Folgert Karsdorp

Meertens Instituut
Koninklijke Nederlandse Akademie van Wetenschappen
folgert.karsdorp@meertens.knaw.nl
www.karsdorp.io

EDUCATION

PhD *Retelling Stories. A Computational-Evolutionary Perpective*, (*Cum Laude*), Radboud University, Nijmegen, 2012–2016

MA Linguistics (*Cum Laude*), Leiden University, 2007–2009

BA Dutch Language and Culture, Leiden University, 2004–2007

ACADEMIC APPOINTMENTS

2020– KNAW Meertens Institute
Senior researcher, Oral Culture Research Group & Affiliated member of the KNAW Humanities Cluster DHlab

2016–20 KNAW Meertens Institute
Tenure Track researcher, Oral Culture Research Group & Affiliated member of the KNAW Humanities Cluster DHlab

2016–16 Radboud University
Research assistant, supervision by Prof. Dr. Antal van den Bosch

2010–11 Instituut voor de Nederlandse Taal
Researcher, (Formerly known as the Instituut voor Nederlandse Lexicologie)

2009–10 Free University Berlin
Research Assistant, Supervision by Prof. Dr. Matthias Hüning

RESEARCH AREAS

I'm a researcher in Computational Humanities and Cultural Evolution at the Meertens Institute of the Royal Netherlands Academy of Arts and Sciences (Amsterdam, the Netherlands). My research focuses on why some cultural phenomena are adopted and persist through time, while others change or disappear. Additionally, I'm interested in measuring cultural diversity and compositional complexity, and how we can account for biases in our estimations of diversity. To do that, I use computational models from Machine Learning, Cultural Evolution, and Ecology. Besides cultural change and diversity, I'm also interested in teaching about computer programming in the context of the Humanities. I published a text book with Princeton University Press about using Python for Humanities data analysis.

PUBLICATIONS

Karsdorp, F., Fonteyn, L., & Manjavacas, E. (2022a). Introducing functional diversity: A novel approach to lexical diversity in (historical) corpora. *Proceedings of the Computational Humanities Research Conference, 2022*.

Karsdorp, F., Wevers, M., & Lottum, J. v. (2022b). What shall we do with the unseen sailor? estimating the size of the dutch east india company using an unseen species model. *Proceedings of the Computational Humanities Research Conference, 2022*.

Kestemont, M., Karsdorp, F., de Bruijn, E., Driscoll, M., Kapitan, K. A., Macháin, P. Ó., Sawyer, D., Sleiderink, R., & Chao, A. (2022). Forgotten books: The application of unseen species models to the survival of culture. *Science*, (6582), 765–769. https://doi.org/10.1126/science.abl7655

Karsdorp, F., & Fonteyn, L. (2021). Telkens weer hetzelfde verhaal. In N. Van der Sijs, L. Fonteyn, & M. Van der Meulen (Eds.), *Wat gebeurt er in het nederlands?! over taal, frequentie en variatie* (pp. 153–157). Sterck & De Vreese.

Karsdorp, F., Kestemont, M., & Riddell, A. (2021). *Humanities data analysis. case studies with python*. Princeton University Press. https://press.princeton.edu/books/hardcover/9780691172361/humanities-data-analysis

Karsdorp, F., Manjavacas, E., Fonteyn, L., & Kestemont, M. (2020). Classifying evolutionary forces in language change using neural networks. *Evolutionary Human Sciences*. https://doi.org/10.1017/ehs.2020.52

Kestemont, M., & Karsdorp, F. (2020). Estimating the loss of medieval literature with an unseen species model from ecodiversity. *Proceedings of the Workshop on Computational Humanities Research (CHR2020)*. http://ceur-ws.org/Vol-2723/short10.pdf

Manjavacas, E., Karsdorp, F., & Kestemont, M. (2020). A statistical foray into contextual aspects of intertextuality. *Proceedings of the Workshop on Computational Humanities Research (CHR2020)*. http://ceur-ws.org/Vol-2723/long28.pdf

Haverals, W., Karsdorp, F., & Kestemont, M. (2019a). Data-driven syllabification for middle dutch. *Digital Medievalist*, (1), 1–23. https://doi.org/10.16995/dm.83

Haverals, W., Karsdorp, F., & Kestemont, M. (2019b). Rekenen op ritme. een datagedreven oplossing voor het automatisch scanderen van de historische lyriek in de dbnl. *Vooys. Tijdschrift voor letteren*, (3), 6–17.

Karsdorp, F. (2019). Willekeur in culturele verandering. *Tijdschrift voor Nederlands Taal- en Letterkunde*, (4), 303–315.

Karsdorp, F., & Fonteyn, L. (2019). Cultural entrenchment of folktales is encoded in language. *Palgrave Communications*, (25). https://doi.org/10.1057/s41599-019-0234-9

Karsdorp, F., Manjavacas, E., & Kestemont, M. (2019a). Keepin' it real: Linguistic models of authenticity judgments for artificially generated rap lyrics. *PLoS ONE*, (10), e0224152. https://doi.org/10.1371/journal.pone.0224152

Karsdorp, F., van Kranenburg, P., & Manjavacas, E. (2019b). Learning similarity metrics for melody retrieval. *Proceedings of the 20th International Society for Music Information Retrieval Conference 2019 (ISMIR 2019)*, 1–8. http://archives.ismir.net/ismir2019/paper/000057.pdf

Kestemont, M., & Karsdorp, F. (2019). Het atlantis van de middelnederlandse ridderepiek. een schatting van het tekstverlies met methodes uit de ecodiversiteit. *Spiegel der Letteren*, (3).

Lassche, A., Karsdorp, F., & Stronks, E. (2019). Repetition and popularity in early modern songs. *Proceedings of the 2019 Digital Humanities conference*. https://pure.knaw.nl/portal/en/publications/repetition-and-popularity-in-early-modern-songs(4a56b6f6-4e6c-4757-a28e-010e8c6b8d28).html

Manjavacas, E., Karsdorp, F., & Kestemont, M. (2019a). Generation of hip-hop lyrics with hierarchical modeling and conditional templates. *Proceedings of the 12th International Conference on Natural Language Generation*. https://www.inlg2019.com/assets/papers/46_Paper.pdf

Manjavacas, E., Kestemont, M., & Karsdorp, F. (2019b). A robot's street credibility: Modeling authenticity judgments for artificially generated hip-hop lyrics. *Proceedings of the 2019 Digital Humanities conference*. https://pure.knaw.nl/portal/en/publications/a-robots-street-credibility-

modeling-authenticity-judgments-for-artificially-generated-hiphop-lyrics(80a790fe-e9cd-459d-9af4-29ccbec320ef).html
Karsdorp, F., Manjavacas, E., Burtenshaw, B., & Kestemont, M. (2017). Synthetische literatuur: Neurale netwerken als coauteur. *DIXIT: tijdschrift over taal- en spraaktechnologie*, *14*, 17–18. https://notas.nl/dixit/dixit_2017_ai.pdf
Manjavacas, E., Karsdorp, F., Burtenshaw, B., & Kestemont, M. (2017). Synthetic literature: Writing science fiction in a co-creative process. *Proceedings of the Workshop on Computational Creativity in Natural Language Generation (CC-NLG 2017)*, 29–37. http://aclweb.org/anthology/W17-3904
Karsdorp, F., & van den Bosch, A. (2016). The structure and evolution of story networks. *Royal Society Open Science*, *3*, 160071. http://dx.doi.org/10.1098/rsos.160071
Kestemont, M., Stover, J., Koppel, M., Karsdorp, F., & Daelemans, W. (2016). Authenticating the writings of julius caesar. *Expert Systems with Applications*, *63*, 86–96. http://dx.doi.org/10.1016/j.eswa.2016.06.029
Meder, T., Karsdorp, F., Nguyen, D., Theune, M., Trieschnigg, D., & Muiser, I. (2016). Automatic enrichment and classification of folktales in the dutch folktale database. *Journal of American Folklore*, *129*, 78–96. https://pure.knaw.nl/portal/files/2337691/JAF_VVB.pdf
Karsdorp, F., Kestemont, M., Schöch, C., & van den Bosch, A. (2015a). The Love Equation: Computational Modeling of Romantic Relationships in French Classical Drama. In M. A. Finlayson, B. Miller, A. Lieto, & R. Ronfard (Eds.), *6th workshop on computational models of narrative (cmn 2015)* (pp. 98–107). Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. https://doi.org/10.4230/OASIcs.CMN.2015.98
Keywords: French drama, social relations, neural network, representation learning
Karsdorp, F., van der Meulen, M., Meder, T., & van den Bosch, A. (2015b). Animacy detection in stories. In M. A. Finlayson, B. Miller, A. Lieto, & R. Ronfard (Eds.), *6th workshop on computational models of narrative (cmn 2015)* (pp. 82–97). Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. http://dx.doi.org/10.4230/OASIcs.CMN.2015.82
Karsdorp, F., van der Meulen, M., Meder, T., & van den Bosch, A. (2015c). Momfer: A search engine of thompson's motif-index of folk literature. *Folklore*, *126*(1), 37–52. https://doi.org/10.1080/0015587X.2015.1006954
Kestemont, M., Karsdorp, F., & Düring, M. (2014). Mining the twentieth century's history of from the time magazine corpus. *Proceedings of the 8th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH), E-ACL*, 62–70. https://pure.knaw.nl/portal/files/1071098/Karsdorp_2014_ACL.pdf
van Kranenburg, P., & Karsdorp, F. (2014). Cadence detection in western traditional stanzaic songs using melodic and textual features. *15th International Society for Music Information Retrieval Conference (ISMIR 2014)*, 391–396. https://pure.knaw.nl/portal/files/943009/Kranenburg_2014_ismir.pdf
Karsdorp, F. (2013). Het is groen en leeft nog lang en gelukkig. Classificatie van volksverhaalgenres op basis van formules. *Tijdschrift voor Nederlands Taal- en Letterkunde*, *129*(4), 274–288. https://pure.knaw.nl/portal/files/2546821/309_449_1_PB.pdf
Karsdorp, F., & van den Bosch, A. (2013). Identifying motifs in folktales using topic models. *Proceedings of BENELEARN 2013*, 41–49. https://pure.knaw.nl/portal/files/458984/Karsdorp_VdBosch.pdf
Karsdorp, F., & Hüning, M. (2012). De relatie tussen schema's en analogische verbindingen. *Nederlandse Taalkunde*, *17*, 261–267. http://neon.niederlandistik.fu-berlin.de/static/mh/Karsdorp_Huening_2012.pdf
Karsdorp, F., van Kranenburg, P., Meder, T., Trieschnigg, D., & van den Bosch, A. (2012a). In search of an appropriate abstraction level for motif annotations. In M. Finlayson (Ed.), *Proceedings of the 2012 computational models of narrative workshop* (pp. 22–26). https://pure.knaw.nl/portal/files/467034/CMN.2012.karsdorp.pdf

Karsdorp, F., van Kranenburg, P., Meder, T., & van den Bosch, A. (2012b). Casting a spell: Indentification and ranking of actors in folktales. In F. Mambrini, M. Passarotti, & C. Sporleder (Eds.), *Proceedings of the second workshop on annotation of corpora for research in het humanities (acrh-2)* (pp. 39–50). https://pure.knaw.nl/portal/files/481270/karsdorp_et_al2012b.pdf

Karsdorp, F., & Beekhuizen, B. (2010). Regelmaat in een regelloos systeem. de nederlandse superlatief. *Voortgang, jaarboek van de Neerlandistiek*, 29–50.

INVITED TALKS

2022 "Forgotten Books The application of unseen species models to the survival of culture". Invited talk at the CUDAN Open Lab Seminar series, Tallinn University, Tallinn, Estonia, September 19, 2022.

2022 "Forgotten Books The application of unseen species models to the survival of culture". Keynote at the Summer School for Literary Studies and DH, Leiden University, Leiden, the Netherlands, June 20, 2022.

2021 "Unseen Species Models from Ecology to Estimate the Losses of Medieval Literature: Advances in an International Comparison". Talk at the International Medieval Conference 2021 (IMC 2021) in the session "Loss and Transmission: Quantitative Approaches to Modelling the Dissemination and Survival of Medieval Literature", 8 July 2021

2021 "The Birds in the Bush. What can Occupancy Models from Ecology Teach us about the Survival of Medieval Literature?", Invited presentation at the symposium The Human in Digital Humanities, June 23 2021.

2021 "Libraries as book traps. Statistical methods from ecology to study the survival of historic literature", Keynote lecture at the conference *Old Books and New Technologies*: Medieval Books and the Digital Humanities in the Low Countries, 6 May 2021

2021 "Estimating the loss of medieval literature with methods from ecology", invited talk in the online lecture series Beyond the Patterns. 10 March 2021, FAU, Nürnberg, DE.

2021 "Estimating the loss of medieval literature with methods from ecology", invited talk in the ATNU/IES Virtual Speaker Series 2020/2021 #6. 21 February 2021, Newcastle, UK

2020 "Blind or directed? Cultural Evolution of Children's Books", Children's Literature and Digital Humanities, University of Antwerp, 22-23 October 2020

2020 "Bias, diversity and survival. Can statistical methods from ecology estimate the loss of medieval literature?" Online Árni Magnússon Birthday lecture, 13 november 2020, The Arnamagnæan Institute, University of Copenhagen

2020 "Estimating the loss of medieval literature with unseen species model from ecodiversity". Presentation at the online conference Dark Archives 20/20: A Voyage into the Medieval Unread and Unreadable, Tuesday 8th September 2020

2019 "Cultural entrenchment of folktales is encoded in language", Presentation at the Lecture series of the Leiden University Centre for Digital Humanities, Leiden, 28 February 2019

2019 "The artificial synthesis of hiphop lyrics". Humlab talk series. Umeå universitet, Sweden. 12 February 2019.

2019 "Keeping it real: the artificial synthesis of hiphop lyrics". Vogin-IP, Amsterdam. 22 March 2019

2018 "Hoe verzin je het?". Kenniscafe de Balie over wetenschap van creativiteit. April 16, 2018,

Amsterdam.

2018 "Cultural Evolution in Children's Literature". Cultural Evolution in Children's Literature, Jul 5, 2018.

2018 "How to Read a Million Stories? Digital Text Analysis for the Study of Children's Literature". Plenary lecture at Children's Literature Summer School 2018, University of Antwerp, 5 July 2018

2018 "Synthesising humanities. Explaining complex models through simple data synthesis". Closing keynote lecture at Digitial Humanities Benelux conference 2018. Amsterdam, The Netherlands, 8 June 2018

2018 "Cultural entrenchment of folktales is encoded in language", Presentation at the workshop Interdisciplinary Workshop on Folk and Fairytales Digital, 15 February 2018

2018 "Willekeur in de Geesteswetenschap". De analyse van emoties en betekenis in tekst en beeld, 15 Jun 2018, VU Amsterdam.

2017 "Een literaire robot als schrijfhulp". Robots: van hulpje tot kunstenaar (Research Files 4), Pakhuis de Zwijger, October 12, 2017.

2017 "Character Bias in Transmitting Folktales". 14th SIKS/Twente Seminar on Searching and Ranking. March 10, 2017.

OPEN DATA SETS

2019 *Supplemental Materials for "Humanities Data Analysis"*
Data discussed in the manuscript *Humanities Data Analysis: Case Studies with Python*. Each folder in this dataset contains data used or discussed in one chapter. Most of the data are texts published before 1900. These texts are in the public domain.
https://doi.org/10.5281/zenodo.891264

2016 *Story network data sets*
Data discussed in Karsdorp & Van den Bosch (2016): The structure and evolution of story networks. Royal Society Open Science, 3, 160071. https://doi.org/10.5281/zenodo.51588

OPEN-SOURCE SOFTWARE

2020–21 Copia. Estimating the survival of cultural heritage artifacts with unseen species models from ecology. Copia is a Python package that can be used for estimating the survival of artifacts from cultural heritage, based on established unseen species models from ecology.
https://copia.readthedocs.io/en/latest/

2018–19 Deep flow. Open source code for the Hip Hop experiment executed at Lowlands 2018.
https://github.com/fbkarsdorp/deepflow

2015–21 Momfer. Meertens Motif Finder. Online application and app for browsing Thompson's *Motif Index*. https://momfer.meertens.knaw.nl/

2013–19 Open source and open access materials for an interactive course about Python in the Humanities. https://github.com/fbkarsdorp/python-course

PUBLIC OUTREACH PROJECTS

2018 *Deepflow: Linguistic models of authenticity judgments for artificially generated rap lyrics*
Collaboration between the Meertens Institute, the University of Antwerp and Lowlands Science on generating artificial hip hop lyrics using neural language models. The project involved an experiment carried out in the context of a large, mainstream contemporary music festival in the Netherlands (Lowlands). We developed an app to study crowd-sourced authenticity judgments for such artificially generated texts.

2017 *Asibot: Writing science fiction in a co-creative process*
Collaboration between the Meertens Institute, the University of Antwerp, the CPNB foundation and writer Ronald Giphart on employing Artificial Intelligence for literary language generation. The result of this experiment was a new story, "De Robot van de Machine is de mens" – written by Giphart and a neural network language generation system – of which 250,000 copies were disseminated to the public during the national campaign "Nederland Leest" in November 2017.

TEACHING (SELECTED COURSES)

2022 "Humanities Data Analsysis wit Python". Digital Curriculums Inspirational Seminar, May 24 2022, Aarhus University, Denmark.

2022 "Humanities Data Analysis with Python". CUSO Winterschool, January 31 – February 4 2022, University of Neuchâtel, Switserland.

2021 "Humanities Data Analysis with Python. Wetting one's appetite with historic cookbooks". Workshop at the Leiden Summer School *Literary Studies & Digital Humanities*, 31 May 2021, Leiden University.

2021 "Automated Authorship Attribution". Guest lecture at Leiden University, April 28 2021, Leiden University.

2021 "Cultural Evolution and the Humanities". Guest lecture at Antwerp University, April 22, 2021. Antwerp University.

2018 "How to Read a Million Stories? Digital Text Analysis for the Study of Children's Literature". Full-day tutorial at the Children's Literature Summer School 2018, July 1 – 5, 2018, University of Antwerp.

2017 "Skills Training 'Introduction to Programming", December 4 – 8, 2017, Doctoral Training Unit 'Digital History & Hermeneutics' (DHH), University of Luxembourg.

2017 "Python programming for the humanities and allied social sciences", Mar 29, 2017 - Mar 31, 2017, Radboud Digital Humanities Spring School 2017

2017 "Text Analysis with Python", January 27, 2017, Digital Humanities Workshop Series, University of Manchester

2016 "Scraping Twitter with Python". Workshop on scraping Twitter data with Python. May 24, 2016. Digital Disruption in Asia Conference, Leiden.

2016 "Cultural Selection of Fairy Tales". Guest lecture on cultural selection and evolution of fairy tales. April 28, 2016, Digital Humanities Lecture Series, Radboud University.

2016 "Digital Text Analysis", Five-day (MA, PhD-level) course on computational text analysis, with a special focus on 'authorship attribution', 'text normalization', and 'linguistic profiling',

January 18 – 22, 2016, LOT Winter School, Tilburg University.

2016 "Python for the Humanities", The objective of the full (MA, PhD-level) course was to familiarize students with the programming language Python for the computational processing of texts. December 2015 – April 2016, Ghent University.

2015 "Python for the Arts and Humanities". Four-day (MA, PhD-level) workshop on programming with Python for the Humanities. March 23 – 26, 2015, Literary Lab, Ghent University.

2015 "Python for the Arts and Humanities". Three-day (MA, PhD-level) workshop on programming with Python in the Humanities and Social Sciences, February 16 – 18, 2015, Maynooth University

2014 "Introduction and tutorial in Humanities Programming with Python". Five-day (MA, PhD-level) workshop on programming with Python for the Humanities and Social Sciences. August 17 – 30, 2014. University of Göttingen.

2014 "Advanced Programming with Python". Five-day (MA, PhD-level) workshop on advanced topics in Humanities programming. European Summer School in Digital Humanities Culture & Technology, July 22 – August 1, 2014.

2014 "Introduction into Python for the Humanities and Social Sciences". March 31 – April 4 2014, University of Antwerp.

2013 "Introduction into Python Programming for the Humanities", Three-day (MA-, PhD-level) workshop on programming with Python for the Humanities. With Maarten van Gompel. April 2013, Radboud University.

THESIS SUPERVISION

2022 Supervision MA thesis by Jurrian Kooiman about detecting and classifying aspects of focalisation in literary texts, Utrecht University;

2021 Supervision MA thesis by Arjan van Dalfsen on applying Unseen Species models to the Short Title Catalog, the Netherlands, Utrecht University;

2020 Supervision Internship Liesje Linden about applying Neural Language Models to classify grammatical connectives in Dutch, Utrecht University;

2020 Supervision MA Thesis Joris Veerbeek on a computational perspective to the evaluation of literary critics, Utrecht University;

2018 Supervision MA thesis by Alie Lassche on Cultural Evolution in Dutch Early Modern Songs. Topical Fluctuations in the Dutch Song Database (1550-1750), Utrecht University;

CURRENT ACADEMIC SERVICE

2021 Guest editor of the *Journal of Open Humanities Data*

2021 Programme Chair of the SciPy 2021 Symposium on Computational Social Science & Digital Humanities

2020 – 2022 Founder and Programme Chair of the *Computational Humanities Research* conference

2017 – 2022 Steering and Programme Committee member of the DH Benelux conference

SELECTED MEDIA COVERAGE

2022 How much medieval literature has been lost? *Scientific American*, March 8, 2022.

2022 'Lost' medieval literature uncovered by techniques used to track wildlife, *Science Magazine*, February 17, 2022.

2022 Seule une infime partie des récits médiévaux a survécu, *Le Figaro*, March 11, 2022.

2022 Study finds 90 percent of medieval chivalric and heroic manuscripts have been lost, *Ars Technica*, February 21, 2022.

2022 De helft van de Nederlandse ridderromans is waarschijnlijk voorgoed verdwenen, *de Volkskrant*, February 18, 2022.

2020 De jacht op de verloren ridderromans. *NRC*, August 24, 2020.

2019 Man (m/v) valt in put... Hebben alle succesvolle verhalen eigenlijk dezelfde onderliggende plot?, *NRC*. July 13, 2019.

2018 Taal is de sleutel tot echte artificiële intelligentie. *De Tijd*. September 15, 2018.

2018 Raprobot krijgt op Lowlands een kans voor het grote publiek. *Trouw*. August 17, 2018.

2018 Rapper MC Turing heeft een indrukwekkende flow, voor een robot. *de Volkskrant*, August 2018.

2017 Guest on the television program *Pauw* to talk about the artificial writer Asibot, October 31, 2017.

2017 Giphart schrijft samen met een robot. *Trouw*, June 16, 2017.

2017 Giphart krijgt hulp van deze literaire robot. *NOS op 3*, June 16, 2017.

2017 Robot schrijft verhaal, Giphart ondersteunt, *Een Vandaag*, July 11, 2017.

2016 De grote boze wolf wordt steeds minder boos. *de Volkskrant*, July 4, 2016.

2016 Pas op, er zit een moraal aan het verhaal. *De Standaard*, June 4, 2016.

2016 New study could explain why we remake certain movies over and over again. Online article in Ars Technica, June 30, 2016.

Updated October 2022