

## Sektion 10

### *Digitale Zugänge zu historischen Textkorpora: Analysen, Praktiken, Reflexionen*

Philip Hillebrand (Universität Münster), Nanette Reißler-Pipka (Max Weber Stiftung, Bonn),  
Susanne Schlünder (Universität Osnabrück) & Christof Schöch (Universität Trier)

#### **Abstracts**

Robin-M. Aust, Marie-Christine Boucher, J. Berenike Herrmann & Kirsten Kramer (Universität Bielefeld)

#### **Ein Ansatz zur mehrsprachigen Modellierung des alpinen Erhabenen**

Das beeindruckende Ausmaß des Alpenraums erzeugte im 19. Jahrhundert eine intensive literarische Resonanz. Die „Suche nach dem Erhabenen“ führte besonders britische Reisende in die Alpenregion und zur Entstehung des Alpinismus (vgl. Bainbridge 2020). Sie schlug sich auch in Texten der Romantik nieder, in der sich die deutsche und die britische Literatur stark annähern: Die gegenseitige Beeinflussung brachte deutsche, französische und englische Texte hervor, die eine Naturästhetik des Erhabenen zu versprachlichen versuchen und in derselben Region verfasst wurden. Damit manifestiert sich das Erhabene als sprachliche Schilderung der Erfahrung komplexer Kombination von Affekten, die von konkreten Räumen verursacht oder auf sie projiziert werden.

Durch seine transnationale Ausbreitung bietet der Alpenraum eine Grundlage für eine multilingual und -kulturell ausgerichtete Fallstudie. Zugleich stellt dies Analysen vor Herausforderungen: Diese Texte befassen sich mit der „Darstellung der Grenze zum Unfaßbaren“ (Braungart und Fricke 2007, S. 492). Mehr noch als im Falle anderer Empfindungen und Affekte stellt sich hier also die Frage, wie sie sprachlich, stilistisch vermittelt wird – und wie dies anhand konkreter textueller Phänomene umrissen und analysiert werden kann.

Unser Zugriff verfährt hier methodenpragmatisch. Bestehende Methoden z. B. zur Sentimentanalyse ermöglichen zwar zuverlässig Analysen von Emotionen in Texten (vgl. Grisot und Herrmann 2023). Aktuell findet sich jedoch kein Zugriff, der kulturell kodierte Emotionskomplexe des Erhabenen noch dazu mehrsprachig analysieren könnte. Daher muss ein Analyse-Workflow von Grund auf multilingual konzipiert werden und *low resource*-Sprachen einbeziehen. Die Komplexität und Stilisierung literarischer, metaphorischer oder bildhafter Sprache erschwert eine lexikonbasierte Untersuchung des Erhabenen.

Aufbauend auf frühere digitale Analysen dieses Phänomens und einer Modellierung komplexerer, in literarischen (Belletristik) und nicht-literarischen (Memoiren, Reiseberichte) Texten des 18. und 19. Jahrhunderts enkodierten Emotionen (vgl. Herrmann und Lüdtko 2023) präsentieren wir einen Ansatz, der über Text Mining und lexikonbasierte Sentimentanalyse hinausgeht, um die folgenden Fragen zu adressieren: Wie wird das Erhabene in den Texten geschildert? Wie lassen sich dessen rhetorische, räumliche und affektive Struktur mit automatisierten und manuellen Methoden adäquat analysieren? Unterscheiden sich die Umsetzungen in unterschiedlichen Sprachen? In einem ersten Schritt hin zu einer Operationalisierung des Erhabenen für ein automatisiertes *distant reading* in Texten sollen sprachliche Merkmale in unserem mehrsprachigen Korpus durch *close reading* klassifiziert, in ihrer Vielfalt erfasst und annotiert werden. Auf dieser Basis testen wir u. a. die Verteilung individueller Emotionen und das Training eines Classifiers, die wir für die automatisierte Analyse in weiteren Projektphasen weiterverfolgen werden.

## Referenzen

- Bainbridge, S. (2020). *Mountaineering and British Romanticism: The Literary Cultures of Climbing, 1770-1836*. OUP Oxford.
- Braungart, Georg und Harald Fricke, Hrsg. (2007). *Reallexikon der deutschen Literaturwissenschaft. 2: H - O / gemeinsam mit Georg Braungart, hrsg. von Harald Fricke*. Berlin New York: de Gruyter.
- Grisot, Giulia und Berenike Herrmann (2023). "Examining the Representation of Landscape and Its Emotional Value in German-Swiss Fiction between 1840 and 1940". In: *Journal of Cultural Analytics* 8.1. doi: 10.22148/001c.84475.
- Herrmann, Berenike und Jana Lüdtker (2023). "A Fairy Tale Gold Standard. "Annotation and Analysis of Emotions in the Children's and Household Tales by the Brothers Grimm". In: doi: 10.17175/2023\_005.

Laura Hernández-Lorenzo (Universidad de Sevilla)

## **Secuencias morfosintácticas en la poesía femenina del Siglo de Oro español: análisis estilístico y estilométrico**

La emergencia de las Humanidades Digitales (Schreibman et al., 2016) ha supuesto la aparición de nuevas metodologías y perspectivas en el estudio de la Literatura. Dentro de estas, destacan los enfoques cuantitativos, que se han visto reforzados por las teorías de análisis distante (Moretti, 2013) y macroanálisis (Jockers, 2013). A pesar de que en los últimos años han aumentado los trabajos que analizan cuantitativamente los textos de escritores españoles, con nuevas aportaciones en cuestiones de autoría (García-Reidy, 2019; Cuéllar & Vega García-Luengos, 2023), movimientos literarios (Hernández-Lorenzo, 2022) o géneros textuales (Calvo Tello, 2019), se ha dedicado una menor atención a las autoras, que además se encuentran infrarrepresentadas en los corpus utilizados debido a varias causas: fundamentalmente, su menor inclusión en el canon, las dificultades para recopilar sus textos y la escasa digitalización de los mismos.

Aunque destacan las investigaciones llevadas a cabo en el proyecto Bibliografía de escritoras españolas (Bieses), dedicadas al estudio de la obra de escritoras de la primera Edad Moderna, y que han dado como fruto una base de datos de las escritoras españolas, así como a estudios especializados en este ámbito (Baranda & Cruz, 2018), en general en el campo de la poesía áurea contamos con menos trabajos que analicen la participación que pudieron tener estas autoras en la evolución del lenguaje poético. En esta ponencia se presenta un trabajo en progreso cuyo objetivo es analizar los rasgos morfosintácticos de la poesía de las escritoras más relevantes del Siglo de Oro y cómo reflejan la evolución del estilo poético en este periodo. Con este fin se combinará el enfoque cuantitativo que nos ofrece la estilometría con los métodos de estilística de corpus, los cuales nos permitirán establecer un puente con el análisis cualitativo de las principales secuencias morfosintácticas.

### **Referencias**

- Baranda, N., & Cruz, A. J. (2018). *The Routledge Research Companion to Early Modern Spanish Women Writers*. Routledge.
- Calvo Tello, J. (2019). Delta Inside Valle-Inclán: Stylometric Classification of Periods and Groups of His Novels. *Romanische Studien. Theorien von Autorschaft und Stil in Bewegung: Stilistik und Stilometrie in der Romania*, 6, 151-164. <http://romanischestudien.de/index.php/rst/article/view/625/1320>.
- Cuéllar, Á., & Vega García-Luengos, G. (2023). «La francesa Laura». El hallazgo de una nueva comedia del Lope de Vega último. *Anuario Lope de Vega Texto literatura cultura*, 29, 131-198. <https://doi.org/10.5565/rev/anuariolopedevega.492>.
- García-Reidy, A. (2019). Deconstructing the Authorship of *Siempre ayuda la verdad*: A Play by Lope de Vega? *Neophilologus*, 103(4), 493-510.
- Hernández-Lorenzo, L. (2022). Stylistic Change in Early Modern Spanish Poetry Through Network Analysis (with an Especial Focus on Fernando de Herrera's Role). *Neophilologus*, 106, 397-417. <https://doi.org/10.1007/s11061-021-09717-2>.
- Jockers, M. L. (2013). *Macroanalysis: Digital Methods and Literary History*. University of Illinois Press.
- Moretti, F. (2013). *Distant Reading*. Verso.
- Schreibman, S., Siemens, R., & Unsworth, J. (Eds.). (2016). *A New Companion to Digital Humanities*. Wiley-Blackwell.

Philip Hillebrand (Universität Münster) & Susanne Schlünder (Universität Osnabrück)

## **El concepto de *naturaleza* en el imperio español del dieciocho. Perspectivas de la historia conceptual y de los estudios culturales críticos**

El concepto de *naturaleza* forma parte del repertorio conceptual fundamental de una sociedad. Durante la Ilustración, no solo adquiere su configuración moderna en una variedad de discursos epistemológicos, económicos y sociopolíticos, sino que, en contextos coloniales como el Imperio español, también expresa y condiciona asimetrías tanto materiales como simbólicas (Carrasco y Schlünder 2022) entre centros y periferias. Esto se hace palpable en la confluencia de discursos naturalistas y raciales en la así llamada *Disputa del Nuevo Mundo* (Castro-Gómez 2010), en la formación de discursos ilustrados patrióticos en vísperas en el proceso de independencia (Nieto Olarte 2007), o en el surgimiento del conservacionismo ambiental como respuesta a la creciente destrucción forestal en la península ibérica y los territorios americanos (Grove 1995; Lucena Giraldo 1991).

A pesar de su relevancia, el concepto de naturaleza ha permanecido en gran medida desatendido en las investigaciones recientes sobre la historia conceptual del Atlántico ibérico (Fernández Sebastián, 2009-2014). En este contexto, la ponencia presenta la concepción y los primeros resultados de un proyecto sobre la semántica histórica del concepto de naturaleza en España e Hispanoamérica (1700-1835). En la elaboración del corpus, se presta especial atención a la integración de tanto obras canónicas como de textos menores del archivo colonial (Gallego Cuiñas 2024), contando hasta ahora con una cantidad de aproximadamente 30.000 documentos de diferentes géneros y temas, como relatos de viajes, diarios, textos literarios, historias naturales, tratados científicos, y mucho más, que se analizan con métodos computacionales (principalmente word embeddings y topic modeling). El objetivo es doble: por un lado, investigar la pluralidad semántica y el desarrollo diacrónico del concepto de *na-turaleza*, contribuyendo de esta manera a la historia conceptual digital (Wevers & Koolen 2020; Martinc, Kralj & Pollak 2020) del mundo hispanohablante; y por otro, explorar el aporte metodológico e interpretativo que pueden brindar conceptos de la teoría crítica cultural latinoamericana y poscolonial (Risam 2019) para el análisis de corpus históricos (coloniales) con métodos digitales.

### **Referencias**

- Carrasco M., Rolando & Susanne Schlünder (ed.): *Asymmetric Ecologies in Europe and South America around 1800*, Berlin/Boston: De Gruyter, 2022.
- Castro-Gómez, Santiago: *La hybris del punto cero*, Bogotá: Editorial Pontificia Universidad Javeriana, 2010.
- Fernández Sebastián, Javier (ed.): *Diccionario político y social del mundo iberoamericano*, 2 Vol., Madrid: Centro de Estudios Políticos y Constitucionales, 2009-2014.
- Ana Gallego Cuiñas & Daniel Torres-Salinas (ed.): *Humanities and Big Data in Ibero-America*, Berlin/Boston: De Gruyter, 2024.
- Grove, Richard H.: *Green imperialism*, Cambridge: University Press 1995.
- Lucena Giraldo, Manuel: *El bosque ilustrado*, Madrid: Instituto Nacional para la Conservación de la Naturaleza 1991.
- Martinc, Matej, Petra Kralj & Senja Pollak: "Leveraging Contextual Embeddings for Detecting Diachronic Semantic Shift", *arXiv*, (2020), <https://arxiv.org/abs/1912.010Espanol72v2>.
- Nieto Olarte, Mauricio: *Orden natural y orden social*, Madrid: Consejo Superior de Investigaciones Científicas 2007.
- Risam, Roopika (2019): *New digital worlds. Postcolonial digital humanities in theory, praxis and pedagogy*, Evanston (Illinois): Northwestern University Press.
- Wevers, Melvin & Marijn Koolen, "Digital begriffsgeschichte: Tracing semantic change using word embeddings", *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 53, N°. 4 (2020), p. 226-243, <https://doi.org/10.1080/01615440.2020.1760157>.

Georg A. Kaiser, Svenja Krieger & Anja Weingart (Universität Konstanz)

### **Konstanz Corpus Annotation Tool (KonCAT): Ein Softwaretool zur Verarbeitung von Daten in (historischen) (Parallel-)Textkorpora romanischer Sprachen**

Ein Problem der Verarbeitung sprachlicher Daten besteht darin, dass sich die Bedingungen der Datenspeicherung ständig ändern und diese laufend auf dem neuesten Stand gebracht werden muss, damit die Daten nicht unbrauchbar werden. Ein ähnliches Problem besteht bei der Datenannotation. Sie hängt vom theoretischen Rahmen ab, der ebenfalls Änderungen unterliegt, die dazu führen können, dass ursprüngliche Annotationen für die Untersuchung neuer Fragestellungen ungeeignet sind.

In unserem Vortrag stellen wir das Software-Tool KonCAT vor, das diese Probleme zu lösen versucht. Es wurde für ein Paralleltextkorpus entwickelt, der u. a. Bibelübersetzungen in alt- und neuromanische Sprachen sowie in einige nicht-romanische Sprachen enthält. Die Übersetzungen sind entsprechend der biblischen Verseinteilung parallel einander zugeordnet und teilweise in Haupt- und Nebensätze unterteilt. Diese Unterteilung und die sprachwissenschaftliche Annotation der Daten wurden bislang manuell durchgeführt, da alle bisherigen automatisierten Verfahren (noch) keine vollständig zufriedenstellenden Ergebnisse hervorgebracht haben (Kalouli et alii 2018).

KonCat ist webbasiert und orientiert sich an den FAIR-Prinzipien. Das Prinzip der Wiederverwendung des Quellcodes ist von zentraler Bedeutung. Daher werden für die Entwicklung der Anwendung nur freie Software und offene Standards verwendet und die Anwendung selbst wird unter der *European Public License* (The EUPL-v12) veröffentlicht. KonCAT zeichnet sich durch die Möglichkeit der selbstständigen Annotation von Daten und deren Änderung aus, was bisher in online annotierten Korpora nicht möglich ist (Pintzuk 2019). Annotationen sind jedoch sehr fehleranfällig und von forschungsspezifischen Anforderungen abhängig. Häufig gibt es, wie illustriert werden soll, unterschiedliche Bewertungen derselben Daten, wie z. B. die unterschiedliche Bestimmung der Grundwortstellung des Altfranzösischen aufgrund unterschiedlicher Auszählungen bei der Bestimmung der Verbposition (Kaiser 2002:78f, Wolfe 2021:51). KonCAT würde es ermöglichen, verschiedene Annotationen zu erfassen und die Gründe für die verschiedenen Analysen sichtbar zu machen (Weingart & Kaiser 2022). Außerdem ermöglicht die Ausnutzung der Möglichkeiten der Webtechnologie eine gemeinschaftliche Bearbeitung von Daten (Erstellung, Kommentierung, Suche, Anzeige), die Trennung von Datenspeicherung und Datendarstellung sowie die Aggregation von Daten mithilfe von Annotationen (Metadaten). Bisher wurden solche Lösungen für die Datenverarbeitung nur selten eingesetzt, weil anpassungsfähige Softwarelösungen nicht frei verfügbar sind und weil es an Wissen im Bereich der Forschungsdaten und der Softwaretechnologie zum Aufbau einer Lösung mangelt (Krefeld & Lücke 2023). Hier bietet sich KonCAT als ein erster Lösungsansatz an.

### **Referenzen**

- Kaiser, G.A. 2002. *Verbstellung und Verbstellungswandel in den romanischen Sprachen*. Tübingen.
- Kalouli, A.-L., K. Kaiser, A. Hautli-Janisz, G.A. Kaiser & M. Butt. 2018. A multilingual approach to question classification. In: N. Calzolari et al. (eds.), *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. Miyazaki, 2715–2720.
- Krefeld, T. & S. Lücke. 2023. Sketching out a virtual framework for Romance geolinguistics. In: E. Pustka et al. (eds.), *Corpus Dialectology*. Amsterdam, 198–216.
- Pintzuk, S. 2019. Adding linguistic information to parsed corpora. *Linguistic Issues in Language Technology* 18, 1–34.
- Weingart, A. & G.A. Kaiser. 2022. Eine FAIRe Anwendungssoftware für textbasierte Forschungsdaten. Das UV2 Annotationstool. *apropos [Perspektiven auf die Romania]* 9, 240–253.
- Wolfe, Sam. 2021. *Syntactic Change in French*. Oxford.

Ines Kremer (Universität Osnabrück)

## **Bekannt, verkannt, vergessen? Concepción de Estevarena (1854-1876) im (digitalen) Spiegel ihrer Zeit**

Zu den *escritoras románticas*, die in jüngerer Zeit vermehrtes Forschungsinteresse erfahren (Kirkpatrick 1989; 2002; 2009), zählt die in Sevilla geborene Concepción de Estevarena (1854-1876). Ihre Dichtung erschien ein Jahr nach ihrem frühen Tod unter dem Titel *Últimas flores* (1877). Maßgeblich getragen wurde diese Veröffentlichung von der Familie Velilla, mit der Estevarena freundschaftlich verbunden gewesen war. Der Prolog stammt von José de Velilla, einem engen Freund der Dichterin, und vermittelt biografische Informationen über Estevarena und ihr Umfeld, die ihre spätere Rezeption maßgeblich bestimmten. In der aktuellen Auflage ihres Gesamtwerks (Ediciones Torremozas 2005; 2018) bleibt dieser Text jedoch ebenso unkommentiert wie ihre Gedichte und die beigegefügte *Corona poética*. Daher soll in Zusammenarbeit mit Prof. Folke Gernert und dem Trier Center for Digital Humanities eine digitale Neu-Edition entstehen, die das Netzwerk Estevarenas auf den Ebenen der Produktion, der Distribution und der Rezeption sichtbar werden lässt (Mayoral 1990; Gronemann/Komorowska 2023). Diese Neu-Edition wird den Prolog als historische Quelle einordnen und erläutern, welche Rolle die Familie Velilla im literarischen Leben Sevillas spielte. Überdies wird sie ihre Korrespondenz integrieren und Hintergrundinformationen zu den Autor:innen der *Corona poética* liefern, deren Verbindungen zu Estevarena in der vorliegenden Ausgabe größtenteils unklar bleiben, sowie zu jenen Personen, denen Estevarena selbst Gedichte widmete. Mit der Verortung Estevarenas im literarischen Feld (Bourdieu 1992) ihrer Zeit leistet das Projekt zugleich einen Beitrag zur Forschung über geschlechtsspezifische Bedingungen von Autorschaft (Gilbert/Gubar 1980; Hahn 1991; Kroll/Zimmermann 1995; Bontrup/Metzler 2000; Gronemann/Schwan/Sieber 2012).

### **Referenzen**

- Bontrup, Hiltrud/Metzler, Christian (Hrsg.): *Aus dem Verborgenen zur Avantgarde. Ausgewählte Beiträge zur feministischen Literaturwissenschaft der 80er Jahre*. Hamburg: Argument Verlag, 2000.
- Bourdieu, Pierre: *Les règles de l'art. Genèse et structure du champ littéraire*. Paris: Seuil, 1992.
- Estevarena, Concepción de: *Últimas flores*. Madrid: Colección Torremozas, 2018.
- Gilbert, Sandra M./Gubar, Susan: *The Madwoman in the Attic. The Woman Writer and the Nineteenth-Century Literary Imagination*. New Haven; London: Yale University Press, 1980.2
- Gronemann, Claudia/Komorowska, Agnieszka (Hrsg.): *Fe/Male Friends: Staging Gender and Friendship in Seventeenth- and Eighteenth-Century Spanish Literature*. Frankfurt am Main/Madrid: Vervuert/Iberoamericana, 2023.
- Gronemann, Claudia/Schwan, Tanja/Sieber, Cornelia (Hrsg.): *Strategien von Autorschaft in der Romania: zur Neukonzipierung einer Kategorie im Rahmen literatur-, kultur- und medienwissenschaftlich basierter Geschlechtertheorien*. Heidelberg: Universitätsverlag Winter, 2012.
- Hahn, Barbara: *Unter falschem Namen. Von der schwierigen Autorschaft der Frauen*. Berlin: Suhrkamp, 1991.
- Kirkpatrick, Susan: *Las Románticas. Women Writers and Subjectivity in Spain, 1835-1850*. Berkeley [u.a.]: University of California Press, 1989.
- *An Anthology of Nineteenth-Century Women's Poetry from Spain*. New York: The Modern Language Association, 2009.
- „The Uses of Romantic Poetry: Feminine Subjects in Modern Spanish Culture“, in: Esterhammer, Angela (Hrsg.): *Romantic Poetry*. Amsterdam, Philadelphia: John Benjamins Publishing Company, 2002, S. 509-524.
- Kroll, Renate/Zimmermann, Margarete (Hrsg.): *Feministische Literaturwissenschaft in der Romanistik. Theoretische Grundlagen – Forschungsstand – Neuinterpretationen*. Stuttgart: Metzler, 1995.
- Mayoral, Marina: „Las amistades románticas: confusión de fórmulas y sentimientos“, in: dies. (Hrsg.): *Escritoras románticas españolas*. Madrid: Fundación Banco Exterior, 1990, S. 43-71.

Daniela Kuschel (Heinrich Heine-Universität Düsseldorf)

## **Reflexionen eines Forschungsdesigns: *Mix Methods* zur Analyse der literatur-, wissens- und kulturgeschichtlichen Bedeutung von *Paralysie*.**

Im Zentrum des diachron angelegten Forschungsprojekts „Körper – Lähmung – Mobilität“ steht die Erforschung des Konzepts der Paralysie im Kontext übergeordneter literatur-, kultur- und wissensgeschichtlicher Fragestellungen. Dabei werden einerseits konkrete Repräsentationen von Lähmungserscheinungen in fiktionalen (größtenteils kanonisierten) französischsprachigen Texten – d.h. Figuren, Motive, Themen – und Bezugnahmen in nicht-fiktionalen Texten, z. B. (auto-)biographische Verweise oder medizinische Ausführungen, in den Blick genommen. Andererseits untersucht die Studie die über die literarischen Figurationen hinausgehende, metaphorisch-figurative Bedeutung von Lähmung als konstitutivem Element von Wissenskulturen und sozialen Ordnungen. Angesiedelt in den (*Literary Disability Studies* (cf. u.a. Hall 2016, Stiker 2013) werden neben der Analyse literarischer Strukturen insbesondere metaphorologische (cf. u.a. Blumenberg 2013, Lakoff/Johnson 2003) und diskursanalytische (cf. Waldschmidt 2008 und Foucault 2017) Theorien und Konzepte fruchtbar gemacht. Sie erlauben es Lähmung als komplexes Konzept zu fassen, das unterschiedlichste Im/mobilitäts-Narrative ausbildet, die wiederum Wissen und Vorstellungen über Behinderung, Krankheit und soziale Teilhabe beeinflussen.

Der Vortrag reflektiert das Forschungsdesign der Studie, seine Potenziale und Schwierigkeiten – praktische und disziplinäre ebenso wie methodische und epistemologische. An prominenter Stelle steht dabei die Kombination aus *distant* und *close reading* Verfahren, die auch als Vermittlerin zwischen disziplinären Perspektiven und Grenzen fungieren kann.

Anhand konkreter Beispiele und bislang erfolgter Auswertungen wird zunächst gezeigt, wie die Integration digitaler Technologien (hier insbesondere aus der Korpus- und Computerlinguistik, cf. Andresen 2024) und ihre Bedingungen die Forschungsfrage maßgeblich mitgestaltet haben; wie die Festlegung des Korpus (Texte des 18.-20. Jh. der digitalen Sammlung „Les classiques de la littérature“ der BnF) den kultur- und wissensgeschichtlichen Aspekt erst ins Zentrum gerückt und wie Entdeckungen im Korpus relevante Fragen, insbesondere der Metaphorik und Koproduktion von Wissen, aufgeworfen haben.

Anschließend soll anhand bislang erfolgter Auswertungen die Bedeutung eines *Mixed Methods-Ansatzes* erörtert werden: die Auseinandersetzung mit Metadaten verändert den traditionellen Blick auf den literarischen Text, der Weg hin zur maschinellen Annotation (zur Unterscheidung von konkreten und figurativen Verweisen) lenkt den Blick darauf, wie (neue) technologische und methodische Einsichten immer wieder neue Fragen, Möglichkeiten, aber auch Probleme aufwerfen. Diesen Raum eröffnen auch gezielte Auswertungen mit Hilfe computergestützter Textanalysetechniken wie *word embeddings*, die exemplarisch diskutiert werden.

### **Referenzen**

- Andresen, Melanie. *Computerlinguistische Methoden für die Digital Humanities*. Narr Francke Attempto, 2024.
- Blumenberg, Hans. *Paradigmen zu einer Metaphorologie*. Suhrkamp, 2013.
- Foucault, Michel. *Naissance de la clinique*. Presses Universitaires de France, 2017.
- Hall, Alice. *Literature and disability*. Routledge, 2016.
- Lakoff, George/Mark Johnson. *Metaphors We Live By*, University of Chicago Press, 2003.
- Stiker, Henri-Jacques. *Corps infirmes et sociétés*. Dunod, 2013.
- Waldschmidt, Anne: „Wir Normalen‘ - ‚die Behinderten‘? Erving Goffman meets Michel Foucault“. *Die Natur der Gesellschaft: Verhandlungen des 33. Kongresses der Deutschen Gesellschaft für Soziologie in Kassel 2006*. Teilbd. 1 u. 2, hg. von Karl-Siegbert Rehberg, Campus Verl., 2008, S. 5799–5809.

Tommaso Meozzi (Universität Wien) & Simone Reborà (Università di Verona)

## **La rappresentazione letteraria del plurilinguismo: indagine sul corpus *BiBit***

Nell'ultimo trentennio l'autobiografia linguistica, ovvero il racconto autobiografico del proprio rapporto con le lingue, ha ricevuto in Italia un'attenzione crescente, sia in ambito pedagogico, che nell'ambito della critica letteraria, grazie alle opere di autori immigrati in Italia che hanno scelto l'italiano come lingua di scrittura.

Per quanto riguarda quest'ultimo aspetto, si è spesso utilizzato il concetto di "letteratura della migrazione" che, se certamente evidenzia il carattere transculturale e identitario delle opere, non si focalizza sull'aspetto del plurilinguismo. Si tratta inoltre di un concetto che tende a separare queste opere dalla prospettiva più generale del canone della letteratura italiana. La nostra ipotesi di lavoro è che la rappresentazione letteraria del plurilinguismo costituisca invece un elemento fondamentale in grado di gettare nuova luce sui processi narrativi di creazione identitaria, e che vada dunque analizzato nel contesto più ampio possibile della letteratura italiana.

In questa prospettiva vengono in aiuto gli strumenti delle *Digital Humanities*, che consentono un approccio integrato di *distant reading* e *close reading*. Si utilizzerà l'archivio *BiBit*, che ospita circa 1600 testi canonici della letteratura italiana, per individuare i brani in cui sono presenti esperienze di plurilinguismo.

La procedura sarà avviata da un primo filtraggio dei testi attraverso strumenti di *Natural Language Processing* (ricerca di parole chiave, riconoscimento automatico di passaggi plurilingui), finalizzato a individuare le parti di testo candidate a ospitare il fenomeno ricercato. Il filtraggio sarà poi raffinato tramite tecniche di *Machine Learning*, procedendo quindi a un'annotazione manuale volta a distinguere i passaggi di interesse dai falsi positivi, per allenare un modello di intelligenza artificiale da applicare su larga scala.

I passaggi estratti dall'archivio *BiBit* saranno infine analizzati combinando due prospettive complementari: *distant reading* delle linee di evoluzione del fenomeno in chiave emotiva e semantica; *close reading* delle caratteristiche stilistiche e tematiche che caratterizzano il multilinguismo nella tradizione letteraria italiana. I risultati offriranno una nuova prospettiva di larga scala su un fenomeno ancora poco studiato, che permetterà anche di sviluppare una serie di ragionamenti teorico- metodologici sull'applicazione dei metodi computazionali allo studio della letteratura.

### **Bibliografia essenziale**

- Blei, D. M. (2012). Probabilistic topic models, *Communications of the ACM*, 55(4), 77–84.
- Gnisci A. (1998). *La letteratura italiana della migrazione*, Lilit Edizioni, Roma.
- Hatzel, H. O., Stiemer, H., Biemann, C., & Gius, E. (2023). Machine learning in computational literary studies, *It - Information Technology*, 65(4–5), 200–217.
- Jänicke, S., Franzini, G., Cheema, M. F., & Scheuermann, G. (2015). On Close and Distant Reading in Digital Humanities: A Survey and Future Challenges, *Eurographics Conference on Visualization (EuroVis)*, 83–103. <https://diglib.org/collections/a1f4a085-d3cf-4864-82a4-873b8570e732>.
- Jezek, E., & Sprugnoli, R. (2023). *Linguistica computazionale. Introduzione all'analisi automatica dei testi*, Il Mulino, Bologna.
- Meozzi, T (2022). Una narrazione autobiografica radicata nell'alterità: la rappresentazione letteraria dell'apprendimento linguistico, *Italiano LinguaDue*, 1, 755-771.
- Reborà, S. (2023). Sentiment Analysis in Literary Studies. A Critical Survey, *Digital Humanities Quarterly*, 017(2). <http://www.digitalhumanities.org/dhq/vol/17/2/000691/000691.html>

Céline Powell & Sascha Resch (Ludwig-Maximilians-Universität München)

## **Die *Africa*-Übersetzung Francesca Roberti Francos und *imitatio*. Eine Beispiel-Analyse mithilfe eines Machine Learning-Modells**

Die Praxis der *imitatio* und ‚imitative‘ (oder auch intertextuelle) Schreibverfahren sind seit der Antike von großer Bedeutung. Ihre Identifikation stellt jedoch vor Herausforderungen und ist gerade bei längeren Texten (oder gar ganzen Textkorpora) mit einem hohen Arbeitsaufwand verbunden (Resch 2025a). Aus diesem Grund scheint es reizvoll, ein System zu nutzen, das mithilfe maschinellen Lernens eine erste Sondierung hinsichtlich ‚imitativen‘ Schreibens ermöglicht. Ein erster ‚*proof of concept*‘ für ein solches System wurde mit Blick auf *imitatio Dantis* und *imitatio Petrarcae* – zwei der wohl einflussreichsten Formen ‚imitativen‘ Schreibens im Kontext italienischsprachiger Literatur – am Institut für Italienische Philologie entwickelt. In diesem Zuge wurden bereits erste Prüfungen auf Validität hin durchgeführt (Resch 2025b). Nun geht es jedoch darum, die Praxistauglichkeit des entwickelten Tools weiter zu erproben und so Schwachstellen und Verbesserungspotenziale zu identifizieren.

In diesem Zusammenhang möchten wir die *Africa*-Übersetzung Francesca Roberti Francos mithilfe des Tools untersuchen und die Ergebnisse (im Sinne eines *close reading*) interpretieren. Im Jahr 1776 erscheint in Padua die Übersetzung *Dell’Africa di Francesco Petrarca libro primo, volgarizzato da Egle Euganea P. A.*, die aus der Feder von Francesca Roberti Franco stammt. Mit dieser poetischen Übersetzung der lateinischen Verse von Petrarca erlangte sie literarischen Ruhm und sie etablierte sich nicht nur als Dichterin und prominente Vertreterin des *petrarchismo femminile* im Settecento, sondern auch als bedeutende Übersetzerin (Powell 2026). Daher würden wir gerne untersuchen: wie sind Aspekte der *imitatio Dantis* und der *imitatio Petrarcae* vor dem historisch-biographischen Hintergrund Robertis zu sehen? Welche Besonderheiten gelten bei der Betrachtung ‚imitativen‘ Schreibens und Übersetzungs-/Translationstätigkeit? Sind die Ergebnisse des Machine Learning-Tools für die literaturwissenschaftliche Arbeit nützlich? Wie können die Ergebnisse des Machine Learning-Tools interpretiert werden? Wo liegen die Grenzen des aktuell entwickelten Verfahrens? Es geht uns somit darum, anhand eines historisch bedeutenden Beispiels die konkrete Integration eines Machine Learning-Tools in die hermeneutische Praxis literaturwissenschaftlichen Arbeitens im Modus einer ‚Fallstudie‘ zu erläutern.

### **Referenzen**

- Powell, Céline (vsl. 2026): *Lettere di traduttrici venete « au tournant des Lumières » ossia l’officina della sensibilità e della sociabilità. Francesca Roberti Franco, Elisabetta Caminer Turra, Gioseffa Cornoldi Caminer* (Dissertationschrift – Abgabe 2025).
- Resch, Sascha (2025a): *Machiavelli und Dantes ‚Commedia‘. Hermeneutische und digitale Perspektiven jenseits der Imitatio*, Berlin/New York: De Gruyter.
- Resch, Sascha (2025b): „Imitatio and Machine Learning: Experiments Regarding Imitatio Dantis and Imitatio Petrarca“, zur Publikation eingereicht und im Peer Review-Prozess bei *Digital Scholarships in the Humanities*.

Natascha Rempel (Leibniz Universität Hannover)

**Von der *Revista Diáspora(s)* hin zu kubanischen Blogs der Gegenwart: Alternative Medien im digitalen Wandel und deren Erschließung mittels Methodenmix**

Der Vortrag zielt darauf ab, zu erörtern, inwieweit es sich als fruchtbar erweist, das kubanische Intellektuellenprojekt *Diáspora(s)* und seine gleichnamige Untergrundzeitschrift (1997-2002), die in acht Ausgaben in La Habana während der Sonderperiode mittels *samisdat* zirkulierte und einen neo-avantgardistischen Gegendiskurs zu verbreiten versuchte, mithilfe eines *mixed-methods*-Ansatzes zu untersuchen.

Neben dieser methodologischen Diskussion, die eine klassische Literaturanalyse um quantitatives Arbeiten und digitale Tools zu erweitern versucht, werden ebenso Herausforderungen der analytischen Texterschließung in Zeiten eines medialen Wandels am Beispiel verschiedener kubanischer (Zeitschriften-)Projekte beleuchtet.

Vereinbar mit rezenten Tendenzen innerhalb der romanistischen Zeitschriftenforschung soll *Diáspora(s)* daher nicht nur als vielschichtiges Netz erfasst werden, sondern auch in den Kontext mit weiteren kubanischen Zeitschriftenprojekten der 1990er- und 2000er-Jahre, E-Zines der Zeit und gegenwärtigen Blogprojekten gesetzt werden. Dabei werden verschiedene Akteur:innen unterschiedlicher Projekte und deren Synergien ermittelt. Weiter wird versucht, thematische und ästhetische Spezifika, Kontinuitäten und/oder Brüche aufzudecken und abzubilden sowie poetologische Transformationen aufzuzeigen.

Nanette Reißler-Pipka (Max Weber Stiftung, Bonn) & José Calvo Tello (Niedersächsische Staats- und Universitätsbibliothek Göttingen)

### **Kriterien für digitale romanistische Korpora in der Literaturwissenschaft - theoretische, methodische und pragmatische Zugänge**

Unabhängig von der Wahl der Methode (hermeneutisch, poststrukturalistisch, kulturanthropologisch, digital oder pluralistisch) gilt in der Literaturwissenschaft die Auswahl des Korpus als eine der entscheidenden Vorarbeiten, die immer auch theoretisch begründet werden muss: Richtet man sich am Kanon aus, grenzt man eher zeitlich oder stilistisch ein - oder geben gar die Gattung und das Medium die einzige feste Vorgabe?

Je nach eigenen zeitlichen, menschlichen und technischen Ressourcen im Projekt stellt sich auch die Frage, ob der Text (oder die sonstige mediale Grundlage der Sammlung) in digitaler Form und im passenden Format vorliegt, kein Hindernis dar: Es kann gescannt, digitalisiert, OCR'd und erschlossen werden, was noch nicht entsprechend vorliegt.

Dennoch erfordern viele digitale Methoden und Werkzeuge entweder eine Mindestanzahl an Texten im Korpus, ein bestimmtes Format, eine zeitliche und stilistische Begrenzung und vor allem eine sprachliche Einheitlichkeit, da sie ansonsten schlicht nicht anwendbar sind oder statistisch keine brauchbaren Ergebnisse liefern können.

Der Beitrag versucht, die Anforderungen aus der romanistisch-literaturwissenschaftlichen Community zu strukturieren und in pragmatische Empfehlungen für die Infrastrukturanbieter im Bereich digitaler Bibliotheken und Repositorien zu überführen. Gleichzeitig möchten wir der romanischen Literaturwissenschaft methodische Überlegungen zur Kriterienbildung vorschlagen, die eine realistische Anwendungsmöglichkeit in zukünftigen oder bestehenden Angeboten aus der Infrastruktur haben.

Als Datengrundlage verwenden wir die Antworten einer Umfrage zur Wichtigkeit von ausgewählten Metadaten, die in Katalogen, digitalen Bibliotheken und Repositorien zur Verfügung gestellt werden könnten, um die Zusammenstellung literarischer Korpora für die Forschung zu erleichtern.

#### **Literatur**

Calvo Tello, José, Nanette Reißler-Pipka, Florian Barth, Kerstin Jung, und Christof Schöch. „Questionnaire of the Survey: How Do You Compose Your Literary Corpus or Literary Collection?“, 4. Dezember 2023. <https://doi.org/10.5281/zenodo.10203624>.

Calvo Tello, José, Nanette Reißler-Pipka, Sally Chambers, Philippe Genêt, und Patrick Helling. „Metadata Desiderata for Literary Corpora: A Survey for Starting a Conversation between Literary Studies, Libraries and Research Data Repositories.“ 16. Mai 2024. <https://doi.org/10.5281/zenodo.11202132>.

Wu, Mingfang, Kathleen Gregory, Felicitas Löffler, Brigitte Mathiak, Fotis Psomopoulos, Uwe Schindler, Amir Aryani, et al. 'Ten Principles to Improve Dataset Discoverability', 28 August 2024. <https://doi.org/10.15497/rda/00120>.

Glenn Roe (Sorbonne Université)

**Donner du sens à de vastes collections textuelles hétérogènes : l'intertextualité dans l'archive du XVIII<sup>e</sup> siècle**

Dans cette présentation, Glenn Roe examinera le projet ModERN, financé par l'ERC, un programme de recherche sur cinq ans visant à conjuguer les technologies informatiques de pointe et les méthodologies traditionnelles afin de reconstituer et d'analyser les réseaux conceptuels et intertextuels du Siècle des Lumières français. Il s'agira en particulier d'aborder certains des défis rencontrés au cours des premières années du projet, notamment la constitution du corpus, l'alignement des textes à grande échelle, la réduction du bruit et les premières analyses des réseaux intertextuels.

Antonio Rojas Castro (Universidad de Tübingen) & Allison Keith (Universidad de Stuttgart)

### **Análisis computacional del gracioso en las comedias calderonianas**

El teatro del Siglo de Oro español se caracteriza por la presencia de personajes que, mediante arquetipos recurrentes, cumplen funciones narrativas y críticas esenciales. Uno de los más destacados es el gracioso, un criado cuya intervención humorística y metateatral no solo aligera la acción, sino que también ofrece comentarios críticos sobre el desarrollo dramático (Lobato 1994; Barone 2013; Nohe 2018).

Esta ponencia tiene como objetivo analizar el arquetipo del gracioso a través de dos experimentos complementarios utilizando el corpus CalDraCor. En primer lugar, se entrena un modelo de clasificación para diferenciar entre criados con funciones humorísticas y metatextuales (graciosos) y aquellos que desempeñan roles convencionales. Utilizando un extenso corpus anotado de comedias calderonianas en formato XML-TEI, el modelo aprende a reconocer rasgos distintivos en el lenguaje. El segundo experimento introduce el análisis de redes como herramienta para examinar las interacciones entre personajes. A partir de la extracción de datos de diálogos y escenas, se construyen redes sociales en la que cada nodo representa a un personaje y los enlaces reflejan interacciones directas o co-apariciones en escena. Mediante este enfoque, se obtienen métricas estructurales como la centralidad de grado, la centralidad de intermediación y la densidad de conexiones, con el fin de determinar el papel estratégico de los graciosos en la trama dramática.

La integración de ambos experimentos –el modelo predictivo y el análisis de redes– permite abordar el estudio del gracioso de manera holística. Mientras que el primer enfoque se centra en la identificación automática y la clasificación de personajes a partir de patrones lingüísticos, el segundo aporta una dimensión estructural que revela cómo se insertan estos personajes en la dinámica social y escénica de la obra.

Los resultados esperados tienen implicaciones tanto en el ámbito de los estudios literarios como en el de las Humanidades Digitales. Por un lado, se prevé que la aplicación del modelo predictivo contribuya a una comprensión más precisa de los rasgos estilísticos y funcionales del gracioso. Por otro lado, el análisis de redes promete revelar la importancia estructural de estos personajes, aportando datos cuantitativos que complementen la interpretación cualitativa de la obra. La sinergia entre ambos métodos posibilita una aproximación interdisciplinaria capaz de enriquecer el conocimiento sobre la evolución y la relevancia de los arquetipos en el teatro clásico.

En síntesis, esta ponencia propone un estudio innovador que combina técnicas de procesamiento de lenguaje natural y análisis de redes para investigar el arquetipo del gracioso. Al abordar tanto la dimensión textual como la estructural de la obra, se pretende ofrecer una visión integral de cómo el humor y la crítica social se entrelazan en la dramaturgia clásica, abriendo nuevas vías de investigación en el campo de las Humanidades Digitales.

### **Referencias**

- Barone, Lavinia. 2013. «La figura del gracioso en las comedias serias de Calderón». En *"Festina lente": actas del II Congreso Internacional Jóvenes Investigadores Siglo de Oro, JISO 2012: BIADIG: Biblioteca áurea digital v.17, 2013, ISBN 978-84-8081-385-3, págs. 23-34, 23-34*. GRISO (Grupo de Investigación Siglo de Oro). <https://dialnet.unirioja.es/servlet/articulo?codigo=5037271>.
- Lobato, María-Luisa. 1994. «Ensayo de una bibliografía anotada del gracioso en el teatro español del Siglo de Oro». <https://www.cervantesvirtual.com/obra/ensayo-de-una-bibliografia-anotada-del-gracioso-en-el-teatro-espanol-del-siglo-de-oro/>.
- Nohe, Hanna. 2018. «El gracioso como personaje metateatral: funciones y desarrollo a lo Largo del Siglo de Oro». *Hipogrifo. Revista de Literatura y Cultura del Siglo de Oro* 6 (1). <https://doi.org/10.13035/H.2018.06.01.44>.

Christof Schöch (Universität Trier)

## **Prolégomènes pour un corpus de référence de l'histoire du roman français (1700–2000)**

L'histoire du roman français représente un terrain d'exploration particulièrement riche pour les études littéraires. Malgré cette richesse, les recherches quantitatives à grande échelle sur l'évolution du roman français demeurent encore rares (voir cependant Paige 2022, Barré 2023). Le projet présenté ici a pour ambition de soutenir ce type de recherches en constituant un corpus de référence, offrant aux chercheurs un outil essentiel pour analyser l'histoire du roman français de manière systématique, sur une période allant de 1700 à 2000. La construction de ce corpus s'appuyera sur plusieurs corpus existants, notamment le corpus Chapitre (Leblond 2022), plusieurs corpus français issues d'ELTeC (Schöch et al. 2021) et d'une collection de romans du vingtième siècle préparée dans le projet *Beyond Words* (e.g. Du et al. 2022), couvrant ensemble plus de 4000 romans. Il reste toutefois des lacunes importantes, notamment pour la période 1700–1750, qui devront être comblées par une numérisation ciblée afin de garantir une meilleure couverture temporelle.

L'objectif est de fournir une base solide pour des analyses variées. Le corpus permettra d'abord des études stilométriques, afin d'identifier les signatures stylistiques et les influences littéraires, voire de développer des méthodes d'attribution d'auteurs. Il offrira également de nouvelles perspectives sur l'histoire des sous-genres romanesques, en retraçant leur diversification et leur évolution. Permettre les analyses de la représentation du discours – discours direct, indirect ou indirect libre – sera une autre ambition projet, afin d'observer comment les pratiques discursives évoluent en fonction des périodes ou des choix narratifs. En complément, le corpus permettra l'emploi de méthodes comme le topic modeling, qui permettra de cartographier les grandes thématiques qui traversent les romans et de suivre leur transformation au fil du temps.

Chaque texte du corpus sera encodé en XML-TEI et enrichi de métadonnées détaillées. Celles-ci décriront non seulement les caractéristiques narratives (sous-genre, perspective narrative, cadre spatio-temporel), mais aussi des éléments quantitatifs, comme le nombre de personnages ou de lieux, ainsi que le degré de canonisation des œuvres. En complément du corpus intégral présenté en texte intégral simple, un échantillon représentatif, composé d'extraits cohérents, sera annoté semi-automatiquement pour identifier des éléments spécifiques tels que le type de discours ou les entités nommées (personnages, lieux). Ce travail d'annotation constituera une base précieuse pour entraîner et/ou évaluer des modèles automatiques d'analyse.

En combinant rigueur philologique et méthodes computationnelles avancées, ce projet vise à poser les bases d'une nouvelle histoire littéraire du roman français en mettant à disposition de la communauté scientifique un corpus large et fiable.

## **Références**

- Barré, Jean; Thierry Poibeau (2023): "Beyond Canonicity: Modeling Canon/Archive Literary Change in French Fiction". *Proceedings of the Computational Humanities Research Conference 2023*, <https://ceur-ws.org/Vol-3558/paper9925.pdf>.
- Leblond, Aude (2022). *Corpus Chapitres : 2000 romans français du 19e et 20e siècles*. Github / ANR Chapitre. URL: <https://github.com/ANRChapitres/2000romans19e20e>
- Du, Keli; Julia Dudar; Christof Schöch (2022): "Evaluation of Measures of Distinctiveness. Classification of Literary Texts on the Basis of Distinctive Words", *Journal of Computational Literary Studies* 1(1). doi: 10.48694/jcls.102.
- Paige, Nicholas (2020). *Technologies of the Novel. Quantitative Data and the Evolution of Literary Systems*. Cambridge UP.
- Schöch, Christof; Roxana Patras; Tomaz Erjavec; Diana Santos (2021): "Creating the European Literary Text Collection (ELTeC): Challenges and Perspectives". *Modern Languages Open*, 1/25, 1–19. DOI: 10.3828/mlo.v0i0.364.