

Digitale Tagung
30. September bis 02. Oktober 2020

Sprache und Wissen hin und zurück – iterative Annotation als linguistische Forschungsmethode

Schon die analoge Kulturtechnik des Annotierens mit ihren antiken Wurzeln (Glossen und Scholien z. B.) stellt eine sprachensible Praktik dar, durch die das Zusammenwirken von verschriftlichter Sprache und kontextualisierend-ergänzendem Wissen explizit und sichtbar gemacht wird. Annotationen dienen einerseits der Anreicherung von Dokumenten mit mehr oder weniger kontextspezifisch relevantem Wissen. Andererseits zielen Annotationen aber auch auf die Erschließung von Texten, das Erkennen von Zusammenhängen und Mustern, die Analyse und Interpretation – also das Gewinnen von Erkenntnissen aus sprachlichen Daten. Digitale Annotation stellt als wissenschaftliche Methode keine rein mimetisch ins Digitale überführte Form einer analog geprägten Praktik dar, sondern hat sich unter den spezifischen Bedingungen der Digitalität (weiter)entwickelt und ausdifferenziert. Insbesondere die Aspekte der Automatisierbarkeit und der algorithmischen Weiterverarbeitung bzw. Analyse eröffnen neue Perspektiven – gerade für die Korpuslinguistik und die digitale Diskurslinguistik. Doch auch in anderen Bereichen der Linguistik (bspw. der Pragmatik oder Soziolinguistik) wird Annotation immer mehr angewendet. Derzeit verändert sich das methodische Konzept des Annotierens in diesen Forschungsrichtungen dahingehend, dass Annotation als Erkenntnisprozess, der selbst Forschung ist, stärker im Vordergrund steht. Das durch die Computerlinguistik schon länger eingeführte Verständnis von Annotation hingegen ist vor allem auf die Auszeichnung meist systemlinguistischer Kategorien nach einem zuvor festgelegten Gold-Standard, die Messung des Inter-Annotator-Agreements und die Automatisierung bzw. automatisierte Auswertbarkeit ausgerichtet. Es hat sich tendenziell dahingehend entwickelt, dass unter Annotation oft ‚nur‘ das routinemäßige Tagging als mehr oder weniger standardisierter Erschließungsschritt angesehen wird, und zwar als ein Arbeitsgang vor der eigentlichen Analyse und Interpretation der Daten.

In linguistischen Disziplinen, in denen hermeneutische Interpretation eine zentrale Rolle spielt und etwa auch implizite sprachliche Phänomene analysiert werden, wie zum Beispiel in der Pragmatik (vgl. Archer/Culpeper/Davies 2008 und Weisser 2015, 2018), ist eine andere Konzeption des Annotierens notwendig. Hier besteht der methodische Kern aus dem Zusammenspiel von deduktiver und induktiver Kategorienbildung vor dem Hintergrund linguistischer Theorien sowie der iterativen Ausdifferenzierung von Tagsets und Guidelines in einem teils kollaborativen und diskursiven Verfahren. Die Operationalisierung von Forschungsfragen und linguistischen Phänomenen in Kategorien-Definitionen und -Systemen sowie in den entsprechenden Annotationsrichtlinien ist nicht mehr nur eine Voraussetzung für die Analyse, sondern selbst eine wichtige Ergebnisdimension (vgl. Bender 2020, Bender/Müller 2020).

Schon 2007 wurde im vielzitierten Sammelband von Hermanns/Holly „Linguistische Hermeneutik“ dazu angeregt (hier Haß 2007), über eine „Korpus-Hermeneutik“ nachzudenken. Wir wollen diesen Gedanken aufgreifen und dafür plädieren, dass sie den aktuellen Forschungsstand zu Verfahren des Annotierens einbezieht, sich interdisziplinär und auch im Bezug zu den Digital Humanities verortet, aber linguistisch präzise positioniert. Ziel dabei sollte sein, das linguistische Wissen, welches wir an die Daten herantragen, transparent, plausibel und praktikabel aufzubereiten, also zu explizieren und zu operationalisieren, und in einen iterativ-dynamischen Forschungsprozess zu integrieren. So werden Forschungsansätze möglich, die über klassische korpuslinguistische Analysen der Frequenz bestimmter sprachlicher Kategorien u.ä. hinausgehen. So können etwa soziolinguistische und pragmatische Parameter inkludiert, auf verschiedene Ebenen angewendet und in Relation gesetzt werden.

Eine weitere Perspektive ist die iterative Verzahnung qualitativ-hermeneutischer Annotationsverfahren mit quantifizierend-algorithmischen Ansätzen. Dabei stellt sich die Frage, welche Kriterien die hermeneutisch-interpretative, qualitative Annotation erfüllen muss, damit sie mit automatisierten Verfahren, etwa des maschinellen Lernens, interagieren kann (vgl. Ide 2017).

Diese Perspektiven stellen den thematischen Kern und den Fokus des Austauschs dar, der durch die digitale Tagung ermöglicht werden soll. Da die Veranstaltung relativ kurzfristig konzipiert wurde und an den Tagen der ursprünglich geplanten Jahrestagung des Forschungsnetzwerkes »Sprache und Wissen« stattfindet, möchten wir anders vorgehen als üblich. Wir geben den thematischen Schwerpunkt vor, sind aber vorerst für alle Beitragsformate offen. Als mögliche Mitgestaltungsformen kommen in Frage:

- Diskussion eines bereits verfassten Beitrags,
- Diskussion eines Problemzusammenhangs aus einem Qualifikationsprojekt an einem Beispiel,
- Diskussion einer Fragestellung/Problemstellung mit Bezug zum Annotieren,
- ausführliche Präsentation einer Studie,
- exemplarischer Vergleich verschiedener Verfahren,
- Kurzpräsentation zu theoretischen oder methodischen Aspekten,
- o.Ä.

Die Tagung adressiert also ausdrücklich auch Early-Career- und Student-Researcher, die z.B. Qualifikationsarbeiten oder Annotationskonzepte bzw. Problemstellungen im Rahmen der Mitarbeit in Forschungsprojekten vorstellen möchten.

(Nachwuchs-)Wissenschaftler*innen aus anderen Disziplinen sind ebenso herzlich willkommen. Unsere Bitte wäre lediglich, dass ein klarer Sprachbezug bzw. Bezug zu linguistischen Vorgehensweisen gewährleistet ist, um einen differenzierten Austausch zu ermöglichen.

Wir möchten alle Annotationsinteressierten ermutigen, sich per Mail bei uns zu melden oder bis zum **15. August 2020** einen konkreten Vorschlag einzureichen.

Mail-Adressen:

michael.bender@tu-darmstadt.de

katharina.jacob@gs.uni-heidelberg.de

In einem ersten Schritt werden wir alle Beitragsformate aufnehmen, auf Qualität und Machbarkeit prüfen, ein Programm erstellen und dann Rückmeldung geben, in welcher Form der Beitragsvorschlag angenommen und umgesetzt werden kann. Bis Ende August werden wir dann ein konkretes Programm erstellen.

Die Tagung findet via heiCONF vom 30. September (gegen Mittag) bis zum 02. Oktober (gegen Mittag) statt. Bei kürzerem Programm beschränken wir die Zeit auf den Donnerstag und Freitag.

Wir freuen uns auf reges Interesse und verbleiben mit den besten Grüßen,
Dr. Michael Bender (TU Darmstadt) und Dr. Katharina Jacob (Universität Heidelberg)

Literatur

Archer, Dawn; Culpeper, Jonathan; Davies, Matthew (2008): Pragmatic Annotation. In: Lüdeling, Anke; Kytö, Merja (Hg.): Corpus Linguistics. An International Handbook. Berlin: de Gruyter, S. 613-641.

Bender, Michael (2020): Annotation als Methode der digitalen Diskurslinguistik. In: Diskurse digital. Theorien – Methoden – Fallstudien. Band 2, Heft 1/2020: 1-35. DOI: <https://doi.org/10.25521/diskurse-digital.2020.140>

Bender, Michael; Müller, Marcus (2020): Heuristische Textpraktiken in den Wissenschaften. Eine kollaborative Annotationsstudie zum akademischen Diskurs. In: Zeitschrift für Germanistische Linguistik 48 (1)/2020: 1-46.

Haß, Ulrike (2007): Korpus-Hermeneutik. Zur hermeneutischen Methodik in der lexikalischen Semantik. In: Hermanns, Fritz; Holly, Werner (Hg.): Linguistische Hermeneutik. Theorie und Praxis des Verstehens und Interpretierens. Tübingen: Niemeyer, S. 241-261.

Hermanns, Fritz; Holly, Werner (Hg.) (2007): Linguistische Hermeneutik. Theorie und Praxis des Verstehens und Interpretierens. Tübingen: Niemeyer.

Ide, Nancy (2017): Introduction: The Handbook of Linguistic Annotation. In: Ide, Nancy; Pustejovsky, James (Hg.): Handbook of Linguistic Annotation. Vol. I, Dordrecht: Springer, S. 1-18.

Weisser, Martin (2015): Speech Act Annotation. In: Ajmer, Katrin; Rühlemann, Christoph (Hg.): Corpus Pragmatics. A Handbook. Cambridge: Cambridge University Press, S. 84-116.

Weisser, Martin (2018): How to Do Corpus Pragmatics on Pragmatically Annotated Data. Amsterdam, Philadelphia: John Benjamins.