

Von A wie *Abraham a Sancta Clara* bis U wie *Unbekannter Verfasser* Annotation und Repräsentation barocker Literatur

Der bekannte Prediger und Augustinermönch Abraham a Sancta Clara (1644-1709) gilt in der Literaturgeschichtsschreibung als einer der sprachmächtigsten Autoren seiner Zeit. Dass seine Schriften bisher in keiner Gesamtausgabe vorliegen, ist darauf zurückzuführen, dass der Umfang seines Œuvres nicht feststeht und sich seine Autorschaft besonders im Spätwerk verunklart: „Je leuchtender der Name und Stil des populären Autors in Erscheinung trat, desto fragwürdiger wurde der Bezug des realen Ordensmannes und Schriftsteller zu seinem Werk.“¹

Das Projekt², das im Vortrag vorgestellt werden soll, hat daher im Rahmen des Vorhabens *ABaC:us – Austrian Baroque Corpus* ein digitales Korpus erstellt, das u.a. eine Auswahl von Abraham a Sancta Clara zugeschriebenen Texten in ihren ersten zu identifizierenden Ausgaben enthält. Es eröffnet der Forschung damit einen neuen, unvoreingenommenen Blick auf diese Quellen³ und verfolgt folgende Ziele:

1 Erstellung einer verlässlich annotierten Textgrundlage

Die diplomatischen Transkriptionen der Frakturdrucke (u.a. von *Mercks Wienn*, *Lösch Wienn*, *Grosse Todtenbruderschaft*, *Augustini Feurigs Hertz* und *Besonders meubliert- und gezierte Todtencapelle*) wurden im XML-Format erstellt, folgen dem international empfohlenen de-facto Standard der Text Encoding Initiative (Version P5) und bilden den historischen Sprachstand der Texte unverändert, d.h. zeichen-, zeilen- und seitengetreu ab. Die linguistische Annotation der historischen Texte im Umfang von 200.000 Token erfolgte semi-automatisch, d.h. die Wortklassenzuordnung⁴ und Lemmatisierung⁵ auf Basis des *Treetagger* (der für die deutsche Gegenwartssprache entwickelt wurde und daher mit den graphematischen Varianten, wie sie in Barocktexten vorkommen, größtenteils nicht umgehen

¹ Franz M. Eybl: Wissenslücken um Abraham a Sancta Clara – Zur Problematik populärer Autorschaft. In: Unterhaltender Prediger und gelehrter Stofflieferant. Abraham a Sancta Clara. Eggingen 2012, S. 104.

² Das Projekt „Texttechnologische Methoden zur Analyse österreichischer Barockliteratur“ (Laufzeit 2012-2014) wird durch den Jubiläumsfonds der Österreichischen Nationalbank, Projektnummer: 14738 gefördert.

³ Die Vielzahl von populären Auswahlgaben oder „Blütenlesen“, die zu Abraham a Sancta Clara bis heute publiziert werden, hat sich für wissenschaftliche Fragestellungen als unzureichend erwiesen.

⁴ Die Wortklassenzuordnung erfolgte auf Basis des 54-teiligen Stuttgart-Tübingen TagSets (STTS), das für die historische Sprachstufe des Älteren Neuhochdeutsch adaptiert und um weitere Tags ergänzt wurde, etwa bei kontrahierten Formen wie *wirstu* (VVFİN_PPER) oder *mans* (PIS_PPER).

⁵ Als Referenzwerke für die Lemmatisierung wurden der „Duden“ sowie das „Deutsche Wörterbuch“ von Jacob und Wilhelm Grimm herangezogen bzw. für lateinische Belege das „Lateinisch-deutsche Schulwörterbuch von Stowasser“. Sogenannte „out-of-vocabulary words“, die in keinem der genannten Wörterbücher vorkommen, tragen einen entsprechenden Vermerk.

konnte⁶) wurden durchgehend von zwei AnnotatorInnen mit Hilfe des an der Österreichischen Akademie der Wissenschaften entwickelten *token_editor* verifiziert beziehungsweise manuell korrigiert.

2 Verbesserung des automatischen Taggings historischer Texte

Die bereits annotierten Daten der ersten Texte wurden in der Folge dafür verwendet, die Erfolgsquote des automatisch generierten Taggings weiterer Texte zu verbessern, indem die bereits identifizierten und systematisierbaren Fehleinträge der ersten Werke in den noch nicht annotierten Textstrecken berücksichtigt wurden. Das verlässlich vollannotierte Korpus könnte in Zukunft maßgeblich dazu beitragen, weitere Texte aus dieser Zeitperiode effizienter zu annotieren bzw. die Leistung verschiedener Tagger daran zu messen.

3 Erweiterung des Wissens über abrahamische Spezifika

Mit dem Aufbau der multifunktionalen Sprachressource hat die Projektgruppe Voraussetzungen für sprach- und literaturwissenschaftliche Fragestellungen geschaffen, die sie selbst exemplarisch beantwortet: Erstmals ist man in der Lage, den Wortschatz des Autors, dessen „Sprachmächtigkeit“ in Literaturgeschichten ausdrücklich gewürdigt wird, zu fassen: Mit Textanalysetools wie der *Sketch Engine* oder *Voyant* können systematische Wortschatzanalysen, Konkordanzen und Type-Token-Relationen erstellt, Frequenzen, Kollokationsprofile und Wortklassenverteilungen ermittelt, sowie musterbasierte Abfragen generiert werden, wodurch sich ausgewählte sprachliche Phänomene und barocke musterhafte, stilbildende Elemente (wie Doppelformeln, Wiederholungs-, Häufungs- und Steigerungerscheinungen) in abrahamischen Texten identifizieren und quantifizieren lassen. „Abrahamischen Stil“ anhand ausgewählter Beispiele zu beschreiben, war bereits das Anliegen einiger älterer Untersuchungen⁷, deren Ergebnisse nun überprüft, auf größere Textmengen bezogen, systematisch ausgewertet und damit auf ein begründetes empirisches Fundament gestellt werden können. Was die bis heute ungeklärte Autorenschaft des angeblich letzten Werks von Abraham a Sancta Clara angeht, möchte der Vortrag erstmals eine Reihe von Argumenten vorbringen, die für bzw. gegen eine tentative Zuschreibung sprechen.

⁶ Dass die Texte von Abraham a Sancta Clara im Vergleich mit anderen (sogar älteren) Texten die höchste Fehlerquote aufwiesen, haben Erhard Hinrichs und Thomas Zastrow bereits festgestellt – vgl. ihre Studie „Linguistic Annotations for a Diachronic Corpus of German“. In: *Linguistic Issues in Language Technology*, Volume 7 (2012), S. 1-16, hier S. 11.

⁷ Vgl. etwa Curt Blanckenburg: *Die Sprache Abrahams a S. Clara. Ein Beitrag zur Geschichte der deutschen Drucksprache*. Halle an der Saale: Ehrhardt Karras 1897; Hans Strigl: *Einiges über die Sprache des P. Abraham a Sancta Clara*. In: *Zeitschrift für Deutsche Wortforschung*. Hrsg. v. Friedrich Kluge. Band 8 (1906), S. 206-312; Margaretha Stiassny: *Das Wortspiel bei Abraham a Sancta Clara*. Phil. Diss. Wien 1939 und Norbert Bachleitner: *Form und Funktion der Verseinlagen bei Abraham a Sancta Clara*. Frankfurt am Main / Bern / New York: Peter Lang 1985.

4 Publikation in einem webbasierten Interface und Integration in europäische Forschungsinfrastrukturen

Das vorläufige Ergebnis des Projektes stellt die Integration von fünf abrahamischen Texten in ein webbasiertes Interface dar, das das Korpus über unterschiedliche Wege des Zugriffs nutzbar macht (vgl. Abbildung). Besonders im Zusammenhang mit zeitentfernten Texten wie diesen ist die Frage nach der Art und Weise ihrer Repräsentation im digitalen Kontext eine wesentliche. Das trifft auf das vorliegende Projekt besonders zu, als es sich zum erklärten Ziel gesetzt hat, die Texte nicht nur anderen, textbezogenen Wissenschaften und der universitären Lehre zur Verfügung zu stellen, sondern sie auch einer interessierten nicht-fachlichen Öffentlichkeit näherzubringen.

Wert gelegt wurde insbesondere darauf, die Texte als Werke in ihrer Eigenständigkeit zugänglich zu machen, ohne jedoch das Korpus als ihren Verbund in den Hintergrund zu rücken. Daher ermöglicht das Interface zum einen den „lesenden“ Zugang zum Material über eine seitenweise synoptische Ansicht von Volltext und Faksimile, die durch Inhaltsverzeichnisse, Personennamen- und Ortsregister erschlossen ist, und auch die Einbeziehung von graphischen und typographischen Eigenschaften der Drucke bei der Textinterpretation ermöglicht. Zum anderen bietet die Plattform auch die Möglichkeit, *ABaC:us* auf Wortformen, Lemmata und Part-of-Speech-Tags einzeln und in Kombination zu durchsuchen sowie Frequenzlisten zu generieren.

Die Funktionalität des digitalen Suchens und Navigierens basiert technisch auf einfachen XPath-Pfadausdrücken, die zur Adressierung von Dokumentteilen dienen. Eine frei konfigurierbare Gruppe solcher Indizes bildet beispielsweise das Gerüst des hierarchischen Inhaltsverzeichnisses oder der synoptischen Ansicht, die dynamisch aus den Gesamttexten extrahiert wird. Mit dem vorliegenden semantischen Markup der Texte wäre es etwa ein Leichtes, mittels eines entsprechenden Pfadausdruckes ein Register von Bibelstellen zu erstellen. Somit ist das Interface für die flexible Erweiterung des Korpus sowohl hinsichtlich seines Umfangs als auch seiner Funktionalität ausgerichtet und ermöglicht, das in den Daten kodierte Wissen über die Texte entsprechend sich verändernden Anforderungen nutzbar zu machen.

Die Weboberfläche von *ABaC:us* setzt auf dem am ACDH entwickelten *cr-xq Content Repository* auf, das ein Teil des modularen *corpus_shell* Frameworks⁸ ist und auf dem *CLARIN Federated Content Search*-Standard⁹ basiert. Durch die Einbettung in eine wachsende europäische Infrastruktur an Forschungsdaten ist die Verfügbarkeit der Projektdaten auch in Zukunft gewährleistet, so dass künftige NutzerInnen der Korpusdaten die Annotationen zeitsparend, zweckmäßig und gewinnbringend für ihre Erkenntnisinteressen einsetzen können.

⁸ <http://www.oeaw.ac.at/icltt/node/4>

⁹ <http://www.clarin.eu/content/federated-content-search-clarin-fcs>

Konkordanz

Link
Zitation: Abraham à Sancta Clara, Löstch Wienn, 1880. (Digitale Ausgabe) Ad Lectorem [S. 1]. In: Abacus – Austrian Baroque Corpus Hrsg. von Claudia Resch und Ulrike Zetzlschner. <http://corpus4.aac.ac.at/wiens/apps/lor-q/mets/abacus/get/abacus.5/bac5_21> abgerufen am 21. 10. 2014