

Artikel aus:  
Zeitschrift für digitale Geisteswissenschaften

Titel:  
»Etwas für alle« – Ausgewählte Texte von und mit Abraham a Sancta Clara digital

Autor/in:  
Claudia Resch

Kontakt:  
[claudia.resch@oeaw.ac.at](mailto:claudia.resch@oeaw.ac.at)

Institution:  
Austrian Centre for Digital Humanities, Österreichische Akademie der Wissenschaften

GND:  
[132312794](#)


ORCID:

---

DOI des Artikels:  
[10.17175/2016\\_005](https://doi.org/10.17175/2016_005)

Nachweis im OPAC der Herzog August Bibliothek:  
[863829074](#)

Erstveröffentlichung:  
10.02.2017

Lizenz:  
Sofern nicht anders angegeben 

Medienlizenzen:  
Medienrechte liegen bei den Autoren

Letzte Überprüfung aller Verweise:  
10.02.2017

GND-Verschlagwortung:  
[Literaturwissenschaft](#) | [Linguistik](#) | [Daten](#) | [Annotation](#) | [Edition](#) | [Frühe Neuzeit](#) |

Zitierweise:  
Claudia Resch: »Etwas für alle« – Ausgewählte Texte von und mit Abraham a Sancta Clara digital. In: Zeitschrift für digitale Geisteswissenschaften. 2017. PDF Format ohne Paginierung. Als text/html abrufbar unter DOI: [10.17175/2016\\_005](https://doi.org/10.17175/2016_005).

Claudia Resch

## »Etwas für alle« – Ausgewählte Texte von und mit Abraham a Sancta Clara digital

---

### Abstracts

In einer web-basierten Edition werden am Austrian Centre for Digital Humanities ausgewählte Texte von und mit Abraham a Sancta Clara mit innovativen, digitalen Methoden erschlossen. Die Texte sind Teil des sogenannten *Austrian Baroque Corpus (ABaC:us)* und stehen anmeldungsfrei unter <https://acdh.oeaw.ac.at/abacus/> online zur Verfügung. Welche grundsätzlichen Entscheidungen in der Praxis der Annotation und bei der Visualisierung der Daten getroffen worden sind, damit die Ressource idealiter »Etwas für alle« bieten und sowohl sprachwissenschaftlichen als auch literatur- und kulturwissenschaftlichen Forschungsinteressen dienlich sein kann, ist eine der Fragen, mit der sich vorliegender Beitrag näher beschäftigt.

This paper aims to introduce a new web-based digital edition of printed German texts written by or ascribed to Abraham a Sancta Clara, who was a renowned Augustinian monk and a widely read author at his time. The texts are part of the thematic memento mori research collection, *Austrian Baroque Corpus*, now freely available through the Austrian Centre for Digital Humanities: <https://acdh.oeaw.ac.at/abacus/>. One of the main considerations of the paper was how to shape the annotation process in a way that would meet the needs of linguistic, literary and cultural studies. This objective necessitated a user-friendly visualization that would be suitable to the research questions of a broader audience.

## 1. Zur Projektgenese

Das in vorliegendem Beitrag beschriebene Projekt ist ein Beispiel für jene Arbeitsvorhaben, die Patrick Sahle in seiner dreibändigen Publikation über digitalen Editionsformen als »kleineres Projekt«<sup>1</sup> bezeichnet: Seine Anfänge gehen auf ein personenbezogenes Forschungsstipendium<sup>2</sup> der Stadt Wien zurück. Der 300. Todestag des bekannten Predigers und Schriftstellers Abraham a Sancta Clara (1644–1709) war in Wien unbeachtet vorübergegangen, als aus den Mitteln dieses Stipendiums erste Image-Scans als Vorlage zur Volltextdigitalisierung angekauft werden konnten. In den Jahren 2012–2014 wurde das Vorhaben maßgeblich durch den Jubiläumsfonds der Österreichischen Nationalbank gefördert. Durch die Bewilligung des Projekts *Texttechnologische Methoden zur Analyse österreichischer Barockliteratur* konnte an der Österreichischen Akademie der Wissenschaften (ÖAW) am Institut für Corpuslinguistik und Texttechnologie (ICLTT) ein interdisziplinäres Projektteam<sup>3</sup> aus den Fachbereichen Literaturwissenschaft, Sprachwissenschaft und Korpuslinguistik zusammenfinden, das einerseits den Einsatz digitaler Methoden, Annotationsstandards und Tools an zeitentfernten

---

<sup>1</sup> Sahle 2013 im Kapitel über »Träger und Akteure«, S. 76ff.

<sup>2</sup> Barocke literarische Totentanz-Texte von und mit Abraham a Sancta Clara (2010).

<sup>3</sup> Aus den Projektmitteln anteilig beschäftigt waren Thierry Declerck (Deutsches Forschungszentrum für Künstliche Intelligenz / Universität Saarbrücken), Barbara Krautgartner, Katharina Wappel und Eva Wohlfarter.

Textquellen aus der Barockzeit erproben und andererseits zuverlässige Erkenntnisse über deren Spezifika gewinnen wollte. Die dafür nötigen konzeptionellen und technischen Fertigkeiten wurden im Verlauf des an Mitteln beschränkten Projekts erworben und entwickelt.<sup>4</sup> Dass die erschlossenen Texte nicht nur beforscht wurden, sondern auch unmittelbar nach Abschluss des Projekts Sichtbarkeit erlangten, indem sie auch digital zur Verfügung standen, war im Projektplan nicht intendiert und wäre ohne institutionelle Unterstützung und die interne intensive Zusammenarbeit am Austrian Centre for Digital Humanities (ACDH) nicht möglich gewesen.

## 2. Texte von und mit Abraham a Sancta Clara

Dem Projekt war nicht daran gelegen, große Textmengen zu erschließen. Vielmehr ging es zunächst um die Sammlung und Zusammenführung von ausgewählten Drucken der Barockzeit<sup>5</sup>, die sich mit der Vorbereitung auf den Tod beschäftigten, namentlich um Memento mori-, Ars moriendi- und Totentanzliteratur. Mit der Wahl dieses thematischen Schwerpunktes wurde spezifischen Forschungsinteressen Rechnung getragen, aber auch der Tatsache, dass todesbezogene Texte innerhalb der Literaturproduktion des Untersuchungszeitraumes einen hohen Anteil und Stellenwert hatten. Die aufgenommenen Texte sind (zum Teil auch weniger beachtete) Vertreter einer reichhaltigen, weit verbreiteten und für die Barockzeit charakteristischen literarischen Tradition. Die entstandene textsortenbestimmte Sammlung trägt den Namen *Austrian Baroque Corpus (ABaC:us)* und enthält derzeit etwa 20 Werke mehrerer Autoren vorwiegend aus dem 17. und beginnenden 18. Jahrhundert.

Unter den Autoren kommt dem heute noch bekannten Augustiner Barfüßerprediger Abraham a Sancta Clara, eigentlich Johann Ulrich Megerle (1644–1709)<sup>6</sup> in dieser digitalen Sammlung barocker Literatur besonderer Stellenwert zu: Fünf seiner, beziehungsweise der ihm zugeschriebenen<sup>7</sup>, todesbezogenen Schriften haben alle Stufen der Annotation durchlaufen und stehen seit 5. Mai 2015 in ihren Erstdrucken als digitale Edition zur Verfügung, darunter Bestseller wie *Mercks Wienn* (1680), aber auch weniger bekannte Neujahrsgaben der in St. Augustin ansässigen Totenbruderschaft, deren geistliches Oberhaupt Pater Abraham war, sowie die emblematische *Todten=Capelle*, die 1710, mit seinem werbewirksamen Namen versehen, als sein angeblich letztes Werk veröffentlicht wurde:

1. Mercks Wienn / Das ist Deß wütenden Todts ein vmbständige Beschreibung In Der berühmten Haut vnd Kayserl. Residentz Statt in Oesterreich / Jm sechzehnen hundert /

---

<sup>4</sup> Auch diesen Umstand bezeichnet Sahle als ein Kennzeichen »kleinerer Projekte«, vgl. Sahle 2013, S. 77.

<sup>5</sup> Der Barockbegriff wurde lange Zeit als »Übergangsbezeichnung« verwendet. In der germanistischen Literaturwissenschaft ist diese Sprachregelung heute kaum noch umstritten und »Barock« hat sich als Verständigungsbegriff für die Literatur des 17. Jahrhunderts und des beginnenden 18. Jahrhunderts beziehungsweise für den mittleren Abschnitt der Frühen Neuzeit etabliert. Vgl. etwa Braungart 2006, Meid 2009 und Niefanger 2012.

<sup>6</sup> Vgl. die Lexikoneinträge von Franz M. Eybl 2008, S. 10–14 oder Wilhelm Friedrich Bautz 1990, S. 10–11. Mit Abraham a Sancta Claras publizistischen Erfolgen hat sich eine ganze Reihe von Wissenschaftlern und Wissenschaftlerinnen beschäftigt. Das Interesse an der Person und Biographie des Predigers und die Auseinandersetzung mit seinen Schriften und deren Rezeption weisen mehrere »Kulminationsphasen« auf, die Franz M. Eybl in seiner Habilitationsschrift eingehend dargestellt hat. Vgl. Franz M. Eybl 1992, S. 6–23.

<sup>7</sup> Zur Problematik populärer Autorschaft vgl. Franz M. Eybl 2012, S. 104–121.

vnd neun vnd sibentzigsten Jahr / [...] Gedruckt zu Wienn / bey Peter Paul Vivian / der löbl: Universitet Buchdrucker 1680.<sup>8</sup>

2. Lösch Wienn / Das ist Ein Bewögliche Anmahnung zu der Kays. Residentz=Statt Wienn in Oesterreich / Was Gestalten / Dieselbige der so viel tausend Verstorbene Bekanten vnd Verwandten nicht wolle vergessen [...] Gedruckt zu Wien / bey Peter Paul Vivian / 1680.<sup>9</sup>

3. Grosse Todten Bruderschaft / Das ist Ein kurtzer Entwurff Deß Sterblichen Lebens / Mit beygefügtten CATALOGO, Oder Verzeichnus aller der Jenigen Herren Brüdern / Frauen / und Jungfrauen Schwesternen / welche [...] von Anno 1679. biß 1680. gestorben seyn. Gedruckt im Jahr 1681.<sup>10</sup>

4. AUGUSTINI Feuriges Hertz Tragt Ein Hertzliches Mitleyden mit den armen im Feeg=Feuer Leydenden Seelen / Das ist / Ein kleiner Haußrath etliche Sentenz auß den Schrifftten vnsers Heil. Vatters. [...] Gedruckt zu Saltzburg bey Melchior Haan [...] Anno 1693.<sup>11</sup>

5. Besonders meublirt- und gezierte Todten=Capelle / Oder Allgemeiner Todten=Spiegel / Darinnen Alle Menschen / wes Standes sie sind / sich beschauen / an denen mannigfältigen Sinnreichen Gemählden das MEMENTO MORI zu studiren [...]. Nürnberg [...] Druckts Marrtin Frantz Hertz. An. 1710.<sup>12</sup>



Abb. 1: LINKS: Portrait Abraham a Sancta Claras, das einem seiner Nachrufe vorangestellt wurde. Quelle: Johann Carl Megerle, ABRAHAM ist gestorben. Wien [1710]. Österreichische Nationalbibliothek Wien, Signatur 220677-C. RECHTS: Titelblätter und Kupferstiche der in der *ABaC:us*-Edition enthaltenen Schriften. Eigene Darstellung © ACDH.

Bisweilen stellt sich für *ABaC:us*-Rezipienten und -Rezipientinnen die Frage, weshalb die im Korpus enthaltenen Drucke nicht historisch-kritisch ediert worden sind. Dazu ist festzuhalten, dass es im Projekt primär darum gegangen ist, eine digitale Sammlung mehrerer Werke aufzubauen, um damit textimmanent und vergleichend sprachliche Phänomene zu untersuchen, und nicht darum, ein Werk in all seinen Facetten zu erschließen. Wollte man für jedes in *ABaC:us* enthaltene Werk alle zur Verfügung stehenden Textzeugen berücksichtigen

<sup>8</sup>Stiftsbibliothek Melk, Signatur 48.022.

<sup>9</sup>Rara II Sammlung der Universität Graz, Signatur I 25.896.

<sup>10</sup>Sammlung von Handschriften und alten Drucken der Österreichischen Nationalbibliothek, Signatur 298.002-A; der Erstdruck dieses Werkes gilt seit dem 2. Weltkrieg als verschollen.

<sup>11</sup>Sammlung von Handschriften und alten Drucken der Österreichischen Nationalbibliothek, Signatur 484.862-A.

<sup>12</sup>Emblemsammlung der Universität Illinois, Signatur 832Ab8. Im Gegensatz zu jenem Exemplar aus der Universitätsbibliothek Halle, das von Wolfgang Harms und Michael Schilling für den Reprint 2003 verwendet wurde, weist dieses keine Fehlbindungen auf und enthält zudem eine Vorrede von Georg Augustin Widtman, die in nur wenigen Drucken enthalten, für die Genese des Werkes als Totenbruderschaftskalender jedoch von essentieller Bedeutung ist.

und bewerten, käme dies einem mehrjährigen Großprojekt gleich, wenn man bedenkt, in wie vielen Ausgaben die Abraham a Sancta Clara zugeschriebenen Werke gedruckt worden sind. Nicht zuletzt deshalb sind die Nachdruckverhältnisse seiner Werke zu Lebzeiten und postum nach wie vor größtenteils völlig unerforscht. Im Rahmen der Möglichkeiten hat das Projektteam dennoch versucht, textkritisch vorzugehen, was im folgenden Abschnitt anhand von Beispielen präzisiert wird.

### 3. Erschließungstiefe und Ebenen der Annotation

Zur Erstellung originalnaher Transkriptionen wurden die Druckvorlagen mit XML und verwandten Technologien zu maschinenlesbarem Text verarbeitet und gemäß international empfohlener Standards (TEI – Text Encoding Initiative in der Version P5 <http://www.tei-c.org/Guidelines/P5/index.xml>) erschlossen. Die digitalen Daten bilden die jeweilige Druckvorlage seiten-, zeilen-, und zeichengetreu ab. Sie haben mehrfache Kollationierungsgänge durchlaufen, sind hinsichtlich ihrer Qualität sorgfältig überprüft und geben den historischen Sprachstand der Texte unverändert wieder.

Die Kodierung ist allerdings nicht nur dokumentarisch, weil sie den Druckvorlagen auf das Genaueste folgt, sondern auch textkritisch, indem sie offensichtliche Textfehler des Originals<sup>13</sup> in der Transkription zwar belässt, aber dennoch ausweist und kommentiert: Editorische Anmerkungen wurden durch rote, hochgestellte Klammern dargestellt, die sich mit einem jeweiligen Korrekturvorschlag öffnen, sobald Benutzer und Benutzerinnen mit dem Cursor in deren Nähe kommen. Wer diese Fehler ehemals verursacht hat – der Autor selbst, der Setzer oder der Drucker? – muss offen bleiben; die Beispiele zeigen eine der sehr häufigen Buchstabenauslassungen sowie einen inhaltlichen Rechenfehler (Abbildung 2 und Abbildung 3), der in der Vorrede zur *Todten-Capelle* fälschlicherweise Abrahams Lebensalter um zwei Jahre verlängert.

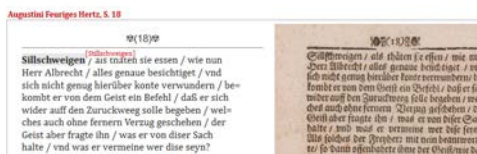


Abb. 2: Kommentierung offensichtlicher Fehler. © ACDH.



Abb. 3: Kommentierung offensichtlicher Fehler. © ACDH.

<sup>13</sup> Unter Berücksichtigung des historischen Sprach- und Schreibstandes ist grundsätzlich zu überlegen, wie tolerant Herausgeber und Herausgeberinnen bei der Einschätzung von Fehlern sein sollen. Allein im Fall von *Mercks Wienn* hat Werner Welzig bislang den Versuch unternommen, ein Register von Fehlern anzulegen (vgl. Deutsche Neudrucke, Reihe Barock, Band 31, 1983, Nachwort S. 13\*f.), das bei der vorliegenden Edition berücksichtigt ist.

In den Texten wurden außerdem historische, biblische und mythologische Personennamen sowie Ortseigenamen annotiert. In der letztgenannten Kategorie wurden die Namen von Bergen, Gewässern, Regionen, Ländern und Kontinenten, Städten und Dörfern sowie Straßen und Plätzen in Wien ausgewiesen.

### 3.1. Linguistische Annotation

Im Hinblick auf die (spätere) Durchsuchbarkeit und Nachnutzung der historischen Texte wurde vergleichsweise viel Zeit in deren linguistische Annotation bestehend aus Wortartklassifizierung (*Part-of-Speech*) und Lemmatisierung investiert: Das gedruckte Ältere Neuhochdeutsch besitzt einerseits zwar bereits hinreichend Ähnlichkeit mit heutigen Texten, sodass sich die Erprobung und der Einsatz von Annotationswerkzeugen lohnt, andererseits weicht es – was Lexik, Flexionsmuster, Syntax und Graphie betrifft – deutlich von allem ab, was Tagger, die auf die Analyse von Gegenwartssprache ausgerichtet sind, (er)kennen können, sodass man beim Einsatz automatischer Tools unweigerlich an Grenzen stößt. Nachdem Zeit und Mittel begrenzt waren und die Datenmenge (180.000 Token) überschaubar schien, wurden zunächst drei der fünf Werke in einem ersten Schritt automatisch annotiert, indem zuerst der Zeichenstrom der Rohtexte auf Wortebene segmentiert und in einem weiteren Schritt jedes einzelne Token mit Hilfe der Software **TreeTagger** einer morphosyntaktischen Wortklasse (unter Verwendung des 54-teiligen **Stuttgart-Tübingen-TagSets**) zugewiesen und mit einem Lemma versehen wurde. In einem zweiten Schritt wurden diese Zuweisungen kontrolliert, manuell nachbearbeitet und stufenweise verbessert. Trotz technischer Erleichterungen durch den Einsatz des am Institut entwickelten *token\_Editor* war dieser Korrekturvorgang aus mehreren Gründen zeitaufwendig:

(1) Zum einen blieben die Leistungen des Taggers erwartungsgemäß weit hinter den erwünschten Ergebnissen zurück.<sup>14</sup> Durch die hohe Anzahl an Fehlzuweisungen mussten jede einzelne Wortart-Attribuierung und das dazugehörige Lemma in seinem Kontext überprüft werden.<sup>15</sup> Die Erfahrungen des Annotatorinnenteams bestätigten, dass Wortformen, die oft nur geringste Abweichungen von der heutigen Norm aufwiesen, vom TreeTagger nicht identifiziert werden konnten und zu falschen Klassifizierungen führten. Die häufigsten fälschlich vergebenen Tags waren Eigennamen (NE), wobei eine daran anschließende Tagstromanalyse zeigen konnte, dass sich die fälschlich vergebenen NE-Tags vor allem auf eigentliche Nomen oder fremdsprachliches Material verteilten.

---

<sup>14</sup> Diese Erfahrungen bestätigen auch Stefanie Dipper 2010, S. 117–121 sowie Erhard Hinrichs und Thomas Zastrow 2012, S. 1–16. Das Testset der Letztgenannten enthielt unter anderen den Abraham a Sancta Clara zugeschriebenen Text *Wunderlicher Traum. Von einem grossen Narren-Nest* (1703) in einer Version des Projekts Gutenberg. Bei der Auswertung hat sich gezeigt, dass dieser Text beim automatischen Tagging-Verfahren ungleich mehr fehlerhafte Annotationen aufweist als etwa Melanchtons (deutlich ältere) *Augsburger Konfession* (1530). Es steht zu vermuten, dass sich – neben den historischen Schreibvarianten und unbekanntem Lexemen – vor allem die Komplexität und die Länge der Sätze in abrahamischen Texten negativ auf die Leistung des TreeTaggers auswirken.

<sup>15</sup> In der Annotationspraxis zeigte sich, dass kaum generelle Entscheidungen bei der Nachkorrektur getroffen werden können: Ein hochfrequentes Wort wie »in« war in den meisten Fällen tatsächlich eine Präposition – im lateinischen Kontext allerdings wäre die Wortform als »fremdsprachliches Material« auszuweisen.

(2) Des Weiteren erschwerten historische Konventionen der Zusammen- und Getrennschreibung die Zuordnung: Für kontrahierte Formen wie etwa »wirst«, »meinst« oder »mustu« waren im standardisierten Stuttgart-Tübingen-TagSet (STTS) keine Kategorien vorgesehen, weshalb in Analogie zu den bestehenden Kategorien neue, kombinierte Tags eingeführt und in den Registern als solche ausgewiesen wurden (für die oben genannten Beispiele etwa VAFINPPER, VVFINPPER und VMFINPPER).

(3) Bei der Lemmatisierung<sup>16</sup> wurde für jedes Token eine normalisierte Grundwortform angesetzt und der Duden sowie das Deutsche Wörterbuch von Jacob und Wilhelm Grimm als Referenzwerke herangezogen. Die mehr als 1000 in den Texten vorkommenden Wortformen, zu denen es in den beiden genannten Referenzwerken keinen entsprechenden Wörterbucheintrag gab, wurden als sogenannte *out-of-vocabulary-words* gekennzeichnet und auf eine naheliegende Grundform zurückgeführt – Beispiele hierfür waren Lexeme wie »beindrechtslerisch«, »brumbrumbrummend« oder »Butterkind« und viele weitere mehr.

(4) Eine besondere Herausforderung stellten Wortspiele dar, die keine eindeutige Zuweisung zu einem Lemma erlaubten: Exemplarisch ist eine Textstelle zu nennen, in der es darum ging, welche Farbe die schönste sei, nämlich die goldene, die in einer betörten Welt den Vorzug hätte gegenüber der »weissen«. Der Autor lässt hier absichtlich beide Lesarten offen, weil sowohl von der »weißen Farbe« als auch von der »weisen Welt« die Rede sein könnte. Bei der Annotation wurde in solchen Fällen wortwitzerhaltend vorgegangen, indem für ein und dieselbe Wortform zwei Lemmata angesetzt wurden, die beide Lesarten zuließen, sodass diese Stelle unter beiden Einträgen – »weiß« oder »weise« – zu finden wäre (vgl. Abbildung 4).

abgeben: Nein / nein / du hast es nicht  
errathen / Goldfarb ist die schönste /  
wer mit solcher angestrichen ist / der  
gefällt jedermänniglich / vnnd schiest  
keine weniger ab als dise / Goldfarb  
hat bey der bethörtē Welt den Vorzug  
vor der weissen / vnnd möchte jemand  
weiß | noch  
weise  
ADJA

Abb. 4: Annotation von abrahamischen Wortspielen – eine Wortform, zwei Lesarten © ACDH.

Die Entscheidungen, die bei der Korrektur der automatischen Annotation getroffen werden mussten, erforderten aber nicht nur Zeit; vielmehr war die linguistische Analyse dieser historischen Sprachdaten an und für sich eine komplexe, höchst anspruchsvolle

<sup>16</sup> Für Texte historischer Sprachstadien, in denen Graphien stärkere Varianz aufweisen als heute, ist die Lemmatisierung unserer Ansicht nach unerlässlich, weil dadurch nicht nur flektierte Formen abgefragt werden können, sondern alle vorkommenden Schreibvarianten, vgl. etwa die Suche nach dem Lemma »Hebräer«, die im konkreten Fall zu den Wortformen »Hebreer«, »Hæbreer«, »Hebrær«, »Ebræer«, »Ebreer« und »Hebr.« führt.

Forschungsaufgabe, die »ein profundes historisch-kulturelles Wissen, solide sprachhistorische Kenntnisse und Vertrautheit mit Texten und Texttraditionen sowie viel Feingefühl bei der Beurteilung«<sup>17</sup> erfordert, wie Wulf Oesterreicher bestätigt. Im *ABaC:us*-Projekt waren neben der Verfasserin vor allem Barbara Krautgartner und Eva Wohlfarter mit dieser Aufgabe betraut, wodurch die Konsistenz der annotierten Daten gewährleistet blieb. Die zahlreichen schwierigen Fälle wurden eingehend besprochen und – wenn es mehrere Lösungsansätze gab – gemeinsam diskutiert und entschieden.

Der oben gezeigte Screenshot (Abbildung 4) zeigt, dass in der Webapplikation jeder Beleg mit Lemma- und PoS-Informationen unterlegt ist, sobald man mit dem Cursor in dessen Nähe kommt: Aus der Sicht der Benutzer kann diese Anzeige (etwa bei heute nicht mehr gebräuchlichen Schreibvarianten) dazu beitragen, das Textverständnis zu verbessern beziehungsweise kann nach diesen Lemmata und Wortarten auch gesucht werden. Gleichzeitig dient diese Anzeige auch der Dokumentation, wodurch die linguistische Annotation als eigene wissenschaftliche Leistung sichtbar und für andere Benutzer und Benutzerinnen nachvollziehbar wird. Mit Sicherheit finden sich unter den 180.000 Token einzelne Fälle, die man anders hätte beurteilen können. Trotz Anwendung eines standardisierten Wortartenklassifizierungssystems bleibt in der Praxis ein gewisser Entscheidungsspielraum offen. Dass Annotation immer auch eine Form der Interpretation ist, muss zukünftigen Benutzern und Benutzerinnen daher jedenfalls bewusst sein.

### 3.2. Multifunktionalität einer Ressource

Die linguistische Annotation von Abraham a Sancta Claras Texten und solchen, die ihm zugeschrieben werden, hat gleichzeitig gezeigt, was ähnlich angelegte Projekte in Zukunft benötigen werden: Tools, mit denen man automatisch erzeugte Taggings benutzerfreundlich und effizient nachbearbeiten kann. Die annotierten Texte waren sicherlich ein Extremfall von Non-standard, dennoch sind die Probleme, die dabei zu lösen waren, weniger periodenspezifisch als erwartet (denn auch in anderen Geltungsbereichen hat man häufig mit Non-standard Varietäten zu tun). Umso erfreulicher ist es, dass vor dem Hintergrund der Annotationserfahrungen von *ABaC:us* am Austrian Centre for Digital Humanities intensiv an einem benutzerfreundlichen *webTokenEditor* gearbeitet wird, mit dem man Daten in tabellarischer Form schnell und unkompliziert im Browser wird korrigieren können.

Das Ergebnis des aufwendigen Annotationsverfahrens ist eine handverlesene Datenbasis, deren Qualität diese digitale Ressource zu einer gefragten Ausgangsbasis für weiterführende Forschungsanliegen macht: Anhand der Datenbasis lässt sich 1. die Leistung automatischer Taggingverfahren evaluieren,<sup>18</sup> 2. die Leistung automatischer Taggingverfahren regelbasiert

---

<sup>17</sup> Oesterreicher 2006, S. 493.

<sup>18</sup> Die Evaluierung des automatischen Taggings und die sich daran anschließenden Versuche, die Leistung des Taggers anhand noch nicht annotierter Texte mit der erarbeiteten Datenbasis als Hilfwörterbuch zu steigern, sind in folgenden Publikationen dokumentiert: Claudia Resch et al. 2014, S. 36–41, besonders Abschnitt 2.1 »Improvement through reliable data«, S. 37 und Claudia Resch et al. 2016. Erste Ergebnisse haben gezeigt, dass mit dem Einsatz zuverlässig korrigierter Daten des gleichen Genres und der gleichen Sprachperiode bei noch nicht annotierten Texten sowohl bei der Wortartenklassifizierung als auch bei der Lemmatisierung wesentlich bessere Ergebnisse zu erzielen sind.



verbessern,<sup>19</sup> und 3. können die generierten, maschinenlesbaren Wortformen- und Lemmalisten dazu beitragen, OCR-Prozesse weiterzuentwickeln.<sup>20</sup>

Darüber hinaus lassen sich die manuell überprüften Daten als verlässliche Basis für eine Vielzahl von (linguistischen) Fragestellungen nutzen. Ihr Wert liegt in der Wiederverwendbarkeit von bereits erarbeitetem, sorgfältig geprüfem Wissen über die Texte, was bedeutet, dass die Projektgruppe selbst, aber auch künftige Nutzerinnen und Nutzer die Annotationen zeitsparend, zweckmäßig und gewinnbringend für ihre Erkenntnisinteressen einsetzen können.

## 4. Etwas für alle: Datenrepräsentation in der ABaC:us-Edition

Seit 5. Mai 2015 stehen die angereicherten Daten als Teil des *Austrian Baroque Corpus* mit erläuternden Texten einem breiteren Benutzerkreis testweise, anmeldungsfrei und kostenlos online unter <https://acdh.oeaw.ac.at/abacus/> unter einer CC BY-NC 4.0 Lizenz zur Verfügung. Sofern vollständig auf die Quelle verwiesen wird, darf ABaC:us als Ressource für individuelle, nicht kommerzielle Forschungszwecke verwendet werden, wofür folgende Zitierweise empfohlen wird: ABaC:us – Austrian Baroque Corpus 2015. Hrsg. von Claudia Resch und Ulrike Czeitschner. <<http://acdh.oeaw.ac.at/abacus/>> abgerufen am [Datum des letzten Zugriffs].

Die Webapplikation, über die Abraham a Sancta Claras Werke zugänglich sind, basiert auf dem modularen Publikations-Framework [corpus\\_shell](#) und wurde in Kooperation mit Matej Ďurčo und Daniel Schopper am ACDH mit der Intention entwickelt, zeitentfernte, in Bibliotheken beherbergte Drucke aus der Barockzeit für wissenschaftliche Fragestellungen zugänglich zu machen. Es hätte viele weitere Möglichkeiten gegeben, die Daten auf andere Weise zu kodieren und zu visualisieren. Das Projektteam hat sich für diese Benutzeroberfläche entschieden, weil sie dazu geeignet ist – einem abrahamischen Buchtitel folgend – »Etwas für alle« (1699) zu bieten, indem sie sich an unterschiedlichen Fragestellungen orientiert und mehrere Nutzungsszenarien ermöglicht und unterstützt. [Abbildung 5](#) zeigt das derzeitige Erscheinungsbild und die Funktionalitäten der ABaC:us-Edition im Überblick.

---

<sup>19</sup> Auf Anfrage des Deutschen Textarchivs an der Berlin-Brandenburgischen Akademie der Wissenschaften stellt ABaC:us seine Daten zur Verfügung, um dessen automatische Taggingwerkzeuge zu optimieren.

<sup>20</sup> Auf Anfrage der Bayerischen Staatsbibliothek stellt ABaC:us seine Daten auch dafür zur Verfügung, im Rahmen des DFG-Projekts »Koordinierte Förderinitiative zur Weiterentwicklung von Verfahren für die Optical-Character-Recognition (OCR)« die Texterkennung zu verbessern.

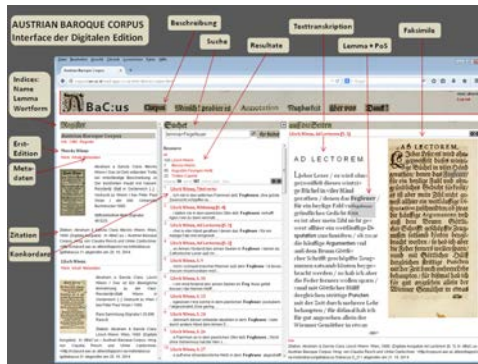


Abb. 5: Screenshot der digitalen *ABaC:us*-Edition © ACDH.

Das Projektteam konnte zwar bestehende Fragen, die an die Daten herangetragen werden, berücksichtigen, aber keine zukünftigen antizipieren, denn: »Die Fragestellungen an eine Edition liegen in der Zukunft und können nicht empirisch ermittelt werden.«<sup>21</sup> Gleichzeitig aber wurde bereits mehrfach der Wunsch geäußert, »[...] dass Editionen nach Möglichkeit für alle Disziplinen nutzbar sein sollten, dass sie allen möglichen Zwecken gleichzeitig dienen sollten.«<sup>22</sup> Diesen Anspruch vertritt auch Harvey: »a proper edition should be adequate for them all, and for many others besides.«<sup>23</sup> Wie also kann eine Edition den unterschiedlichen (und sich womöglich verändernden) Fragestellungen, die in den geisteswissenschaftlichen Disziplinen entwickelt werden, entsprechen?

Da sich die Sprachwissenschaft für andere Fragen interessiert als die Literaturwissenschaft oder andere textbasiert arbeitende Wissenschaften (Theologie, Kunstgeschichte), war den Herausgeberinnen primär daran gelegen, die Originaldokumente möglichst authentisch, quellennah und vorlagengetreu abzubilden: Die zuverlässige Wiedergabe der Texte und die Wahrung des historischen Sprachstandes waren dafür wesentliche Voraussetzungen. Die seitenweise Verknüpfung von ediertem Volltext und dem Digitalisat der Originaldrucke lässt sich im digitalen Medium problemlos herstellen und unterstützt vor allem jene Formen der Analyse und Interpretation,<sup>24</sup> die bereits bei der Betrachtung der Quelle ihren Ausgang nehmen.

Im Grunde ermöglicht die realisierte digitale Repräsentation der Drucke zwei grundverschiedene Rezeptionshaltungen: das Lesen und das Suchen in den Texten. Beide Nutzungssituationen werden durch das adaptive Layout unterstützt, das die Inhalte je nach verfügbarer Bildschirmgröße entweder automatisch anordnet oder durch die Wahl von ein, zwei oder drei Spalten festgelegt werden kann. Den Lesenden bietet es einerseits den Komfort

<sup>21</sup> Sahle 2013, S. 129.

<sup>22</sup> Sahle 2013, S. 133.

<sup>23</sup> Harvey 2001, S. 11.

<sup>24</sup> Da in dieser Art von Literatur Texten und Kupferstiche beziehungsweise Vignetten eine enge Verbindung eingehen und die Gesamtbotschaft nur durch das Zusammenwirken beider semiotischer Partner verständlich wird, dürfen die Bild-Digitalisate der Originale, auf denen Vergänglichkeitsymbole und Totentanzmotive zu sehen sind, künftigen Nutzerinnen und Nutzern jedenfalls nicht vorenthalten werden.

einer digitalen Lektüreansicht, den Suchenden stellt es andererseits Navigationsinstrumente zur Verfügung, die den direkten Zugriff auf bestimmte Stellen im Text erlauben. Während Lesende aus der Übersicht ein Werk, ein Kapitel oder eine Seite wählen, um dann in der synoptischen Ansicht den elektronischen Lesetext mit der digitalen Druckvorlage vor Augen zu haben und mit der Blätterfunktion im Text vorankommen, gelangen Suchende über verschiedene Register zu ihren Fundstellen: etwa über die alphabetisch sortierten und mit Blätterfunktion ausgestatteten Register der Lemmata und der Wortformen oder über das Register der Wortarten, das zusätzlich zu den im *STTS* enthaltenen Kategorien die zuvor beschriebenen kombinierten Formen ausweist (Abbildung 6).

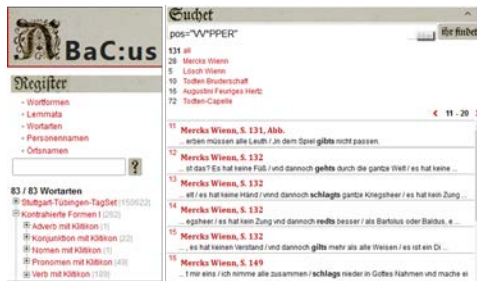


Abb. 6: LINKS: Register, insbesondere Register der Wortarten mit *STTS* und Ergänzungen © ACDH. RECHTS: Trefferliste der Kategorie »Verb mit Klitikon« mit Belegstellen. © ACDH.

Jeder Registereintrag ist mit Frequenzangaben versehen; beim Klicken auf einen Eintrag wird automatisch in der Suchleiste eine Abfrage generiert, die zu den gewünschten Ergebnissen führt. Eine zweite Möglichkeit der Texterschließung ist die freie Suche über die Suchleiste. Sie erlaubt einerseits die Volltext-Suche nach einzelnen Worten, Phrasen bzw. Kombinationen von Wörtern, andererseits eine Index-Suche, bei der spezifisch nach Wortform, Lemma, Wortart, Lemma und Wort kombiniert, Wortteilen, Personennamen oder Ortsnamen gesucht werden kann. Beginnt man mit der Eingabe, wird automatisch eine Liste von möglichen Abfragen vorgeschlagen. Diese Liste ist mit Frequenzangaben versehen und typisiert, d.h. es ist zugleich ersichtlich, ob es sich etwa um einen Personennamen, ein Lemma oder eine Wortform handelt (Abbildung 7).

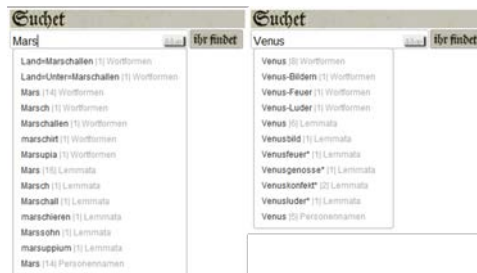


Abb. 7: LINKS: Vorgeschlagene Liste an Abfragen beginnend mit »Mars« © ACDH. RECHTS: Vorgeschlagene Liste an Abfragen beginnend mit »Venus«. © ACDH.

Die Suchergebnisse werden chronologisch nach Werktitel und mit Frequenzangaben gelistet. Durch den Klick auf eines der Werke lassen sich die Ergebnisse bei Bedarf weiter filtern. Unter der Übersicht werden die Einzelergebnisse in chronologischer Reihenfolge ihres Vorkommens innerhalb der Texte in der sogenannten *keyword in context*-Ansicht dargestellt. Das gesuchte Schlüsselwort ist dabei fett hervorgehoben und wird zentriert in seiner jeweiligen Textumgebung angezeigt, sodass sich ein erster Sinnzusammenhang erschließt. Die Fundstelle jedes Belegs wird mit Werkkurztitel und Seitenzahl unmittelbar darüber angegeben; klickt man auf diese rot markierte Zeile, öffnet sich auch die angegebene Textseite in der synoptischen Volltext- und Faksimileansicht, wo das gesuchte Wort grau unterlegt in seinem größeren Seitenkontext erscheint (vgl. Abbildung 8).



Abb. 8: Kombinierte Abfrage aller Nomen mit dem Suffix »-heit« mit *keyword in context*-Ansicht (samt Frequenzangaben), rechts einer der Belege (grau unterlegt mit Lemma und PoS) © ACDH.

## 5. Fazit und Ausblick

In den letzten drei Jahren hat sich das *ABaC.us*-Team ein profundes Wissen über die Literatur und Sprache dieser Zeitperiode angeeignet und kann deren periodenspezifische Herausforderungen durch diese Vorarbeiten realistisch einschätzen. Die Erstellung der Ressource selbst wird mittelfristig dazu beitragen, weitere Texte dieser Zeit effizienter und schneller digital zu erschließen. Vor diesem Hintergrund wäre es sinnvoll, diese erste Version der *ABaC.us*-Edition als vorläufiges Etappenziel zu sehen. Eine Erweiterung wäre in mehrere Richtungen denkbar: Solange es keine verlässliche Ausgabe zu Abraham a Sancta Clara's Schriften gibt, könnten weitere integriert werden (z.B. um das Frühwerk zu vervollständigen), auch ließe sich natürlich das *memento mori*-Genre durch weitere wichtige Vertreter ergänzen, etwa durch Autoren wie Johannes Khuen, Jakob Balde, Florentius Schilling. Gleichzeitig könnte auch das Wissen, mit dem die Daten angereichert wurden, näher spezifiziert werden (etwa durch die Annotation von Bibelstellenangaben oder durch eine granulare morphologische Annotation). Dass sich aus dieser Edition ein größeres Ganzes formen kann, indem noch weitere Ressourcen zusammengeführt und gleichermaßen annotiert werden, wäre zu wünschen, wird jedoch nicht zuletzt von den Mitteln abhängen, die dem Projekt in Zukunft zuerkannt werden.

Anne Baillot und Markus Schnöpf sehen wissenschaftliche digitale Editionen grundsätzlich »immer noch in den Kinderschuhen«<sup>25</sup> und weisen damit darauf hin, dass die Möglichkeiten digitaler Editionen noch bei weitem nicht ausgeschöpft werden. In diesem Zusammenhang wird zu beobachten sein, ob und in wie weit die in *ABaC:us* verwirklichten methodischen und technischen Ziele auch die Ansprüche anderer Projekte beeinflussen werden beziehungsweise besteht natürlich auch für das *ABaC:us*-Team selbst die Chance, seine Arbeit an barocken Texten als einen »Prozess fortschreitender Erschließung und Verarbeitung«<sup>26</sup> zu begreifen und gemeinsam mit anderen innovativen Vorhaben zu verbessern und weiterzuentwickeln.

Inzwischen bleibt zu hoffen, dass die digital verfügbaren Schriften Abraham a Sancta Claras in Forschung, Lehre und Unterricht rezipiert, erprobt und beforscht werden, und dass nicht nur in der historischen Sprachforschung und in der Literaturwissenschaft Fragen gestellt werden, die sich mit dieser Edition beantworten lassen, sondern auch die positiven Rückmeldungen aus der Abraham-Forschung, der Totentanz-Forschung, der Kunstgeschichte und der Religionswissenschaft als Zeichen des Interesses gewertet werden dürfen. Die Reaktionen auf *ABaC:us* in sozialen Medien wie Twitter (Abbildung 9) bestätigen die Relevanz der Ressource für die in unterschiedlichen Disziplinen arbeitende Gemeinschaft von Forschenden.



Abb. 9: Tweets als erste Reaktionen auf die *ABaC:us*-Edition, Eigene Darstellung © ACDH.

Wenn die zeitentfernten, aber nach wie vor faszinierenden Texte dann noch auf die Neugier einer interessierten, web-affinen Öffentlichkeit trafen, die Abraham a Sancta Clara heute noch kennt und seine Gedanken über den Tod nicht nur cursorisch oder anekdotisch nachlesen, sondern sie im unveränderten Wortlaut und unter Angabe von Belegstellen finden möchte, dann wäre das abrahamische »Etwas für alle« mehr als 300 Jahre nach seinem Tod tatsächlich erneu(er)t glücklich.

<sup>25</sup> Vgl. Baillot / Schnöpf 2015, S. 139.

<sup>26</sup> Sahle 2013, S. 137.

## Bibliographische Angaben

- Anne Baillot / Markus Schnöpf: Von wissenschaftlichen Editionen als interoperable Projekte, oder: Was können eigentlich digitale Editionen? [\[online\]](#) In: Digital Humanities. Praktiken der Digitalisierung, der Dissemination und der Selbstreflexivität. Hg. von Wolfgang Schmale. Stuttgart 2015, S. 139–156. [\[Nachweis im GBV\]](#)
- Wilhelm Friedrich Bautz: Abraham a Sancta Clara. In: Biographisch-Bibliographisches Kirchenlexikon. Band I (1990), Sp. 10–11. [\[Nachweis im GBV\]](#)
- Georg Braungart: Artikel Barock. In: Literaturwissenschaftliches Lexikon. Grundbegriffe der Germanistik. 2. Auflage. Berlin 2006, S. 44–50. [\[Nachweis im GBV\]](#)
- Stefanie Dipper: POS#Tagging of Historical Language Data: First Experiments. In: Semantic Approaches in Natural Language Processing. Proceedings of the 10 Conference on Natural Language Processing. Hg. von Manfred Pinkal / Ines Rehbein / Sabine Schulte im Walde / Angelika Storrer. Saarbrücken 2010, S. 117–121. [\[Nachweis im GBV\]](#)
- Franz M. Eybl: Abraham a Sancta Clara. Vom Prediger zum Schriftsteller. Tübingen 1992. [\[online\]](#) [\[Nachweis im GBV\]](#)
- Franz M. Eybl: Abraham a Sancta Clara. In: Killy Literaturlexikon Volume 1 (2008), S. 10–14. [\[Nachweis im GBV\]](#)
- Franz M. Eybl: Wissenslücken um Abraham a Sancta Clara. Zur Problematik populärer Autorschaft. [\[Nachweis im GBV\]](#) In: Abraham a Sancta Clara. Vom barocken Kanzelstar zum populären Schriftsteller. Hg. von Anton Philipp Knittel. Eggingen 2012, S. 104–121. [\[Nachweis im GBV\]](#)
- P. D. A. Harvey: Editing Historical Records. London 2001. [\[Nachweis im GBV\]](#)
- Erhard Hinrichs / Thomas Zastrow: Linguistic Annotations for a Diachronic Corpus of German. In: Linguistic Issues in Language Technology 7 (2012), S. 1–16. [\[online\]](#)
- Volker Meid: Die deutsche Literatur im Zeitalter des Barock. Vom Späthumanismus zur Frühaufklärung. München 2009. [\[Nachweis im GBV\]](#)
- Dirk Niefanger: Barock. 3. Auflage. Stuttgart, Weimar 2012. [\[Nachweis im GBV\]](#)
- Wulf Oesterreicher: Korpuslinguistik und diachronische Lexikologie. Fallbeispiele aus dem amerikanischen Spanisch des 16. Jahrhunderts. In: Lexikalische Semantik und Korpuslinguistik. Hg. von Wolf Dietrich / Ulrich Hoinkes / Bärbara Roviró / Matthias Warnecke. Tübingen 2006, S. 479–493. [\[Nachweis im GBV\]](#)
- Claudia Resch / Thierry Declerck / Barbara Krautgartner / Ulrike Czeitschner: ABaC:us revisited – Extracting and Linking Lexical Data from a historical Corpus of Sacred Literature. In: Proceedings of the 2 Workshop on Language Resources and Evaluation for Religious Texts. Hg. von Eric Atwell / Claire Brierley / Majdi Sawalha. Reykjavik 2014, S. 36–41. [\[Nachweis im GBV\]](#)
- Claudia Resch / Wolfgang U. Dressler: Zur Pragmatik der Diminutive in frühen Erbauungstexten Abraham a Santa Claras. Eine korpusbasierte Studie. In: Linguistische Pragmatik in historischen Bezügen. Hg. von Peter Ernst / Martina Werner. Berlin 2016, S. 235–249. [\[Nachweis im GBV\]](#)
- Claudia Resch / Ulrike Czeitschner / Eva Wohlfarter / Barbara Krautgartner: Introducing the Austrian Baroque Corpus: Annotation and Application of a Thematic Research Collection. In: Proceedings of the Third Conference on Digital Humanities in Luxembourg with a Special Focus on Reading Historical Sources in the Digital Age Luxembourg, Luxembourg, December 5-6, 2013. Hg. von Lars Wieneke / Catherine Jones / Marten Düring / Florentina Armaselu / René Leboutte. 2016 [\[online\]](#)
- Patrick Sahle: Digitale Editionsformen, Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels. 3 Bände. Norderstedt 2013. [\[online\]](#) [\[Nachweis im GBV\]](#)

## Weiterführende Literatur

- Ulrike Czeitschner / Thierry Declerck / Karlheinz Mörth / Claudia Resch: Linguistic and Semantic Annotation in Religious Memento Mori Literature. In: Proceedings of the LREC 2012 Workshop: Language Resources and Evaluation for Religious Texts. Paris 2012, S. 49–52. [\[online\]](#)
- Martin Durrell: »Representativeness«, »Data«, and legitimate expectations. What can an electronic historical corpus tell us that we didn't actually know already (and how)? In: Historical Corpora. Challenges and Perspectives. Hg. von Jost Gippert / Ralf Gehrke. Tübingen 2015, S. 13–33. [\[Nachweis im GBV\]](#)
- Claudia Resch / Ulrike Czeitschner: Repräsentation von | in barocken Buch-Totentänzen im digitalen Medium. In: Repräsentation(en). Interdisziplinäre Annäherungen an einen umstrittenen Begriff. Hg. von Gernot Gruber / Monika Mokre. Wien 2016, S. 35–41. [\[Nachweis im GBV\]](#)

## Abbildungsnachweise und -legenden

Abb. 1: LINKS: Portrait Abraham a Sancta Claras, das einem seiner Nachrufe vorangestellt wurde. Quelle: Johann Carl Megerle, ABRAHAM ist gestorben. Wien [1710]. Österreichische Nationalbibliothek Wien, Signatur 220677-C. RECHTS: Titelblätter und Kupferstiche der in der *ABaC:us*-Edition enthaltenen Schriften. Eigene Darstellung © ACDH.

Abb. 2: Kommentierung offensichtlicher Fehler. © ACDH.

Abb. 3: Kommentierung offensichtlicher Fehler. © ACDH.

Abb. 4: Annotierung von abrahamischen Wortspielen – eine Wortform, zwei Lesarten © ACDH.

Abb. 5: Screenshot der digitalen *ABaC:us*-Edition © ACDH.

Abb. 6: LINKS: Register, insbesondere Register der Wortarten mit *STTS* und Ergänzungen © ACDH. RECHTS: Trefferliste der Kategorie »Verb mit Klitikon« mit Belegstellen. © ACDH.

Abb. 7: LINKS: Vorgeschlagene Liste an Abfragen beginnend mit »Mars« © ACDH. RECHTS: Vorgeschlagene Liste an Abfragen beginnend mit »Venus«. © ACDH.

Abb. 8: Kombinierte Abfrage aller Nomen mit dem Suffix »-heit« mit *keyword in context*-Ansicht (samt Frequenzangaben), rechts einer der Belege (grau unterlegt mit Lemma und PoS) © ACDH.

Abb. 9: Tweets als erste Reaktionen auf die *ABaC:us*-Edition, Eigene Darstellung © ACDH.