

Korpushermeneutik - Ansatz und Werkzeug zur Analyse großer Textkorpora

,
jan.ruediger@uni-kassel.de
Universität Kassel, Deutschland

Der Vortrag fußt auf drei Säulen: Theorie, Forschungspraxis und Hochschullehre. Sie werden im Vortrag einzeln ausgeführt, dann kombiniert.

Theorie: *Korpuslinguistik* mit *Hermeneutik* zu verbinden, ist keine grundsätzlich neue Idee. Die bisherigen Vorschläge (z. B. Haß 2007; Teubert 2006) führen aber in ihrer Konsequenz zu einer einseitig gelagerten Korpuslinguistik, die entweder *corpus-driven* oder *corpus-based* orientiert ist.

Bei Haß (2007) werden wichtige Grundüberlegungen der *Korpus-Hermeneutik* diskutiert. Im Abschnitt Haß (2007: 248-258) erfolgt eine Beispielanalyse, deren Methoden fast ausschließlich dem *corpus-driven* Spektrum zuzuordnen sind. Ermittelte statistische Werte werden zwar interpretiert, jedoch führt dies nicht zu weiteren Forschungskonsequenzen. Gerade aber in der zyklischen Interpretation liegt die Stärke der Korpushermeneutik.

Bei Teubert (2006) ist der Blick auf den Sichtbereich des *corpus-based* Methodenapparats beschränkt. Korpusmaterial dient in dieser Arbeit als eine Art Steinbruch, in dem man nach Belegen schürft. *Text-Mining* ist zwar ein Aspekt der Korpushermeneutik – es darf aber nie das alleinige Merkmal sein.

Daher plädiere ich für grundlegend *neue und praktikable Korpushermeneutik*, die sowohl klassische als auch computergestützte Analyseverfahren vereint. Einen zentralen Punkt nimmt dabei die (Weiter-)Entwicklung des bestehenden Wissens ein. Annahmen, Beobachtungen und Ergebnisse werden zu Wissensmodellen korreliert und durch einen zyklisch organisierten Analyseprozess falsifiziert. Zum jetzigen Zeitpunkt ergeben sich drei grundlegende Forderungen an eine Analyse, wenn Sie unter dem Begriff *Korpushermeneutik* firmieren soll:

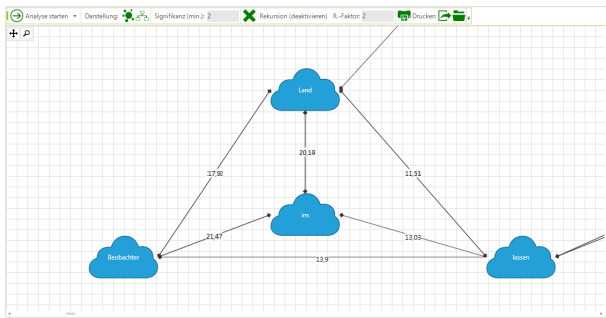
- **Die Analyse muss mehrere, abwechselnde und aufeinander aufbauende Zyklen durchlaufen.**
Diese Idee lehnt sich an der bereits von Felder (2016: 124) aufgestellten Beobachtung an: „Für die praktische Textanalyse sind beide Herangehensweisen erkenntnistiftend – insbesondere ein stetiger Wechsel zwischen der relativ induktiven (*corpus driven*) und der eher deduktiven (*corpus-based* oder *corpus-assisted*) Vorgehensweise ist vielversprechend.“
- **Die Entwicklung von Wissen geschieht durch Falsifikation** - wie sie u. a. durch Popper (2005) und Albert (1969) gefordert wird. Daher sind Fragen,

Thesen und Methoden so zu wählen, dass Vorwissen geprüft und hinterfragt wird.

- **Der Analyseprozess ist transparent zu gestalten - transparent in Durchführung und Fehlerbetrachtung – Die Handlungsmöglichkeiten sind aufzuzeigen.** Jeder Durchführungszyklus geschieht unter der Prämisse der Falsifikation. Jede Methode ist ebenso kritisch zu hinterfragen wie die daraus resultierenden Ergebnisse. Durch einen Zyklus werden neue Handlungsmöglichkeiten offengelegt, die es zu überprüfen gilt. Daher muss transparent gemacht werden, welche Ansätze mit welchen Mitteln verfolgt werden und welche Fragen am Schluss offenbleiben.

Forschungspraxis: Gerade in den letzten fünf bis zehn Jahren ist die Möglichkeit stark gewachsen, große (linguistische) Datenmengen zu erheben und auszuwerten. Text-/Sprachdaten können fast ohne Limitierung für die Forschung erhoben werden. Die darauf aufbauenden Datenmodelle erreichen eine immer höhere Komplexität. Daher bedarf es neuer Methoden, diese zu strukturieren und teilweise auch zu reduzieren (z. B. durch Algorithmen oder Visualisierungen), damit sie (er-)fassbar für den Anwender werden. Ein Problem bei der Umsetzung der korpushermeneutischen Theorie ist die bisher existierende Softwarelandschaft der Computer-/Korpuslinguistik. Viele Programme sind notwendig, um aus einem einfachen Rohtext ein visuelles Ergebnis zu erzeugen. Die Programme sind teilweise untereinander inkompatibel - oder sie folgen ausschließlich einem der beiden Paradigmen. Im Vortrag wird ein von mir entwickeltes Programm vorgestellt, das diese Arbeit übernimmt und korpushermeneutische Analysen ermöglicht. Der *CorpusExplorer* ist kostenfrei verfügbar und übernimmt alle nötigen Arbeitsprozesse – angefangen bei der Textaufbereitung, Trennung von Text und Metadaten, Annotation, bis hin zur Auswertung und Visualisierung (über 30 unterschiedliche Auswertungsmodule). Alles mit einem Tool, mit nur wenigen Mausklicks und vereint unter einer intuitiven Benutzeroberfläche. Der *CorpusExplorer* erlaubt sowohl *corpus-driven* als auch *corpus-based* Analysen und durch die zyklische Verschränkung der Werkzeuge die angestrebte korpushermeneutische Analyse. Im Vortrag wird auf konkrete Praxisbeispiele eingegangen und gezeigt, wie sich eine korpushermeneutische Analyse entwickelt. Ein exklusiver Vortragspunkt wird sein, dass neben dem Programm das *CorpusExplorer*-Framework erstmalig vorgestellt wird. Mit diesem werden zwei Dinge möglich. Zum einen kann der *CorpusExplorer* mit eigenen Funktionen erweitert werden (z. B. schreiben / anbinden neuer Tagger / Dateiformate, entwickeln eigener Analysemodule, uvm.). Zum anderen kann man den *CorpusExplorer* in eigene Programme integrieren. Ein Teil des Quellcodes (Import- / Export-Funktion), sowie Quellcode von An-Projekten wurde bereits veröffentlicht. Der Quellcode des Frameworks wird nach Abschluss des Promotionsprojekts freigegeben.

N-Gramm-Graph
Verknüpfung von N-Grammen auf Basis von POS-Tags
Graph: Grün: N-Gramm-Kopf, Blau: N-Gramm-
Zwischenteil, Rot: N-Gramm-Ende



41–60 http://www.linguistik-online.de/28_06/teubert.html
[letzter Zugriff 09. Februar 2016].

Kookkurrenzgraph (Ausschnitt)

Das Beispiel zeigt einen per Rekursion ermittelten
Teilausschnitt, der auf die Phrase: „Beobachter / ins /
Land / lassen“ rekurriert.

Fußnoten

1. z. B. kann die Ausgabe des einen Programms nicht vollumfänglich von einem anderen eingelesen werden.
2. Gemeint sind hier die Paradigmen corpus-driven oder corpus-based.
3. Aktuell verfügbar: TreeTagger, TnT, Stanford-Tagger oder gar *Keine Annotation*.

Bibliographie

Albertr, Hans (1969): *Traktat über kritische Vernunft*. Tübingen: J.C.B. Mohr (Paul Siebeck).

Bubenhof, Noah (2009): *Sprachgebrauchsmuster*. Korpuslinguistik als Methode der Diskurs- und Kulturanalyse. Berlin: de Gruyter.

Dang-Anh, Mark / Rüdiger, Jan Oliver (2015): “From Frequency to Sequence: How Quantitative Methods can Inform Qualitative Analysis of Digital Media Discourse”, in: *10plus1* 1: 57–73.

Gardt, Andreas (2007): “Linguistisches Interpretieren: Konstruktivistische Theorie und realistische Praxis”, in: Hermanns, Fritz / Holly, Werner (eds.): *Linguistische Hermeneutik*. Theorie und Praxis des Verstehens und Interpretierens. Tübingen: Niemeyer 263–280.

Haß, Ulrike (2007): “Korpus-Hermeneutik: zur hermeneutischen Methodik in der lexikalischen Semantik”, in: Hermanns, Fritz / Holly, Werner (eds.): *Linguistische Hermeneutik*. Theorie und Praxis des Verstehens und Interpretierens. Tübingen: Niemeyer 241–261.

Popper, Karl R. (2005): *Gesammelte Werke*. 3: Logik der Forschung Tübingen: Mohr Siebeck.

Runkler, Thomas (2010): *Data Mining*. Methoden und Algorithmen intelligenter Datenanalyse. Wiesbaden: Vieweg+Teubner.

Teubert, Wolfgang (2006): “Korpuslinguistik, Hermeneutik und die soziale Konstruktion der Wirklichkeit”, in: *Linguistik Online* 28, 3: