

Tagungsbericht für den DHd-Blog

11. März 2016: Vorträge_5b: Digitale Edition III

Melanie Mihm, Justus-Liebig-Universität Gießen

Als Masterstudentin der Computerlinguistik und Texttechnologie an der Justus-Liebig-Universität Gießen habe ich an der diesjährigen DHd in Leipzig (<http://dhd2016.de/>) teilgenommen, die zum Thema „Modellierung – Vernetzung – Visualisierung: Die Digital Humanities als fächerübergreifendes Forschungsparadigma“ vom 7. bis 12. März 2016 über 460 Teilnehmer anzog. In diesem Blogposting möchte ich von meinen Eindrücken im Hinblick auf die Vortragsreihe „**Vorträge_5b: Digitale Edition III**“ erzählen, in der drei Beiträge aus den Bereichen der Korpuslinguistik, der Digitalen Editorik und der Geschichtswissenschaften vorgestellt wurden. Der Fokus lag bei allen drei Vorträgen auf den Möglichkeiten der Implementierung von literarischen wie historischen Textkorpora und digitalen Editionen in Infrastrukturen. Zunächst wurde der jeweilige Gegenstand und die entsprechende Infrastruktur skizziert und anschließend ein Anwendungsfall samt Live-Demonstration simuliert, bei dem die benutzten Tools und die gezeigten Arbeitsumgebungen unter die Lupe genommen wurden. Alle drei Vorträge stellten interessante und hilfreiche Tools vor, die mitunter auch innovative, explorative Zugänge zu Textkorpora ermöglichen. In den folgende drei Absätzen gehe ich kurz auf jeden einzelnen Vortrag ein und hebe solche Aspekte hervor, die für mich persönlich ein Zugewinn waren.

Dirk Goldhahn (Universität Leipzig) ging in seinem Vortrag „**Operationalisierung von Forschungsfragen in CLARIN-D - Der Anwendungsfall Ernst Jünger**“ der übergeordneten Fragestellung nach, inwieweit linguistische Ressourcen genutzt werden können, um Forschungsfragen der Digital Humanities zu beantworten. In Zusammenarbeit mit Thomas Gloning (Universität Gießen) und unter der Verwendung der Infrastruktur CLARIN-D (<http://www.clarin-d.de/de/>) wurde ein Anwendungsfall aus der germanistischen Praxis entwickelt, der aufzeigte, wie ein systematisches Vorgehen und die Ermittlung der Daten, der Referenzdaten und der Algorithmen sowie die Visualisierung der Ergebnisse in einem Workflow innerhalb einer Forschungsinfrastruktur kombiniert werden können. Die dazugehörige Entwicklungsarbeit der Projektinhaber Dirk Goldhahn, Thomas Eckart und Gerhard Heyer (alle Universität Leipzig) stellt die Webanwendung *Corpus Diff* (<http://corpusdiff.informatik.uni-leipzig.de/corpora/>) dar. Ein generisches Tool, das zur Differenzanalyse und Visualisierung der Ergebnisse dient. Interessant bei diesem methodischen Vorgehen einer Differenzanalyse ist die Möglichkeit, die Veränderungen des Vokabulargebrauchs in Texten zu ermitteln und somit die Unterschiede und Ähnlichkeiten zum Referenzkorpus zu quantifizieren. Anschaulich wurde das Ausmachen von Referenzdaten mithilfe des sogenannten CLARIN *Virtual Language Observatory* (VLO) (<https://vlo.clarin.eu>), eine spezielle Suchmaschine für sprachbasierte Ressourcen, vorgestellt. Für Sprach- und Literaturwissenschaftler_innen ist die Infrastruktur CLARIN-D nützlich, vor allem aufgrund von

WebLicht (http://weblicht.sfs.uni-tuebingen.de/weblichtwiki/index.php/Main_Page), welches das Ansteuern von webbasierten Prozessketten ermöglicht und somit auch die Aufbereitung des Rohmaterials. Dirk Goldhahn demonstrierte, wie die Differenzanalyse für die Ernst Jünger-Daten angestoßen wird und als Ergebnis eine Matrix der paarweisen Korpusähnlichkeiten mit verschiedenen Farbschemata angezeigt wird. Sollte das Interesse einem bestimmten Wort gelten, kann sich der Benutzer über eine Detailansicht auch individuelle Wortlisten der Korpora anzeigen lassen. Neben der Ähnlichkeitsmatrix können je nach Bedarf ein Dendrogramm und ein Leitliniendiagramm angezeigt werden. Mit der Webanwendung *Corpus Diff* (<http://corpusdiff.informatik.uni-leipzig.de/corpora/>) konnte erfolgreich gezeigt werden, dass ein komplexer Workflow mit unterschiedlichen Einzelschritten und der Hinzunahme weiterer Tools benutzerfreundlich kombinierbar und realisierbar ist. Meiner Meinung nach punktet die Webanwendung vor allem aufgrund der Benutzerfreundlichkeit und den drei Visualisierungsformen: Matrix, Dendrogramm und Liniendiagramm. Im Hinblick auf die Nutzung von Wortlisten, die erstens leicht zu verarbeiten sind und zweitens den leichteren Zugriff auf zeitgenössische Literatur des 20. Jahrhunderts mit besonderen Lizenzbedingungen ermöglichen, unterstreicht dieser Ansatz selbst seine Hervorhebung.

Susanne Schütz und Marcus Pöckelmann (beide Martin-Luther-Universität Halle-Wittenberg) stellten im Vortrag „**LERA - Explorative Analyse komplexer Textvarianten in Editionsphilologie und Diskursanalyse**“ ihr Tool LERA (<http://lera.uzi.uni-halle.de/>) vor. LERA steht für „Locate, Explore, Retrace and Apprehend complex text variants“ und ist eine interaktive, webbasierte Arbeitsumgebung für die Kollationierung verschiedener Textfassungen. Die jüngsten Innovationen innerhalb dieser Arbeitsumgebung sind drei Visualisierungstools für digitale Editionen: (1) Eine Suchfunktion samt (2) Übersichtsleiste für das Auffinden von Schlagworten namens *CATview* (<http://catview.uzi.uni-halle.de/>), die in Form von Rechtecken eine strukturierte Übersicht über den gesamten Text oder Textsegmente gibt und das Navigieren innerhalb der Einheiten durch eine Verlinkungsstruktur ermöglicht. (3) Interaktive Wortwolken können individuell generiert werden. Susanne Schütz machte stark, dass die Benutzeroberfläche individuell anpassbar sei, und dies im Besonderen für kulturwissenschaftliche Herangehensweisen und Forschungsfragen hilfreich sei, wenn man beispielsweise orthografische und drucktechnische Informationen ausblenden lassen möchte und stattdessen die inhaltlichen Veränderungen auf der Wortebene angezeigt bekommen möchte. Marcus Pöckelmann stellte den innovativen Ansatz vor, die drei Visualisierungstools miteinander zu kombinieren, der meines Erachtens das Erkunden umfangreicherer Texte in digitalen Editionen erheblich erleichtert. Eine Besonderheit von LERA ist die Integration von weiteren Sprachen in die Arbeitsumgebung, was unter anderem für Germanisten_innen und Hispanisten_innen von Interesse sein kann.

Bryan Jurish (Berlin-Brandenburg Akademie der Wissenschaften) und Thomas Werneke (Zentrum für Zeithistorische Forschung) schlossen mit „**DiaCollo: diachronen Kollokationen auf der Spur**“ die

Vortragseinheit ab. Jurish erläuterte, dass *DiaCollo* (<http://de.clarin.eu/de/kollokationsanalyse-in-diachroner-perspektive>) ein Softwarewerkzeug ist, mit dem typische Wortverbindungen auf Grundlage verschiedener Textkorpora in einem bestimmten Zeitraum gefunden und so Bedeutungsverschiebungen ausgemacht werden können. *DiaCollo* wurde innerhalb der Infrastruktur von CLARIN-D entwickelt, um Historikern zu helfen, Änderungen in den Diskursen der Quellen aufzufinden, indem effizient die Kollokationen extrahiert und zum diachronen Vergleich und zur Analyse ausgegeben werden. Für die Aufbereitung und Ausgabe der Ergebnisse bietet das Tool Zeitreihen-, Tag-Cloud- und Bubble-Visualisierungen an. Die technischen Hintergründe des Werkzeuges und die Implementierung sind als modulare Perl Bibliothek aufgebaut. Laut Thomas Werneke ist der Hintergrund des Projektes, die historische Dimension von auf den ersten Blick synchron erscheinenden Korpora wieder sichtbar zu machen. Das interessante an *DiaCollo* ist dessen prinzipielle Offenheit für unterschiedliche Wissenschaftsbereiche. Im Feld der Geschichtswissenschaften hob Werneke unter anderem den Nutzen des Tools für die Auswertung einer Kulturgeschichte eines politischen und sozialen Diskurses einer bestimmten Epoche hervor. Ferner kann das Werkzeug umfangreiche Quellenkorpora stemmen, was besonders für das Close Reading sehr nützlich sein kann. In einem Screencast simulierten Jurish und Werneke eine Suchanfrage für „Krise“ (<http://kaskade.dwds.de/dstar/zeit/diacollo/?query=Krise&format=cloud>) in den Korpora *Die Zeit* und *DDR-Presse*. Die Ergebnisse werden in einer Wortwolke ausgegeben, die auf einer Zeitleiste abgespielt wird und mit einer Farbskalierung unterschiedlich starke oder schwache Kollokationen markiert. Zusammenfassend wurde die flexible Einstellung der Zeitabschnitte gelobt, um den semantischen Wandel in den Korpora anzeigen zu lassen – ein Prozess, der mit relativ wenig Rechenleistung von statten geht und ein Vorteil dieses Tools darstellt. Werneke beendete den Vortrag mit dem Appell, dass das „Tool *DiaCollo* nicht das Close Reading ablösen soll, sondern eine Verschränkung von Distant Verfahren mit dem traditionellen hermeneutischen Zugang der Historiker an Quellen ermöglicht“. Abschließende Worte, mit denen ich diesen Blogeintrag schließen möchte – auch für die beiden anderen Tools gilt dasselbe. Alles in allem konnten alle drei Vorträge mit ihren vorgestellten Tools punkten und eröffneten einen interessanten Blick auf die aktuellen Entwicklungen in der Digitalen Editorik. Man darf gespannt sein, wie die Entwicklung innerhalb eines Jahres bis zur nächsten DHd 2017 in Bern (<http://www.dig-hum.de/dhd-2017>) voranschreitet und weitere nützliche Tools und Optimierungen zu Tage fördert.