

Der falsche Quijote? Autorschaftsattribute für spanische Prosa der frühen Neuzeit.

,
nanette.rissler-pipka@gmx.de
Universität Siegen, Deutschland

Hintergrund

Das bis heute bekannteste Werk der spanischen Literatur ist weltweit als „Don Quijote“ von Miguel de Cervantes geläufig. Dass Cervantes aber, wie kurz zuvor Mateo Alemán, mit einem Fälscher zu kämpfen hatte, der den ersten Band des „Quijote“ ungefragt weiter dichtete und ein Jahr vor der eigenen Fortsetzung durch Cervantes einen zweiten Band aus eigener Feder unter dem Namen Alonso Fernández de Avellaneda heraus brachte, wissen die wenigsten Leser des „Quijote“. Es ist im strengen Sinne des Wortes auch keine Fälschung oder ein Plagiat, sondern die freie Fortsetzung eines erfolgreichen Romans unter eigenem Namen bzw. in diesem Fall unter Pseudonym. Die Identität Avellanedas ist bis heute unbekannt.

Fragestellungen / Stand der Forschung

Angesichts der historischen Publikationsbedingungen in Spanien der frühen Neuzeit (Chartier 2006), stellt sich die Frage, ob eine Autorschaftsattribute mithilfe stilometrischer Methoden auf der Grundlage aktuell zugänglicher digitalisierter Buchausgaben überhaupt möglich ist. Doch auch Patrick Juola und Christopher Coufal nahmen 2010 die Ausgabe des „Don Quijote“, die im Project Gutenberg zugänglich ist als Grundlage ihrer Analyse, die als Ergebnis hatte, dass Cervantes nicht der Autor der letzten 69 (von 74) Kapitel des 2. Bandes des „Don Quijote“ sei (Coufal / Juola 2010). Aus dieser provokanten These entwickelte sich aber keine wissenschaftliche Debatte innerhalb der internationalen Hispanistik und auch die DH-Experten Juola und Coufal erweiterten ihre Fragestellung nicht auf die sich unmittelbar anschließende Frage, wer der Autor des „falschen“ Quijotes von Avellaneda sei. Auch eine weitere statistische Untersuchung der beiden Teile des „Quijote“ von Cervantes berücksichtigt nicht Avellaneda (López Quintero 2011). Die Tatsache, dass sich der Stil Cervantes' innerhalb des zweiten Teils des „Quijote“ ändert, ist in der Hispanistik anerkannt und wird im Allgemeinen mit dem

Erscheinen der apokryphen Fortsetzung durch Avellaneda in kausalen Zusammenhang gebracht (Strosetzki 1991: 93; Ehrlicher 2008: 42ff.; Blasco 2007: XVII; Gómez Canseco 2008). Bislang wurde jedoch der Stil Cervantes' gerade im Vergleich zu seinem Nachahmer Avellaneda zumeist als in jeder Hinsicht überragend dargestellt (vgl. zu einer kompakten Darstellung dieser Missachtung Avellanedas: Alvarez Roblin 2014). Erst durch die jüngste Reihe von neuen Ausgaben der Avellaneda-Fortsetzung wird dessen literarische Leistung in der Fachwelt anerkannt (Gómez Canseco 2014; Alvarez Roblin 2009; Suárez Figaredo 2014; López-Vázquez 2011). Mit dem wachsenden Interesse an Avellaneda nimmt auch die Suche nach dessen Identität mithilfe digitalisierter Korpora zu. Zum größten Teil stützen sich diese Autorschaftsattribute auf schlichte Recherchemöglichkeiten von CORDE (Corpus diacrónico del español), CREA (Corpus de Referencia del Español Actual) und GoogleBooks (Suárez Figaredo 2011; Madrigal 2009; López-Vázquez 2011; Blasco 2005; Jiménez 2007) oder auf philologisch-historische Recherchen (Cruz Casado 2008; Sánchez Portero 2006). Insgesamt werden im Laufe der Recherche nach der wahren Identität Avellanedas 39 Namen ins Spiel gebracht. Von 18 dieser Kandidaten liegen digitalisierte Texte frei zugänglich vor. Dennoch nutzt keiner der Autorschaftsdetektive aktuelle Methoden der DH zur Autorschaftsattribute (wie z. B. JGAAP oder Stilometrie mit R; vgl. Juola 2012; Eder 2015). Neben der Autorschaftsattribute bzgl. Avellanedas apokrypher Fortsetzung des „Quijote“, stellt die Hauptfrage dieses Papers, die stilistische und stilometrische Unterscheidung zwischen Cervantes und Avellaneda dar. Es kann gezeigt werden, dass nur mithilfe stilometrischer Methoden, die festgefahrene Fachdiskussion neue Perspektiven und unerwartete Ergebnisse erhält.

Die Methode

Mithilfe des stylo-Pakets für R (Eder / Rybicki 2011), das für spanischsprachige Texte noch vergleichsweise wenig getestet wurde, sollen zum einen die vorliegenden Autorschaftsattribute falsifiziert und die Fragen nach stilistischer Nähe zwischen Cervantes, Avellaneda und anderen zeitgenössischen Autoren spanischer Prosa geklärt werden. Somit wird auch die Methode hinsichtlich ihrer Kompatibilität mit spanischsprachigem Korpus überprüft. Dazu konnte die neueste Version des stylo-package (0.6.0) und die von Jannidis et al. (2015) vorgestellte Cosine Distance genutzt werden.

Zu diesem Zweck wurde zunächst ein passendes Korpus erstellt, das repräsentativ für die Zeit von 1585-1630 spanische Prosawerke enthält und sich auch aus den Kandidaten für die Autorschaft des apokryphen „Quijote“ zusammensetzt. Die Texte stammen aus digitalen Editionen von cervantesvirtual, Wikisource und Project Gutenberg. Sie wurden einheitlich in plain-text-Format abgespeichert und von Textteilen, die nicht von selben Autor stammen

(wie z. B. einleitende Bemerkungen des Herausgebers) befreit.

Um zunächst ein sicheres Set zu haben, das sowohl die Autorschaft als auch Genre und Epoche betreffend vergleichbar ist, wurden zunächst nur 4 Autoren mit unterschiedlich vielen Texten ausgewählt (insgesamt 32 Texte). Eine Vergleichbarkeit die Textlänge betreffend hätte bedeutet entweder nur die Novellen oder nur die Romane miteinander zu vergleichen. Da es auch mit unterschiedlicher Textlänge zu guten Ergebnissen kam, wurde auf diese Angleichung verzichtet (vgl. Eder 2010). Das Korpus wurde mit zwei verschiedenen und anerkannten Distanzmaßen (Eder's Delta und Cosine) und 100-5000 MFW als Cluster-Analyse ausgewertet. Schrittweise wurden dann weitere Kandidaten hinzugefügt und die Ergebnisse bewertet. Die Problematik, die sich dabei ergab, war, dass es wenig Sinn macht, Autoren mit ins Korpus zu nehmen, von denen nur ein Textbeispiel vorliegt, da diese keinem „Partner“ im Dendrogramm zugeordnet werden können und somit fälschlicherweise eigentlich weiter voneinander entfernte Texte zusammen geclustert dargestellt werden. Um dieses Problem der Cluster-Analyse zu umgehen, wurde im Vergleich eine Principle Component Analysis (PCA) durchgeführt, die eine bessere Darstellung der Distanzen zwischen den einzelnen Texten im Raum zeigt.

Die Ergebnisse

Im ersten Korpus mit 32 Texten funktioniert die Zuordnung sehr gut. Die Cervantes-Texte sind trotz ihrer starken Größenunterschiede (Novellen mit ca. 7000 und Romane mit ca. 200.000 Wörtern) klar zusammen geclustert.

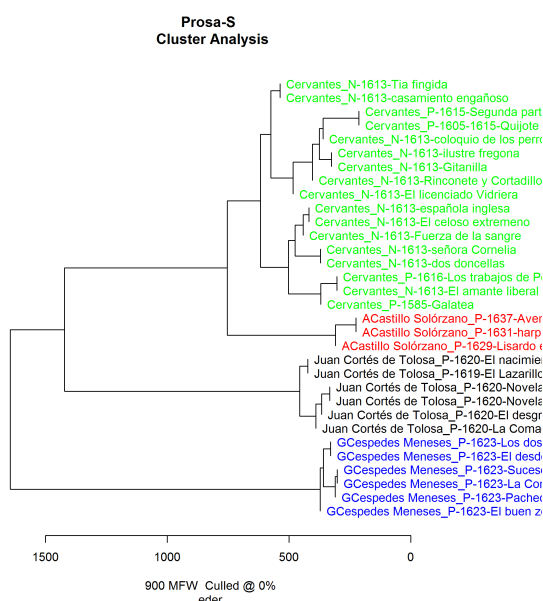


Abb.1: Cluster Analysis mit R (stylo), 900 MFW, Eder's Delta

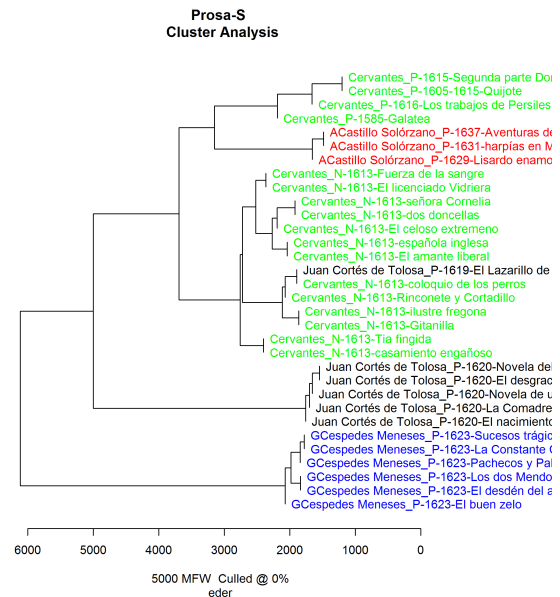


Abb. 2: Cluster Analysis mit R (stylo), 5000 MFW, Eder's Delta

Ab 900 MFW und darunter ist die Autorschaftszuordnung mit Eder's Delta einwandfrei, bei darüber liegenden Zahlen schlich sich beständig der Nachahmer-, „Lazarillo“ von Cortés de Tolosa und auch der Block der Werke von Castillo Slórzano zwischen die Cervantes-Werke. Zum Vergleich wurde der gleiche Versuch mit Cosine Distance durchgeführt und brachte keine größere Veränderung in der Darstellung. Weiterhin blieb der Roman von Cortés de Tolosa hartnäckig bis zum 300 MFW unter den Cervantes-Werken, jedoch konnte Castillo Solórzano früher heraus genommen werden.

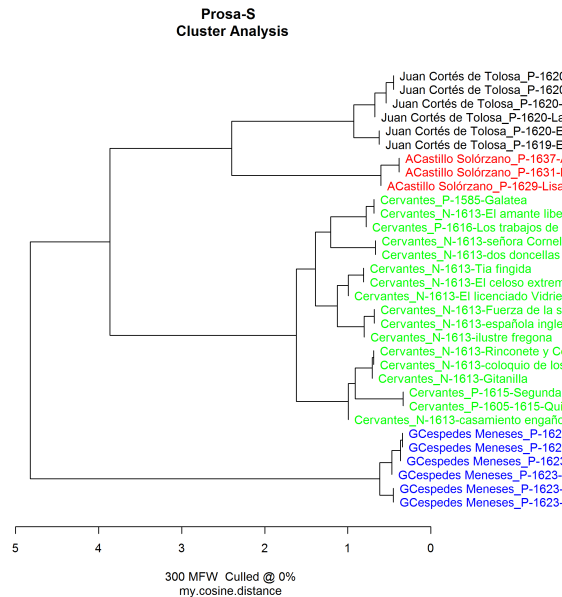


Abb. 3: Cluster Analysis mit R (stylo), 300 MFW, Cosine Distance

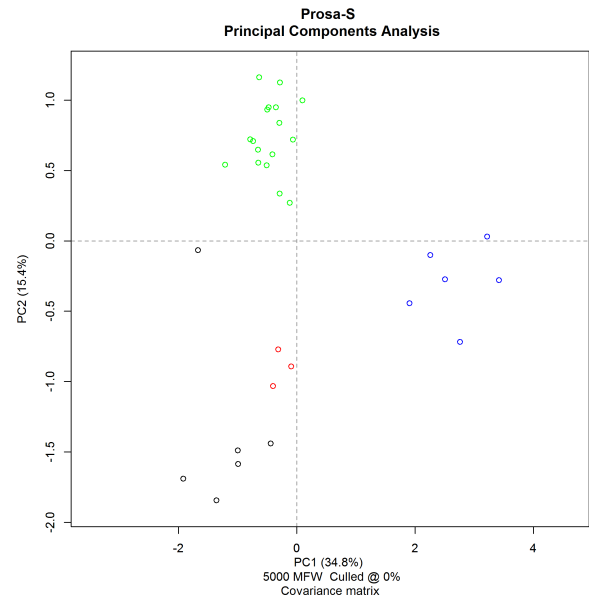


Abb. 5: PCA mit R (stylo), 5000 MFW, Cosine Distance

Deutlich wird vor allem, dass der „Lazarillo“, den Cortés Tolosa in Nachahmung des Originals (anonym 1554) geschrieben hat, sehr von seinem übrigen Werk abweicht, aber ebenso noch deutlich von den Cervantes-Werken entfernt ist. Das erklärt sich allerdings leicht durch die sprachliche Unterscheidung der vorliegenden Edition des Werkes, die eine ältere Form des Kastilischen verwendet und daher in der Vergleichbarkeit eingeschränkt bleibt.

Spannender werden die Ergebnisse, wenn man Avellaneda und weitere Kandidaten hinzunimmt:

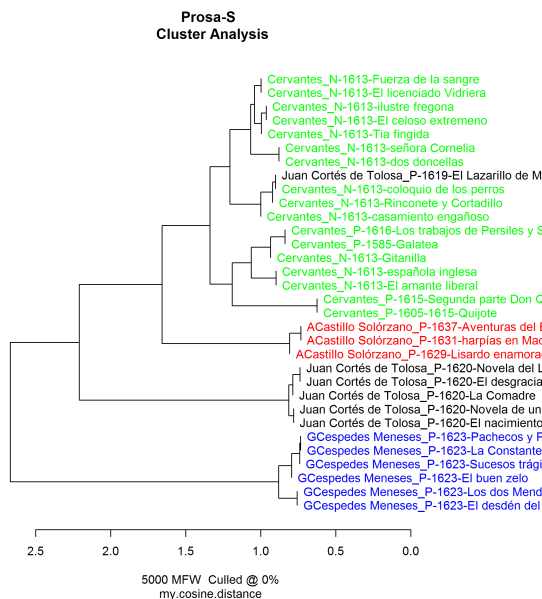


Abb. 4: Cluster Analysis mit R (stylo), 5000 MFW, Cosine Distance

Interessant ist hier, dass die PCA mit denselben Variablen zeigt, wie weit entfernt der „Lazarillo“ von Cortés de Tolosa doch von den übrigen Werken Cervantes‘ entfernt ist (und zwar bei allen 100-5000 MFW):

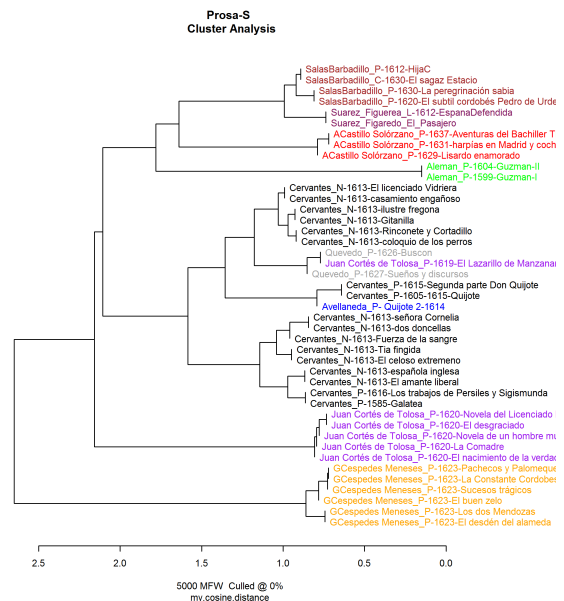


Abb. 6: Cluster Analysis mit R (stylo), 5000 MFW, Cosine Distance (Korpuserweiterung: 43 Texte)

In einer nächsten Korpuserweiterung nehmen wir weitere Kandidaten mit Prosabeispielen hinzu, von denen nur ein Text zur Verfügung steht.

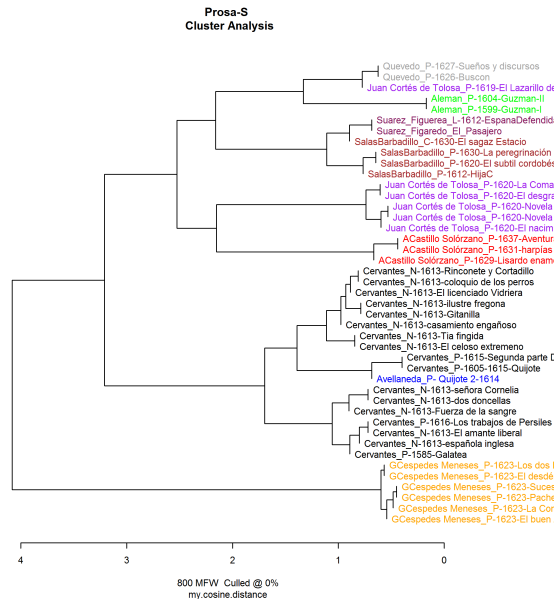


Abb. 7: Cluster Analysis mit R (stylo), 800 MFW, Cosine Distance

Auch mit der Korpuserweiterung bleiben die Zuordnungen relativ stabil (außer Cortés de Tolosa, wie zuvor). Keiner der möglichen Kandidaten wird jedoch Avellaneda zugeordnet, sondern der apokryphe „Quijote“ wird mit den anderen beiden Teilen von Cervantes zusammen geclustert – dies bleibt über 100-5000 MFW konstant. Auch mit dem einer PCA bestätigt sich dieses Ergebnis (vgl. Fig. 8: der blaue Punkt bei den schwarzen zeigt Avellaneda im Umkreis der Cervantes-Werke):

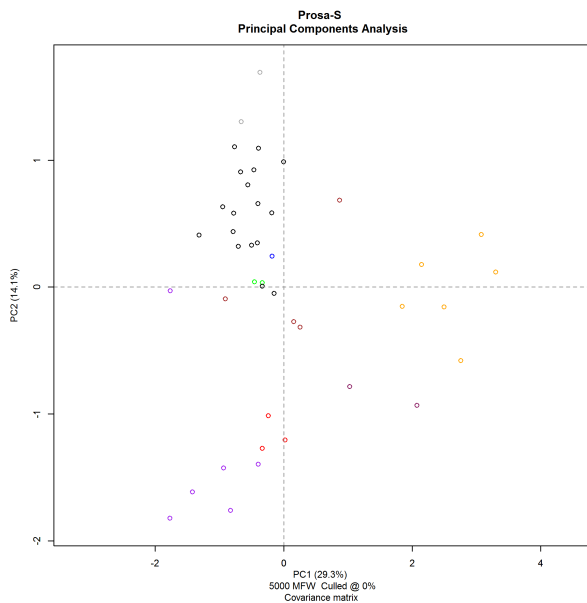


Abb. 8: PCA mit R (stylo), 5000 MFW, Cosine Distance

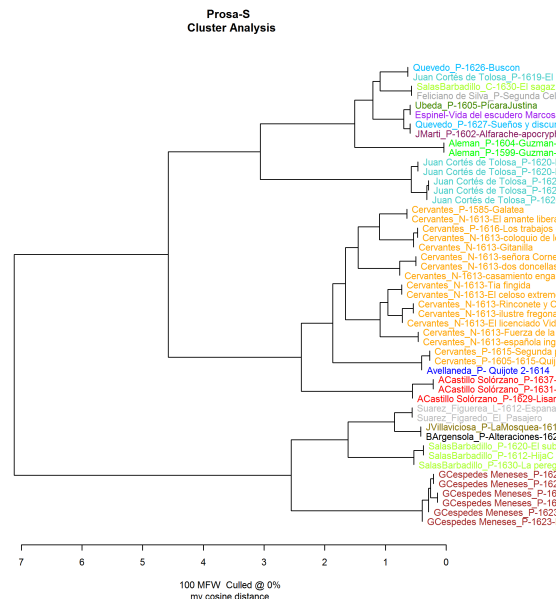


Abb. 9: Cluster Analysis mit R (stylo), 100 MFW, Cosine Distance (Korpuserweiterung: 49 Texte)

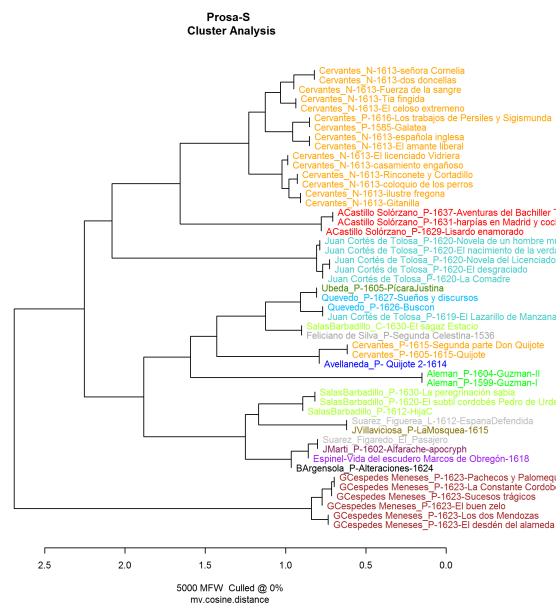


Abb. 10: Cluster Analysis mit R (stylo), 5000 MFW, Cosine Distance

Es wird schnell deutlich, dass die Einzelgänger unter den Texten, das zuvor stabile Gefüge auseinanderbringen und offensichtlich nicht für die statistische Untersuchung zu gebrauchen sind. Erstaunlich bleibt jedoch, dass selbst in diesem sehr vagen Clustering (je nach MFW schieben sich die einzelnen Texte mal dort mal dort hin), die drei

„Quijote“-Texte konsistent bei 100-5000 MFW ein Cluster bilden. Kein anderer Kandidat schafft es in die Nähe von Avellaneda.

Auch die PCA bringt keine entscheidenden Vorteile gegenüber der Cluster Analysis. Espinel und Úbeda gruppieren sich zwar in die Cervantes-Gruppe, aber nicht direkt zu Avellaneda.

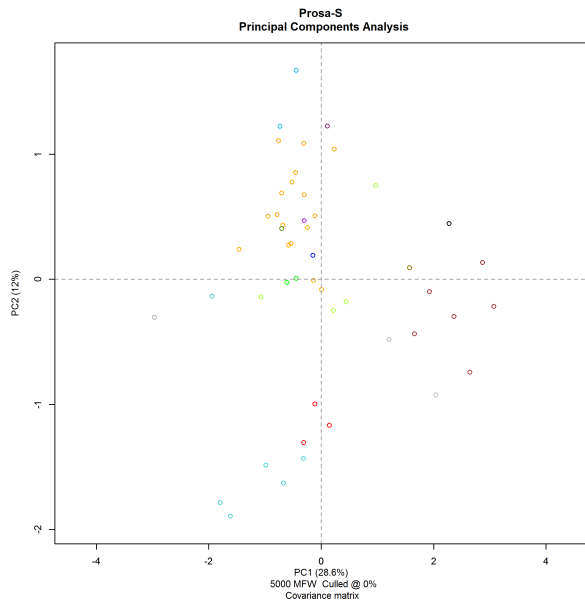


Abb. 11: PCA mit R (stylo), 5000 MFW, Cosine Distance

Schlussfolgerungen

Wenn überhaupt ein Autornamen anstelle desjenigen Avellanedas genannt werden sollte, müsste es nach den vorliegenden Ergebnissen Cervantes selbst sein, der in einem gekonnten Spaß ganz im Sinne seines Quijote die Leser mit der eigenen falschen Fortsetzung an der Nase herum führt. Oder aber beide Autoren haben ihren Stil gegenseitig aufeinander abgestimmt, dass sie sich derart im Wortgebrauch ähneln. Es würde sich als folgende Untersuchung ein rolling delta anbieten, um eine kollaborative Autorschaft ausmachen zu können und um den vorgeblichen Stilwechsel im zweiten Teil von Cervantes' „Quijote“ genauer definieren und mit Avellaneda in Zusammenhang bringen zu können.

Bibliographie

Alvarez Roblin, David (2014): *De l'imposture à la création. Le Guzmán et le Quichote apocryphes*. Madrid : Casa de Velázquez.

Avellaneda, Alonso Fernández de (2014): *Segundo tomo del ingenioso hidalgo don Quijote de la Mancha*. Edición de Gómez Canseco, Luis. Madrid: Real Acad. Española, Centro para la Ed. de los Clásicos Españoles.

Avellaneda, Alonso Fernández de (2011): *El Quijote apócrifo*. Edición de López-Vázquez, Alfredo Rodríguez. Madrid: Cátedra.

Avellaneda, Alonso Fernández de (2014): *El Quijote apócrifo* (= Lemir 18: Conmemoración iv Centenario del Quijote de Avellaneda) http://parnaseo.uv.es/Lemir/Revista/Revista18/Textos/06_Quijote_Avellaneda_Figaredo.pdf [letzter Zugriff 15. Oktober 2015].

Blasco, Javier (2005): “La lengua de Avellaneda en el espejo de ‘La pícara Justina’”, in: *Boletín de la Real Academia Española* 85, 291-292: 53-109 <http://uvadoc.uva.es/bitstream/10324/2436/1/LA%20LENGUA%20DE%20AVELLANEDA...%28Javier%20Blasco%29.pdfOCR.pdf> [letzter Zugriff 15. Oktober 2015].

Chartier, Roger (2006): “Materialidad del texto, textualidad del libro”, in: *Orbis Tertius* 11, 12: <http://www.orbistertius.unlp.edu.ar/article/view/OTv11n12a01/3774> [letzter Zugriff 8. Januar 2016].

Coufal, Christopher / Juola, Patrick (2010): “Authorship Discontinuities of El Ingenioso Hidalgo don Quijote de la Mancha as detected by Mixture-of-Experts”, in: *Digital Humanities 2010 Conference Abstracts*, London <http://dh2010.cch.kcl.ac.uk/academic-programme/abstracts/papers/html/ab-755.html> [letzter Zugriff 15. Oktober 2015].

Cruz Casado, Antonio (2008): “Revisión de una hipótesis. Juan Valladares de Valdelomar; autor del ‘Quijote’”, in: Dotras Bravo, Alexia (ed.): *Tus obras los rincones de la tierra descubren*. Actas del VI congreso internacional de la Asociación de Cervantistas, Alcalá de Henares 29-44 http://cvc.cervantes.es/literatura/cervantistas/congresos/cg_VI/cg_VI_18.pdf [letzter Zugriff 15. Oktober 2015].

Eder, Maciej (2010): “Does size matter? Authorship attribution, short samples, big problem”, in: *Digital Humanities 2010. Conference Abstracts*, London: 132-35.

Eder, Maciej / Kestemont, Mark / Rybicki, Jan (2013): “Stylometry with R: a suite of tools”, in: *Digital Humanities 2013. Conference Abstracts*, University of Nebraska, Lincoln 487-89.

Ehrlicher, Hanno (2008): „Der andere Autor im eigenen Werk. Medialisierte Autorschaft bei Mateo Alemán und Miguel de Cervantes“, in: Dünne, Jörg / Moser, Christian (eds.): *Automedialität*. München: Fink 27-52.

Gómez Canseco, Luis (2008): “1614: Cervantes escribe otro ‘Quijote’”, in: Dotras Bravo, Alexia (ed.): *Tus obras los rincones de la tierra descubren*. Actas del VI congreso internacional de la Asociación de Cervantistas, Alcalá de Henares 29-44 http://cvc.cervantes.es/literatura/cervantistas/congresos/cg_VI/cg_VI_05.pdf [letzter Zugriff 15. Oktober 2015].

Jannidis, Fotis / Pielström, Steffen / Schöch, Christof / Vitt, Thorsten (2015): “Improving Burrows’ Delta – An empirical evaluation of text distance measures”, in: *Digital Humanities 2015. Conference Abstracts*, Sydney

<http://dh2015.org/abstracts/index.php> [letzter Zugriff 15. Oktober 2015].

Jiménez, Alfonso Martín (2007): “Cotejo por medios informáticos de la Vida de Pasamonte y el Quijote de Avellaneda”, in: *Etiópicas* 3: 69-131 <http://rabida.uhu.es/dspace/bitstream/handle/10272/1616/b1551216.pdf?sequence=1> [letzter Zugriff 15. Oktober 2015].

López Quintero, Freddy (2011): “Diferencias internas en Don Quijote. Cambios en proporciones y cambios estructurales”, in: *Lemir* 15: 259-270 http://parnaseo.uv.es/Lemir/Revista/Revista15/13_Lopez_Freddy.pdf [letzter Zugriff 15. Oktober 2015].

López-Vázquez, Alfredo Rodríguez (2011): “Introducción”, in: Avellaneda, Alonso Fernández de: *El Quijote apócrifo*. Edición de López-Vázquez, Alfredo Rodríguez, Madrid: Cátedra 13-84.

Madrigal, José Luis (2009): „Tirso, Lope y el Quijote de Avellaneda“, in: *Lemir* 13: 191-250 http://parnaseo.uv.es/lemir/revista/revista13/11_madrigal_jose.pdf [letzter Zugriff 15. Oktober 2015].

Rybicki, Jan / Eder, Maciej (2011): “Deeper Delta across genres and languages: do we really need the most frequent words?”, in: *Literary and Linguistic Computing* 26, 3: 315-21.

Sánchez Portero, Antonio (2006): *El autor del ‘Quijote’ de Avellaneda es Pedro Liñán de Ríaza, poeta de Calatayud*. Edición digital por cortesía del autor para la Biblioteca Virtual Miguel de Cervantes, Alicante <http://www.cervantesvirtual.com/obra/el-autor-del-quiote-de-avellaneda-es-pedro-lin-de-riaza-poeta-de-calatayud-0/> [letzter Zugriff 15. Oktober 2015].

Strosetzki, Christoph (1991): „Der Roman im Siglo de Oro“, in: Strosetzki, Christoph (ed.): *Geschichte der spanischen Literatur*. Tübingen: Niemeyer 84-118.

Suárez Figaredo, Enrique (2011): “Sobre la atribución del Quijote apócrifo a José de Villaviciosa”, in: *Lemir* 15: 135-146 http://parnaseo.uv.es/Lemir/Revista/Revista15/05_Suarez_Enrique.pdf [letzter Zugriff 15. Oktober 2015].