - [<https://de.dariah.eu/dariah-de-in-kurze>].

Mit dem Begriff *DH* wurde zunächst auch mehr auf die Forschung fokussiert. Erst in den letzten Jahren erscheinen Publikationen, die als Einführungsliteratur für den Bereich gelten. Seitdem scheint sich der Anglizismus im deutschsprachigen Raum auch breiter, entsprechend auch im Zusammenhang der Lehre in den Geisteswissenschaften (oder gar als Studienfach)

durchzusetzen, nachdem in Bezug auf die digitalisierte zuvor schon fachübergreifend Begriffe wie *E-Learning* und *Blended Learning* weite Verbreitung gefunden hatten. Daran lässt sich auch ablesen, dass der Zusammenhang von Forschung und Lehre in den Anfängen kaum eine Rolle spielte. So entstanden auf der einen Seite Forschungsbereiche wie die Computerlinguistik, die nur unter Spezialisten diskutiert wurden und werden und auf der anderen Seite wurde unter Einbezug einer breiteren Öffentlichkeit der effizienzsteigernde Einsatz digitaler Technologien zu Lehrzwecken diskutiert und zunächst recht enthusiastisch als *E-Learning* propagiert, dann etwas bescheidener unter dem Begriff des *Blended Learning*.

Nicht zu vergessen ist hier eine in den Geisteswissenschaften oft vorzufindende reservierte Haltung gegenüber den digitalen Rechenverfahren bis hin zu ausgesprochener Kritik. Dies hat damit zu tun, dass Informationen in den Geisteswissenschaften einerseits möglicherweise nicht exakt genug sind und andererseits nicht einfach objektiv gegeben sind, sondern von einem erkennenden Subjekt erst

S. 380

hervorgebracht werden. Diese Perspektivenabhängigkeit erweise sich als Problem bei der Datenmodellierung. (Jannidis 2017:107)

Jannidis erwähnt zudem eine grundsätzliche Spannung des „systembedingten Primats des Individuellen vor dem Allgemeinen“, so dass die Modellierung, weil diese eben das Individuelle vernachlässige, in Geisteswissenschaften als „zu oberflächlich wahrgenommen“ werde. Er meint jedoch, dass gerade dort, „wo das geisteswissenschaftliche Verständnis eines Weltausschnitts und die Kompetenz zur formalen Modellierung zusammentreffen“, „im besten Fall neue Fragestellungen und neue Forschungen“ entstünden. (Jannidis 2017:107f.)

B. Erkundung von Nutzungsmöglichkeiten in der Auslandsgermanistik

Im Folgenden geht es um einige Anwendungen und insbesondere Text-*Tools* im Bereich des Übergangs von allgemeinen Kulturtechniken zum und im sogenannten *low-end*-Bereich selbst. Die Nutzung ist also möglich ohne Voraussetzungen an Vorkenntnissen sowie ohne Notwendigkeit des Erwerbs spezieller *Hard-* oder *Software* und ohne längeres Einarbeiten.

Im Rahmen dieses kurzen Aufsatzes kann nur rudimentär auf einige ausgewählte Anwendungen eingegangen werden, im Vortrag ist auch eine Demonstration möglich. Ohnehin sei dem Leser empfohlen, einige *Tools* einmal selbst auszuprobieren. Daher sind hier die Erkundungs-Links (alle zuletzt abgerufen am 31. Juli 2019) direkt im Text angegeben, während die Hintergrundtexte im Literaturverzeichnis am Ende angegeben sind. Für den Bereich Text-*Tools* gibt es zudem mit dem *Text Analysis Portal for Research* (TAPoR 3) eine zentrale Anlaufstelle im Netz. [<http://tapor-test.artsrn.ualberta.ca/home>]

Beispiel 1: Projektentwicklung mit dem Tiddlywiki [<http://tiddlyspot.com/>]

Ein Tiddlywiki kann man aus Nutzerperspektive noch zum Bereich der allgemeinen Kulturtechniken rechnen, eine fachübergreifende oder gar germanistische Qualifikation ist nicht notwendig. Es ähnelt nämlich dem guten alten Zettelkasten. Ich selbst nutze es

gelegentlich in der Vorbereitung von Seminaren oder anderen Projekten, besonders bei der Einarbeitung in neue Themen und deren Terminologie, also als eine Art Projekt-Tool.

Ein Tiddlywiki ist ein persönliches Wiki, also ein vereinfachtes, bearbeitbares *Content-Management-System*. Es dient vor allem der Organisation sowie der Mitteilung komplexer Informationen, die in Hypertextstruktur miteinander verknüpft sind. Die Anwendung ist *Javascript-* und *Cascading-StyleSheet*-basiert und nutzt nur eine einzige *Hypertext-Markup-Language*-Seite (HTML). Die Abfolge, in der sich Informationen abrufen lassen, ist nicht vorgegeben. So wie man sich im Internet oft von einer Seite zur nächsten durchklickt, so springt man im Tiddlywiki, von einer Infobox zur nächsten bzw. lässt sich auch mehrere Boxen gleichzeitig anzeigen. Es kann ohne Kenntnis der genannten *Script*-Sprachen genutzt werden. Die *HTML*-Seite kann entweder kostenlos *online* oder auf dem eigenen Computer bzw. Netzwerk abgelegt werden. Daher ist die Nutzung der Anwendung nur daran gebunden, dass ein *Web-Browser* mit den genannten gängigen *Scripts* zur Verfügung steht sowie im Falle der *Online*-Bearbeitung natürlich auch ein Internetzugang.

Anstelle von Karteikarten gibt es *Tiddler* genannte Textblöcke, welche man anlegt bzw. editiert, indem man einen Titel vergibt, dann Informationen zum Titelbegriff einträgt und in einer Unterzeile *Tags* (=Auszeichnung von Daten mit Zusatzinformationen) anlegt. Das Datum des Editierens wird automatisch mitgespeichert. Darüber hinaus kann man sehr schnell zu Wörtern innerhalb des Textfeldes automatisch einen neuen verlinkten *Tiddler* generieren, indem man innerhalb des Begriffes einen GroßBuchstaben (wie im vorangehenden Wort) verwendet oder den Begriff in eckige [[Doppelklammern]] setzt.

S.381

Neben der einfachen Verlinkung innerhalb der eigenen Seite sind auch Verlinkungen auf externe Seiten möglich. Jede Linkangabe im Textfeld wird automatisch als solche erkannt und ist ohne weitere *Scripts* auch automatisch aktiv. Daneben gibt es weitere Layout-Optionen, zu deren Nutzung man nur recht einfache Regeln kennen sollte. Ausrufezeichen vor einer Zeichenfolge variieren zum Beispiel die Schriftgröße und ermöglichen damit Überschriften, eckige Doppelklammern sind eine weitere Möglichkeit zur Verlinkung innerhalb des TiddlyWiki, vorangestzte Sternchen werden eingesetzt, um Listen zu erzeugen. Kenntnisse in der *HTML*-Programmierung können für komplexere Bearbeitungen des *Layouts* genutzt werden, sind aber keine Voraussetzung für die Nutzung der Grundfunktionen. Auch für die einfachen Regeln des *Layoutens* gilt, diese können, müssen aber nicht notwendig genutzt werden. Nach Fertigstellung eines *Tiddlers* oder mehrerer, muss die ganze Seite jeweils neu gespeichert werden. Wichtig ist, dass man jeweils speichert, bevor man die Seite verlässt. Die *Tiddler* lassen sich auf verschiedenen Wegen auffinden: Durch Eingabe eines Suchwortes, durch eine *Timeline*, durch eine alphabetische Auflistung oder durch vergebene *Tags*. Man kann auch Projekte, Aktionen oder *Reviews* anlegen und so *Tiddler* mit einer solchen Übersichtsfunktion herausheben.

Neben der Vorbereitung von Forschung und Lehre kann das Tiddlywiki auch in der Lehre Verwendung finden. So können Studierende den Arbeitsauftrag erhalten, mittels eines Tiddlywiki den Wortschatz während des Semesters selbst zu erarbeiten, das heißt Begriffsklärungen so dokumentieren, dass auch die Bezüge unter den Wörtern durch entsprechende Querverweise wie *Links* und *Tags* deutlich werden. Eine solche eigenständige Erarbeitung des Wortschatzes sollte sich auch positiv auf die Erweiterung des *mentalnen Lexikons* (persönlichen Wortschatzes) auswirken. Statt des Auswendiglernens vorgegebener Vokabellisten, haben sich Studierende hier aktiv mit den Wörtern und ihrer Stellung im Wortschatz auseinanderzusetzen. Dies ginge natürlich auch mit Karteikarten, die ebenso eine kontinuierliche Umstrukturierung und Neugruppierung der Wortkarten erlaubt. In der digitalen Form sind diese mehrfachen Korrekturen innerhalb eines *Tiddlers* aber deutlich einfacher zu editieren als auf herkömmlichen Papierkarten. Zudem können über eine Suchabfrage alle *Tiddler* mit dem entsprechenden Suchbegriff gleichzeitig aufgerufen werden. In Verbindung mit dem Digitalen Wörterbuch der Deutschen Sprache können neben der Wortbedeutung nach und nach weitere Aspekte wie Beispielsätze, Wortverlaufskurven, Kollokationen etc. auf Basis verschiedener Korpora erarbeitet und eingetragen werden. Studierende simulieren so die Arbeit eines Lexikographen. Ein gezielter Einsatz im Teilgebiet Lexikologie bietet sich an.

Im Internet gibt es eine Reihe von (leider verschwindenden) Angeboten zum Herunterladen eines Wikis oder auch zum *Hosting* einer solchen Datei, wie zum Beispiel das kostenlose *Tiddlyspot*. Der *Hoster* (siehe Link oben im Titel) erlaubt das Herunterladen und den Aufbau eines eigenen öffentlichen oder privaten Wikis ohne jegliche Installation.

Beispiel 2: Lernniveaugemessene Textauswahl mit Textanalyse-Tools

[\[https://www.psychometrica.de/lix.html\]](https://www.psychometrica.de/lix.html)

Im Internet sind eine Reihe von Textanalyse-Tools – zumeist im Zusammenhang mit der Thematik *Search Engine Optimization* (SEO) zu finden. Ihre Kernfunktion ist die Ermittlung eines Lesbarkeitsindexes. Am bekanntesten ist der Flesch-Index, der von Rudolph Flesch im Jahre 1948 zunächst für das Englische entwickelt wurde. Daneben findet man auch gelegentlich den Lesbarkeitsindex oder kurz LIX, welcher von dem schwedischen Entwickler Björnsson 1968 für das Schwedische entwickelt wurde. In beiden Fällen handelt es sich um Berechnungsformeln, die mit Satz- und Wortlängen eines Textes operieren, und daher von Rechnern prozessiert werden können. Die Formeln sind auch entsprechend für die deutsche Sprache angepasst worden.

S.382

Beide Indizes arbeiten mit Quantifizierungen von Satzlänge und Wort- oder Silbenlänge. Die Ergebnisse bewegen sich auf einer Skala zwischen 0 und 100, allerdings in gegensätzlicher Ausrichtung. Während hohe Werte nach dem Flesch-Index bedeuten, dass der Text leicht lesbar ist, zeigen hohe Werte beim LIX an, dass der Text schwierig ist. Wer sich für die genauen Formeln interessiert, die auf den ersten Blick mit ihren festen Faktorzahlen mit einer Reihe von Nachkommastellen etwas fragwürdig erscheinen, findet etwa bei Lenhard

und Lenhard (2014-2017) [<http://www.psychometrica.de/lix.html>] oder bei Wikipedia unter „Lesbarkeitsindex“ [<https://de.wikipedia.org/wiki/Lesbarkeitsindex>] kurze Darstellungen.

In einem von mir durchgeföhrten Experiment zeigte sich, dass die mittels solcher Anwendungen ermittelten Lesbarkeitsindizes weitgehend mit *ad hoc*-Einschätzungen der Textschwierigkeit durch Studierende (und Lehrende) in der Auslandsgermanistik übereinstimmen. Die Angaben der Textschwierigkeit nach der Skala des Gemeinsamen Europäischen Referenzrahmens (GER), die man zu Lerntexten auf DaF-Seiten im Internet finden, sind dagegen teilweise weniger zutreffend. (Fricke 2019)

Einen LIX-Rechner [<https://www.psychometrica.de/lix.html>] stellt das Institut für psychologische Diagnostik *online* sowie als kostenloses Programm zum Herunterladen zur Verfügung. Alternativ ist der LIX-Rechner von Supertext [<https://www.supertext.ch/tools/lix>] verwendbar. Der Flesch-Index kann bei diversen anderen *Tool*-Anbietern *online* ermittelt werden – etwa auch den im folgenden Absatz genannten Tools mit Zusatzfunktionen.

Der Lesbarkeitsindex kann Lehrenden eine gute Orientierung bei der Auswahl eines geeigneten Textes geben. Das Auffinden authentischer Texte zu einem bestimmten Thema ist angesichts der vielfach kostenlosen Angebote im Internet kaum noch ein Problem. Oft findet man mehrere Texte und so stellt sich die Frage der Auswahl. Neben den Inhalten ist dabei das Lernniveau der Studierenden zu beachten. Der Lesbarkeitsindex hilft dabei, eine lernniveauangemessene Auswahl unter Texten zu treffen. Andere *Tools* geben neben einem Lesbarkeitsindex zudem einige konkrete Schwierigkeiten im Text an, wie etwa zu lange Sätze, Nominalisierungen, Passivsätze, unpersönliche Sprache, Phrasen oder Füllwörter. Dadurch soll eine gezielte Bearbeitung im Sinne einer Textvereinfachung erleichtert werden. Beispiele hierfür sind die *Tools* von Schreiblabor [<http://www.schreiblabor.com/textanalyse/>] und Wortliga [<https://wortliga.de/textanalyse/>].

Macht man fortgeschrittene Studierende mit den Möglichkeiten vertraut, können diese selbst an Textvereinfachungen arbeiten. Sie erhalten durch die Anwendung eine Anleitung, welche Textstellen leichter lesbar formuliert werden können. Dies ist meines Erachtens eine gute Übung und Vorbereitung auf die Anfertigung von größeren schriftlichen Arbeiten, wie etwa zum Studienabschluss. Wenn Studierende mit wissenschaftlichen Texten arbeiten, neigen sie nicht selten dazu, zu eng am Wortlaut des Originals zu bleiben, aber ebenso den – gerade in deutschsprachigen akademischen Texten – komplizierten Schreibstil der rezipierten Autoren nachzuahmen, anstatt in einem ihren Fähigkeiten angemessenen eher einfachen Stil zu referieren. Der Einsatz des Wortliga-*Tools* könnte Studierenden helfen, eine Strategie zur einfacheren Formulierung von im Ausgangstext komplizierten Konstruktionen auszubilden.

Beispiel 3: Wortschatzanalyse mit dem *Language Level Evaluator*

[<https://lle.derdiedaf.com/>]

Das Schwierigkeitsniveau allein aufgrund eines auf Quantifizierung beruhenden Lesbarkeitsindexes zu ermitteln, mag trotz der oben erwähnten Resultate des Experiments, weiter kritisch betrachtet werden. So spielen sicher auch einerseits eine Passung von Leserinteressen und Thema eine Rolle. Andererseits stellen unbekannte Wörter eine Heraus-

enthalten darf, in akademischen Texten gar nur 2%, damit Fremdsprachenlerner eine gute Chance haben, einen Text ohne didaktische Begleitung noch ausreichend erfassen zu können (Schmitt u.a. 2011).

Welchen Wortschatz ein/e individuelle/r LeserIn beherrscht mag sehr unterschiedlich sein. Doch in der Arbeit mit Studierendengruppen kann man sich an den Niveaustufen des Gemeinsamen Europäischen Referenzrahmens (GER) orientieren, zumal in vielen Studiengängen der Auslandsgermanistik mit entsprechend abgestimmten Sprachlehrwerken gearbeitet wird. Hat man nun ein Wörterbuch in digitaler Form, das für jedes Wort eine entsprechende Zuordnung zu einer Niveaustufe beinhaltet, so ist es möglich, für jeden beliebigen Text, dessen Wortschatz nach den Niveaustufen zu sortieren. Für eine Studierendengruppe auf Niveau B1 darf dann nur ein kleiner Teil mit den vermutlich neuen Vokabeln der Niveaustufe B2 oder höher vorhanden sein, damit die LeserInnen den Text nicht als Überforderung empfinden. Der Klett-Verlag, welcher das Tool zur Optimierung seiner Lernmedien nutzt, stellt ein entsprechendes Tool unter dem Namen *Language Level Evaluator* im Internet nach Registrierung kostenlos zur Verfügung. Zum Hintergrund gibt es auch ein Interview im Videoformat [<https://www.youtube.com/watch?v=xdLO3teK7N0>].

„Der *Language Level Evaluator* [LLE] ermittelt:

- die Gesamtzahl der Wörter eines Textes,
- die absolute Zahl der einzigartigen Wörter (Lemmata) eines Textes,
- das Sprachniveau eines jeden Wortes (passives Leseverständnis gemäß GER, die prozentuelle Verteilung aller Vokabeln eines Textes nach Sprachlernniveau,,
- die Anzahl der Sätze eines Textes,
- die Anzahl der Wörter pro Satz und die prozentuelle Verteilung aller Sätze des Textes nach Satzlänge“ [KLETT-Verlag: <https://www.derdiedaf.com/c-191>].

Diese Angaben würden auch erlauben, einen der oben ermittelten Lesbarkeitsindizes zu ermitteln und damit einige der aufgelisteten Kenngrößen in einen Wert zu überführen, der nicht nur eine Einordnung von Texten nach Niveaustufen im Hinblick auf den Wortschatz, sondern auch im Hinblick auf Wort- und Satzstruktur erlaubt.

Der *Language Level Evaluator* ist also ein weiteres nützliches Tool, die Schwierigkeit eines Textes, hier anhand der (semantischen) Niveaustufen des GER, annäherungsweise zu bestimmen und Lehrenden bei der Textauswahl zu helfen oder den Text anzupassen. Auch für Studierende mag sich die Nutzung des *LLE* lohnen. Im Abgleich mit den eigenen Kenntnissen können sie Lücken erkennen, die auf dem formal erreichten Lernniveau im eigenen Wortschatz noch vorhanden sind. Wenn es einem Lerner wiederum auf B2 Niveau, noch deutlich an Kenntnissen des im Text vorhandenen Wortschatzes auf den Niveaustufen A1 bis B1 fehlt, ist dies sicher ein wichtiger Hinweis, im Bereich des Wortschatzes nachzuarbeiten.

Beispiel 4: Annotation mit dem TXT Werk [<https://services.neofonie.de/txt-demo>]

Eine andere Art der „Textanalyse“ bietet das Annotationstool TXT Werk von Neofonie. Auch hier kopiert man den Lesetext in ein Eingabefenster und der Text wird sekundenschnell annotiert. Das heißt, wenn man mit dem *Cursor* über das Wort fährt, wird automatisch eine Wörterklärung angeboten. Diese Form der Annotation ist unter anderem auch vermehrt in Wikipedia-Artikeln zu finden, ohne jedoch auf einem Automatismus zu basieren. Für eine Automatisierung müssen Schnittstellen zu Datenbanken angelegt sein. Das *Tool* zieht also die Angaben aus den Datenbanken und macht diese im Text auf dem Monitor sichtbar. Es erledigt also in Sekunden oder sogar weniger und für den ganzen

S. 384

Text die Aufgabe der Suche in diversen Nachschlagewerken, die den Leser sonst deutlich länger beschäftigen würden. Gleichzeitig werden am Bildschirmrand Angaben zu dem Ressort, aus dem der Text wahrscheinlich stammt, zu Schlagworten und Personen, Orten und Konzeptbegriffen aufgelistet. Diese Angaben sind meiner Erfahrung nach (noch) nicht immer überzeugend. Es macht auch den Eindruck, dass das *Tool* schon länger nicht optimiert wird, obwohl es seit einiger Zeit auch kommerziell angeboten wird. Interessant erscheint es mir aber durch die aufgezeigten und sicher noch optimierbaren Möglichkeiten.

Ein solches Tool kann die Vorbereitung des Wortschatzes durch Lehrende oder das selbstständige Erarbeiten von Texten durch Studierende entlasten. Die Zeiterparnis durch das nicht mehr notwendige Nachschlagen, das gleichwohl auch geübt werden sollte, kann in die Erstellung einer einsprachigen Vokabelsammlung bzw. eines Glossars investiert werden, um ein Systematisieren des Wortschatzes zu üben.

Beispiel 5: Text-Mining mit dem Voyant-Tool [<https://voyant-tools.org/>]

Voyant ist ein kostenloses, webbasiertes *Tool* zur Analyse von Texten bzw. Korpora, welches außer einem *Browser* und Internetanschluss nur noch die *Javascript*-Anwendung benötigt. Eigentlich ist es eine Sammlung von kleineren *Tools*, welche die Analyseergebnisse in unterschiedlicher Form visualisieren. Der zu untersuchende Text oder, wenn man mit einem kleinen Korpus arbeitet, auch eine Vielzahl von Texten kann entweder über *copy & paste* in ein Eingabefeld eingefügt oder auch aus eigenen Dateien hochgeladen werden. Die gängigen Dateiformate (rtf, doc[x], pdf, html) werden dabei akzeptiert. Darüber hinaus kann zum Laden eines Textes auch ganz einfach die *Linkadresse* (der *Uniform Resource Locator*, URL) eines *Online*-Textes eingegeben werden.

Die quantitativen Analysen von (relativen) Wortfrequenzen werden in mehreren *Panels* in unterschiedlicher Form visualisiert. Es gibt eine *Word-Cloud*, eine *Keyword in Context*-Funktion (unter Korpuslinguisten als *KWIC* abgekürzt), eine *Word Trend*-Darstellung, welche die Wortfrequenz für Textsegmente anzeigt oder auch die Frequenzen von Kollokationen unterschiedlicher Länge.

Der Erkenntnisgewinn bei der Analyse kurzer Einzeltexte ist meines Erachtens nicht besonders hoch, wenn man davon ausgeht, dass man diesen Text ohnehin selbst gründlich

lesen wird (*close reading*). Anders sieht es aus, wenn man mit einer umfangreichen Sammlung von Texten zu tun hat, die man vergleichen möchte. In diesem Zusammenhang ist der Begriff *Text-Mining* angebracht und bei großen Textsammlungen, die zu groß sind, um jeden Text gründlich zu lesen hat sich der Terminus des *distant reading* etabliert. Hier können zwei Beispiele für das methodische Arbeiten mit dem Tool angeführt werden. Das erste Beispiel wird in einem *Blog* von Robert Yeates (2013) dargestellt, der ein ganzes Textgenre, nämlich jenes der *apocalyptic fiction*, auf thematische Trends innerhalb von gut 200 Jahren hin untersucht, um eine interessante Periode für seine Analyse zu identifizieren.

Ein anderes Beispiel aus der Geschichtswissenschaft stammt von Alyssa Anderson, die mehr als 2.500 *runaway slave advertisements* des 19. Jahrhunderts in drei US-Staaten untersuchte. Als Ergebnis wird u.a. formuliert: „*The word »African« appears most frequently in Texas compared to the other states, ... This confirms our suspicions from close-reading that Texas had higher numbers of Africans than the other states.*“ Hier dient die Analyse (des *distant reading*) also der Bestätigung einer Hypothese (basierend auf *close reading*). In Bezug auf die Nutzung des Voyant-Tools heisst es außerdem: „*that digital text mining tools can not only help confirm hypotheses from close-reading but also raise previously unthought of questions for further close-reading.*“ Das Ergebnis einer Untersuchung ohne ein schon vorhandenes Erkenntnisinteresse, so soll hier betont werden, mag hingegen oft enttäuschend bzw. nichtssagend bleiben.

Damit ist auch klar, dass der Nutzen dieses *Tools* deutlich im Bereich der Forschung liegt, in der Lehre

S. 385

also wahrscheinlich nur dann zum Einsatz kommt, wenn diese – etwa im Masterstudium – forschungsorientiert angelegt ist.

Beispiel 6: *Culturomics* mit dem *N-Gram-Viewer* [<https://books.google.com/ngrams>]

Auch der *NGram-Viewer* von Google hat mit Wortfrequenzen zu tun und bietet sich zur schnellen Überprüfung von Hypothesen an. N-Gramme bezeichnen hier Wortfolgen. Ein Unigramm ist ein Wort, ein Bigramm ist eine Folge von zwei Wörtern, ein Trigramm ist eine Folge von drei Wörtern usw.

Die Analyse des *N-Gram-Viewers* basiert auf dem riesigen Korpus der von *Google* gescannten Bücher. Es muss also nicht erst ein eigener Korpus zusammengestellt werden. Da aber die Auswahlkriterien für den Korpus von *Google* nicht bekannt sind, kann nicht von einem repräsentativen Korpus ausgegangen werden. Zudem ist die automatische Texterkennung (*Optical Character Recognition*), gerade auch bei Texten mit älteren Lettern, recht fehleranfällig. So ist der *N-Gram-Viewer* wirklich nur für eine erste Analyse geeignet, die spätere genauere Analysen mit anderen Mitteln erfordern mag.

Die Analysen lassen sich für verschiedene Sprachen, unter anderem auch für Deutsch, und aktuell für den Zeitraum der gescannten Publikationen von 1600 bis 2008 durchführen. Diese Art der Analyse wird gelegentlich auch mit dem Begriff *Culturomics* bezeichnet, da sich

kulturelle Trends über den Sprachgebrauch rekonstruieren lassen. Auch dieses *Tool* ist eher zu Forschungszwecken brauchbar. Eine Orientierung zum Umgang mit dem *Tool* gibt *Google* ebenfalls im Internet [<https://books.google.com/ngrams/info>]. Die bessere Alternative dazu ist aber die Möglichkeit, das Internet umfassend und in systematischer Weise als Korpus zu verwenden. Ansätze hierzu sind für das Englische mit dem *i-Web corpus* vorhanden [<https://www.english-corpora.org/iweb/>].

C. Schluss

Damit ist die Erkundung angesichts der Seitenvorgabe für diesen Aufsatz abgeschlossen. Es könnte auf weitere Anwendungen eingegangen werden: Eine Liste mit *Onlinewörterbüchern* findet man im Lexikographieblog [<https://lexikographieblog.wordpress.com/deutsche-worterbucher-online/>], darunter auch in diachronischer Perspektive interessante Wörterbücher, beispielsweise das Digitale Wörterbuch der Deutschen Sprache (DWDS) [<https://www.dwds.de/>] oder das Wörterbuchnetz [<http://www.woerterbuchnetz.de>], welches 29 Wörterbücher (historisch und regional) und Nachschlagewerke in die Suche integriert.

Tools für die Vorbereitung von Aufgaben für Sprachlerner finden sich auf der Seite *Compleat Lexical TUTOR* von Tom Cobb [<https://www.lextutor.ca>]. Die vorrangig auf Englisch ausgelegten Tools, lassen sich teilweise auch für Deutsch nutzen oder adaptieren.

Zudem wächst die Zahl an *online* kostenlos verfügbaren Korpora, die kostenlos nutzbar sind. Diese Angebote können sicher auch selbstständig erkundet werden. In das „Lehren und Lernen mit Korpora im DaF-Unterricht“ mag Franziska Wallners Beitrag (2014) einführen.

Literatur *[Abruf aller angegebenen URLs letztmals am 31. Juli 2019]

Anderson, Alyssa (2014): Using Voyant for Text Analysis. In: *Digital History Methods*. Rice University. URL*: <http://ricedh.github.io/02-voyant.html>

Fricke, Uwe (2019, im Erscheinen): Internettools als Hilfe bei der lernniveauangemessenen Textauswahl in der Auslandsgermanistik? In: *Dokumentation der Internationalen Konferenz des Indonesischen Germanistenverbandes vom 28.-29. Sept. 2018 in Yogjakarta, Indonesien*

Jannidis, Fotis; Kohle , Hubertus; Rehbein, Malte (2017): Digital Humanities: eine Einführung. Darin: *Teil 4 mit Unterkapiteln von verschiedenen Autoren zu „Digitale Methoden“* S.386

Jannidis, Fotis (2017): Grundlagen der Datenmodellierung. In: *Jannidis, Fotis; Kohle , Hubertus, Rehbein, Malte: Digital Humanities: eine Einführung. –Stuttgart* S.99-108

Lenhard, Wolfgang; Lenhard, Alexandra (2014-2017): Berechnung des Lesbarkeitsindex LIX nach Björnson. URL*: <http://www.psychometrica.de/lix.html>

Sahle, Patrick (2015): Digital Humanities? Gibt's doch gar nicht! In: *Grenzen und Möglichkeiten der Digital Humanities*. Hrsg. C. Baum und Th. Stäcker. (= Sonderband der Zeitschrift für digitale Geisteswissenschaften) URL*: http://www.zfdg.de/sb001_004

Schmitt, Norbert; Jiang, Xiangying; Grabe, William (2011): The Percentage of Words Known in a Text and Reading Comprehension. In: *The Modern Language Journal* 95, 1, S.26-43
URL*: https://www.lextutor.ca/cover/papers/schmitt_et.al_2011.pdf

Svensson, Patrik (2016): Sorting out the Digital Humanities. In: Schreibman, Susan; Siemens, Ray; Unsworth, John: *A New Companion to Digital Humanities*. –Oxford S. 476-492

Thaller, Manfred (2017): Geschichte der Digital Humanities. In: Jannidis, Fotis; Kohle, Hubertus; Rehbein, Malte: *Digital Humanities: eine Einführung*. –Stuttgart S.3-12

Wallner, Franziska (2014): Lehren und Lernen mit Korpora im DaF-Unterricht. In: Goethe-Institut: *Magazin Sprache*. URL: <https://www.goethe.de/de/spr/mag/20454877.html>

Yeates, Robert (2013): Project Voyant Tools. In: *Post-Apocalyptic Cities. A research blog by Robert Yeates*. URL*: <https://postapocalypticcities.wordpress.com/2013/05/02/voyanttools>