

Tools zur elektronischen Text- und Sprachanalyse

Vorbemerkungen

Im Bereich der sog. Content Analysis, die im Kontext sozialwissenschaftlicher Fragestellungen schon seit längerem einen festen Platz behaupten kann, begegnet man Verfahren sowohl der quantitativen als auch der qualitativen Textanalyse. Spezialisierte Computerprogramme haben die Durchführung dieser Analysen weiter automatisiert und damit beträchtlich vereinfacht. Hinzu treten – insbesondere von (computer-)linguistischer Seite – Methoden der Prozessierung sprachlicher Korpora über Annotationssysteme, Satzgrammatiken und verschiedene Techniken der Informationsextraktion. Morphologisch-lexikalische und phonetische Analysesysteme natürlicher Sprachen schließlich runden das Bild „nach unten“ hin ab.

Welche Tools nun sinnvoll in Literaturwissenschaft und Linguistik eingesetzt werden können, entscheidet sich an Untersuchungsziel und Methodik der jeweiligen Forschungsarbeit. Generell werden wohl zunächst eher Fragen der Kontextanalyse (über Konkordanzen und Folgerelationen) und der Segmentierung von Text- bzw. Worteinheiten im Vordergrund des Interesses stehen. Davon abgesehen lohnt sich zudem gewiss ein Blick auf verschiedene Möglichkeiten der elektronischen Beschreibung des Wortschatzes, da auch hier Verwendungskontexte (semantische und pragmatische Relationen) eine wichtige Rolle spielen (sehr aufschlussreich ist in diesem Zusammenhang übrigens das „Wortschatz-Lexikon“ der Uni Leipzig).

Adressen zur Sichtung und Beschreibung textanalytischer Tools

Content Analysis Resources

URL: <http://www.car.ua.edu/>

Content-analysis.de

URL: <http://www.content-analysis.de/>

CTI Textual Studies: Guide to Digital Resources

URL: <http://users.ox.ac.uk/~ctitext2/resguide/resources/index.html>

Inhaltsanalyse.de

URL: <http://www.inhaltsanalyse.de/>

SIL Software Catalog

URL: <http://www.sil.org/computing/catalog/index.asp>

Text Analysis Info

URL: <http://www.textanalysis.info/>

The Content Analysis Guidebook Online

URL: <http://academic.csuohio.edu/kneuendorf/content/>

The Linguist List – Web Resource Listings

URL: <http://linguistlist.org/sp/Software.html>

Kurze Vorstellung einiger relevanter Programme

- TextSTAT (ein einfaches, aber effizientes Tool zur Indizierung und Konkordanzanalyse; freie Software)

URL: <http://www.niederlandistik.fu-berlin.de/textstat/software-en.html>

- Simple Concordance Program (Programm zur Kontextanalyse von Schlüsselwörtern, Erstellung von Frequenzprofilen mit einigen textstatistischen Extras, zahlreiche eingebaute Alphabete, Alphabeteditor; Freeware)
URL: <http://www.textworld.com/scp/>
- AntConc (Freeware-Programm mit stetig gewachsenem Repertoire textstatistischer Analysemöglichkeiten – Erstellung von Wortlisten, Konkordanzen, Kollokationen, Clustern, Plotverteilungen, Schlüsselwörtern –, im Umfang durchaus vergleichbar mit den WordSmith Tools, sehr flüssig bedienbar)
URL: <http://www.antlab.sci.waseda.ac.jp/software.html>
- AntMover (in Perl geschriebenes Freeware-Programm zur automatisierten Analyse von Textstrukturen, noch ziemlich am Anfang der Entwicklung stehend)
URL: <http://www.antlab.sci.waseda.ac.jp/software.html>
- Oxford WordSmith Tools (zahlreiche textstatistische Features; eingeschränkte Demoversion frei herunterladbar, für den tatsächlichen Einsatz kostenpflichtig)
URL: <http://www.lexically.net/wordsmith/index.html>
- CoAn (umfängliches Tool zur Content Analysis; freie Demoversion, sonst kostenpflichtig)
URL: <http://www.coan.de/>
- TextQuest (textanalytisches Tool mit vielen Funktionen, Demoversion herunterladbar)
URL: <http://www.textquest.de/>
- Concordance (Programm zur Erstellung interaktiver (Web-)Konkordanzen; zeitlich eingeschränkte Demoversion herunterladbar)
URL: <http://www.concordancesoftware.co.uk/>
- EventParser / EpiTest (narratologisches Tool-Gespann zur aktionslogischen Analyse „literarischer“ Texte; unter Copyleft stehend, also frei verfügbar)
URL: <http://www.jcmeister.de/html/epitest.html>
- TempusMarker (Tool, das dem Markup von Temporalitätseffekten in narrativen Kontexten dienen soll; Copyleft)
URL: <http://www.jcmeister.de/html/tempusmarker1.html>
- Kura (linguistisches Datenbankprogramm zur Beschreibung natürlicher Sprachen; freie Software)
URL: <http://www.ats.lmu.de/kura/index.php>
- Toolbox (frei verfügbares ethnologisch-linguistisches Tool zur Segmentierung und Klassifizierung sprachlicher Einheiten, sehr geeignet für morphologische Untersuchungen und die Erstellung von Lexika; die Benutzung will erlernt sein)
URL: <http://www.sil.org/computing/toolbox/>
- WordSurv (Hilfsprogramm zur Bestimmung linguistischer Verwandtschaften über den Vergleich von Wortlisten; unter SIL-Freeware-Lizenz stehend)
URL: http://www.sil.org/computing/catalog/show_software.asp?id=67
- Lexique Pro (Programm zur Editierung, Darstellung und Weitergabe lexikalischer Daten; frei verfügbar unter SIL-Freeware-Lizenz)

URL: <http://www.lexiquepro.com/>

- Linguistic Tree Constructor (der Name ist Programm; freie Software unter der GPL)
URL: <http://ltc.sourceforge.net/>
- RSTTool (ein graphisches Interface zur Auszeichnung von Texten im Sinne der Rhetorical Structure Theory; Freeware)
URL: <http://www.wagsoft.com/RSTTool/index.html>
- Systemic Coder (Tool zur Erleichterung der linguistischen Kodierung von Korpusmaterial; Freeware)
URL: <http://www.wagsoft.com/Coder/index.html>
- UAM Corpus Tool (neu entwickelte Programm-Umgebung zur Annotation von Textkorpora mit Hilfe selbst entworfener Annotationsschemata, Möglichkeit der Recherche über mehrere Ebenen, komparative Statistiken, Speicherung der Annotationen in XML-Dokumenten; Freeware)
URL: <http://www.wagsoft.com/CorpusTool/index.html>
- Unitex (Korpus-Prozessor zur Analyse von Texten unter Nutzung linguistischer Ressourcen, Erstellung von Indizes, Konkordanzen, Satzbauplänen, etc.; freie Software unter GPL-Lizenz)
URL: <http://www-igm.univ-mlv.fr/~unitex/>
- CLaRK System (XML-basiertes System zur Entwicklung und Untersuchung linguistischer Korpora; freie Nutzung im akademischen Rahmen)
URL: <http://www.bultreebank.org/clark/index.html>
- GATE (eines der führenden Programme aus dem Bereich des Natural Language Engineering; freie Software unter GPL-Lizenz)
URL: <http://gate.ac.uk/>
- Ellogon (eine weitere wichtige Arbeitsumgebung im Umkreis des Language Engineering; freie Software, LGPL-Lizenz)
URL: <http://www.ellogon.org/>
- EXMARaLDA (frei verfügbares System zur Transkription und Annotation gesprochener Sprache, Verwendung XML-basierter Dateiformate)
URL: <http://www1.uni-hamburg.de/exmaralda/>
- PRAAT (Programm zur phonetischen Analyse gesprochener Sprache; freie Software unter GPL-Lizenz)
URL: <http://www.fon.hum.uva.nl/praat/>