

Kohei Watanabe, PhD

Email: watanabe.kohei@gmail.com
Website: <http://blog.koheiw.net>
Github: <https://github.com/koheiw>

Profile

Nationality: Japanese
Languages: **Japanese** (native), **English** (fluent)

Education

PhD in Social Research Methodology, Department of Methodology, **London School of Economics and Political Science (LSE)**, 2013 – 2018

- Learned quantitative social research methodology based on econometrics.
- Developed two semi-supervised machine learning models for textual data (both published as R packages and research papers).
- Developed software programs for text analysis in Python, R, and C++ (quanteda has been downloaded over 600,000 times; see the Software Development section)
- Applied quantitative text analysis tools to study Russia's information operations during the Ukraine crisis (published in *European Journal of Communication*).

Pre-doctoral Student, Institute of Communication Studies, **University of Leeds**, UK, 2012 – 2013

- Learned the international news production and distribution mechanism.

MA in Political Science, Department of Political Science, **Central European University**, 2010 – 2011, Budapest, Hungary

- Learned statistical analysis of survey, network, and textual data in political research.
- Performed dictionary analysis of international news in PHP and SQL (published in *International Communication Gazette*).

MA in Sociology, **Musashi University**, 2007 – 2009, Tokyo, Japan

BA in Literature, Department of Literature, **Seikei University**, 2001 – 2005, Tokyo, Japan

Skills

- R, C++, Python, JavaScript, PHP, MySQL, MongoDB, HTML, CSS

Research Interests

- Quantitative text analysis (text mining)
- Statistical software development
- International political communication
- National security perception
- Disinformation on social media

Career History

Lazard Asset Management

UK, 2020/03 – Present

Senior Data Scientist

Fulltime (2021/06 – present):

- Analyse reports, earning-call transcripts, and news articles about publicly traded companies in the US, Japan and Europe in terms of their financial and ESG performance.
- Employ semi-supervised machine learning tools (e.g. LSS, seeded-LDA) in multi-lingual text analysis.
- Work closely with the Director of Quantitative Equity Research in London as a member of the data science team.

Part-time (2020/03 – 2021/05):

- Developed a system to apply dictionary analysis and industry clustering using R packages that I develop (quanteda and proxyC).
 - We found that dictionary analysis has significant performance implication in the US market.
 - We found that results of dictionary analysis of the US and Japanese documents are comparable.
- Developed a system to collect, parse and clean 10K forms from EDGAR (1996 to 2020) using R and MongoDB on Azure Cloud.
- Developed a system to collect, parse and clean annual financial report from EDINET (2015 to 2020) using R and MongoDB on Azure Cloud.
- Integrated the data collection system and MSSQL for automated analysis of EDGAR documents (for production test).
- Developed a system to parse and clean European financial documents (2003 to 2020) using R and MongoDB on Azure Cloud.
 - Applied above-mentioned dictionary analysis to European documents.
- Applied latent semantic scaling (LSS) to measure companies' policy towards carbon emission reduction.
- Applied semi-supervised LDA to earning calls to identify topics using an R package that I developed (seededlda)

University of Innsbruck

Austria, 2019/06 – 2021/03

Senior Assistant Professor, Digital Science Center & Department of Political Science (fulltime, tenure track)

- Teach courses on text mining, data management and political communication; develop new ML algorithms for textual data analysis; study American Twitter users' engagement with disinformation.

Waseda University

Japan, 2018 – 2019

Assistant Professor, Waseda Institute for Advanced Study (fulltime)

- Developed methodology for quantitative analysis of Asian-language texts.
- Organized an international conference on quantitative text analysis (POLTEXT) for the first time in Asia.
- Teach quantitative text analysis methodology at the School of Political Science and Economics.

London School of Economics and Political Science

UK, 2014 – 2018

Research Officer, Department of Methodology & US Centre

- Developed a quantitative text analysis package (Quanteda) in R and C++ (with Prof. Kenneth Benoit), and applied the package to study US foreign policy (with Prof. Peter Trubowitz).

Research Assistant, International Relations Department & Social Psychology Department

- Developed automatic news download systems from online databases and websites.

Oxford University

UK, 2015 – 2017

Research Assistant, Centre for Experiment Social Science

- Developed interactive user interfaces on the Qualtrics survey platform for the Financial Conduct Authority.

Freelance

Japan, 2009 – 2010

Website Developer

- Worked with Inter Media Japan, the largest website development company in Japan.

Entrepreneur

Japan, 2005 – 2007

- Co-founded Japan's first audio book production and distribution company, Otobank (<https://www.otobank.co.jp>), whose clients include Toyota Motors and Japan Airlines.
- Co-founded a personal tutor agent, Todai Net (<http://www.todainet.com>), to recruit and dispatch Tokyo University students.

Other Affiliation

London School of Economics and Political Science

UK, 2019 – Present

Visiting Research Fellow, US Centre

Measure America's perception of geo-political threat since through the New York Times Archive.

Waseda University

Japan, 2019 – Present

Adjunct Researcher, Institute for Advanced Study

- Research coverage of nuclear threats by Japanese and Hebrew newspapers.

Quanteda Initiative (non-profit organization)

UK, 2018 – Present

Chief Technological Officer

- Maintain the quanteda package and its learning materials for text analysis.

Teaching Experience

- Lecturer, Text Mining, **University of Innsbruck**, Austria, 2019 – 2020
- Lecturer, Political Communication and Electoral Studies, **University of Innsbruck**, Austria, 2019 – 2020
- Visiting Lecturer, Statistics for Social Sciences, **Hifotsubashi University**, Japan, 2020/02
- Visiting Lecturer, Introduction to Text Analysis & Advanced Text Analysis, ECPR Method Summer School, **Central European University**, Hungary, 2019/08 - 09
- Lecturer, Political Text Analysis, **Waseda University**, Japan, 2019/04 – 06
- Visiting Lecturer, Quantitative Political Methodology, **Kobe University**, Japan, 2018 – 2019
- Visiting Lecturer, workshop on quantitative analysis using R, **Hungarian Academy of Science**, Hungary, 2018/05
- Visiting Lecturer, workshop on quantitative analysis using R, **Tel Aviv University**, Israel, 2018/03
- Visiting Lecturer, workshop on quantitative analysis using R, **WZB Berlin Social Science Centre**, Germany, 2018/01
- Teaching Assistant, LSE, Computer Programming (MY470), Department of Methodology, **LSE**, UK, 2017

- Instructor, workshop on quantitative analysis using R, European Symposium Series on Societal Challenges in Computational Social Science, **Alan Turing Institute**, UK, 2017/11
- Visiting Lecturer, workshop on quantitative text analysis, **Kobe University**, Japan, 2017/06
- Visiting Lecturer, workshop on quantitative text analysis, **Waseda University**, Japan, 2017/06
- Tutor, workshop on quantitative analysis using R, **LSE**, UK, 2017/05
- Teaching Assistant, Computing Packages for Applied Analysis (MY591), Department of Methodology, **LSE**, UK, 2016
- Teaching Assistant, Introduction to Quantitative Analysis (MY451), Department of Methodology, **LSE**, UK, 2014

Publication and Presentation

Peer-Review Journal Articles

- Kohei Watanabe, Elad Segev and Atsushi Tago. (Forthcoming) Discursive Diversion: Manipulation of Nuclear Threats by the Conservative Leaders in Japan and Israel, **International Communication Gazette**.
- Elad Segev, Atsushi Tago, and Kohei Watanabe. (2022), Could Leaders Deflect from Political Scandals? Cross-National Experiments on Diversionary Action in Israel and Japan, **International Interactions**, <https://doi.org/10.1080/03050629.2022.2044326>.
- Tobias Heidenreich, Olga Eisele, Kohei Watanabe and Hajo Boomgaarden. (2022) Exploring Engagement With EU News on Facebook: The Influence of Content Characteristics, Politics and Governance, <https://doi.org/10.17645/pag.v10i1.4775>.
- Peter Trubowitz and Kohei Watanabe. (2021), The Geopolitical Threat Index: A new text-based computational approach to identifying foreign threats, **International Studies Quarterly**, <https://doi.org/10.1093/isq/sqab029>.
- Kohei Watanabe. (2020). Latent Semantic Scaling: A Semisupervised Text Analysis Technique for New Domains and Languages, **Communication Methods and Measures**, <https://doi.org/10.1080/19312458.2020.1832976>.
- Kohei Watanabe and Yuan Zhou. (2020). Theory-Driven Analysis of Large Corpora: Semisupervised Topic Classification of the UN Speeches, **Social Science Computer Review**, <https://doi.org/10.1177/0894439320907027>.
- Amy Catalinac and Kohei Watanabe. (2019). Quantitative Text Analysis in Japanese. **WIAS Research Bulletin** 2018.
- Kenneth Benoit, Kohei Watanabe, Haiyan Wang, Paul Nulty, Adam Obeng, Stefan Müller, and Akitaka Matsuo. (2018). Quanteda: An R package for the quantitative analysis of textual data, **Journal of Open Source Software**, <https://doi.org/10.21105/joss.00774>.
- Kohei Watanabe. (2018). Newsmap: A semi-supervised approach to geographical news classification, **Digital Journalism**, <https://doi.org/10.1080/21670811.2017.1293487>.
- Tomila Lankina and Kohei Watanabe. (2017). 'Russian Spring' or 'Spring Betrayal'? The Media as a Mirror of Putin's Evolving Strategy in Ukraine, **Europe-Asia Studies**, <https://doi.org/10.1080/09668136.2017.1397603>.
- Kohei Watanabe. (2017). Measuring news bias: Russia's official news agency ITAR-TASS's coverage of the Ukraine crisis, **European Journal of Communication**, <https://doi.org/10.1177/0267323117695735>.
- Kohei Watanabe. (2017). The spread of the Kremlin's narratives by a western news agency during the Ukraine crisis, **Journal of International Communication**, <https://doi.org/10.1080/13216597.2017.1287750>.
- Kohei Watanabe. (2016). Computer-aided dictionary making: An efficient dictionary construction technique for content analysis, **Proceedings of the International Conference on the Advances in Computational Analysis of Political Text**.
- Kohei Watanabe. (2013). The western perspective in Yahoo! News and Google News: Quantitative analysis of geographic coverage of online news, **International Communication Gazette**,

<https://doi.org/10.1177/1748048512465546>.

Public Policy Reports

- Maria Snegovaya and Kohei Watanabe. (2021). The Kremlin's Social Media Influence inside the United States: A Moving Target, **Free Russia Foundation**.

Book Chapters

- Kohei Watanabe (in-press). Text-as-Data, Encyclopedia of Technology & Politics, **Edward Elgar Publishing**.
- Tomila Lankina, Kohei Watanabe and Yulia Netesova. (2020). Framing, Fabricating, and Leveraging Protest: How Russian State Media Manipulate Public Discontent. In Valerie Bunce, Karrie Koesel and Jessica Chen Weiss (Eds.) Citizens and the State in Authoritarian Regimes: Comparing Russia and China, **Oxford University Press**.

Doctoral Thesis

- Kohei Watanabe. (2017). Measuring Bias in International News: A Large-scale Analysis of News Agency Coverage of the Ukraine Crisis, <http://dx.doi.org/10.21953/lse.6dcccptor2ki6>.

Selected Presentation

- Quantitative Text Analysis in Economics, Japanese Economic Association Spring Meeting (online), Kwansei Gakuin University, Hyogo, Japan, 2021/05
- Mediation of Exception and Unease: Analysis of British Newspapers Coverage of Refugees in 2006-2018, International Conference on Resources and Human Mobility, Mahidol University International College, Bangkok, Thailand, 2019/12
- Application of distributional semantics in social sciences: Analysis of news coverage of the refugee crisis using word-embedding techniques, Austrian Linguistics Conference, Salzburg, Austria, 2019/12
- Searching for 'submarine' propaganda outlets: An efficient method to identify sources of false information on the web, EPSA Annual Conference, Queen's University Belfast, UK, 2019/06
- Making a topic dictionary for semi-supervised classification of the UN speeches, QTA-Dublin, University College Dublin, Ireland, 2019/06
- How to make a seed dictionary: Semi-supervised topic classification of the UN speeches, Waseda-Essex Scientific PoliSci/IR Collaboration and Exchange, Essex University, Colchester, UK, 2019/03
- Historical Analysis of New York Times Foreign News using Semi-Supervised Models, Society as Data, Waseda University, Tokyo, Japan, 2019/01
- Quantitative Text Analysis in International Political Research, The Japan Association of International Relations, Omiya, Japan, 2018/11
- Conspiracist propaganda: How Russia promotes anti-establishment sentiment online?, ECPR General Conference, Hamburg, Germany, 2018/08
- Quantitative Analysis of Textual Data with R, Tokyo R user group meeting, Yahoo Japan, Tokyo, Japan, 2018/07
- Internationalizing Text Analysis, Collaborative Research on Political Information Transmission, Waseda University, Tokyo, Japan, 2018/06
- Introduction to Machine Learning for Text Analysis, POLTEXT 2018, Hungarian Academy of Science, Hungary, 2018/05
- Building text analysis models using Quanteda, LSE Computational Social Sciences Hackathon, LSE, UK, 2018/04
- Ten Years of Protest: What Have We Learned?, PONARS Conflict, Contestation and Peacebuilding in Eurasia, LSE, UK, 2018/03
- Quanteda: quantitative text analysis of textual data, London R user group meeting, LSE, UK, 2018/01
- Analysis of hate speeches using machine learning: inferring polarity of reader comments on news sites using Wordfish, Alsace Japanese Study Seminar, Centre Européen d'Études Japonaises

d'Alsace, France, 2017/09

- Data-driven approach to bilingual text analysis: representation of US foreign policy in Japanese and British newspapers in 1985-2016, research seminar at Waseda University, Japan, 2017/05
- Britain in the World: Representation of international relations by the UK media in 2012-2016, PSA Media and Politics Group annual conference, London, UK, 2016/12
- Measuring news bias Russia's official news agency ITAR-TASS's coverage of the Ukraine crisis, IAMCR Annual Conference 2016, Leicester, UK, 2016/07
- Computer-aided dictionary making: An efficient dictionary construction technique for content analysis, PolText 2016, Dubrovnik, Croatia, 2016/07
- How to collect, code and analyse 200,000 news stories on Ukraine crisis, Amsterdam Text Analysis Conference, Netherlands, 2016/06
- Mapping International New: Lexicon-based and dictionary-based approaches for geographical classification, ICA Annual Conference, Fukuoka, Japan (Best Paper Award was granted), 2016/06
- Making domain-specific sentiment dictionary, Mapping the Cultural Authority of Science (MACAS) project workshop, Bilgi University, Turkey, 2015/03
- Reinventing dictionary-based content analysis: A new computer-aided dictionary making technique for social sciences, Mapping the Cultural Authority of Science (MACAS) project workshop, London School of Economics and Political Science, 2014/12

Additional Information

Professional Service

Academic Journal Review

- I have served as a reviewer for various academic journals such as Communication Research, Political Analysis, International Communication Gazette, Political Analysis, Journalism Studies, Security Studies, Post-Soviet Affairs, Communication Methods and Measures, and Online Information Review.

Conference Organization

- I served as the main organizer of POLTEXT 2019 conference at Waseda University (Japan) focusing on multi-lingual quantitative text analysis. The event attracted more than 60 presenting and 30 non-presenting participants from 26 countries.

Awards

- The Statistical Software Award 2020, The Society for Political Methodology, **The American Political Science Association**, 2020
- The Best Paper Award 2016, The Research Methodology Group, **The International Communication Association**, 2016

Research Grants

- Grants-in-Aid for Scientific Research (1,7000,000 JPY or approx. 140,000 EUR), **Japan Society for the Promotion of Science**, 2019 – 2021
- Symposium and Seminar Grant (800,000 JPY or approx. 6,500 EUR), **The Telecommunication Advancement Foundation**, 2019
- Murata Science Foundation Research Grant H26-020 (900,000 JPY or approx. 8,000 EUR), **Murata Science Foundation**, 2014 – 2015
- LSE Doctoral Scholarship (tuition waiver and stipend), **London School of Economics and Political Science**, 2013 – 2017

Membership

- European Consortium for Political Research

Software Development

quanteda

- Award-winning NLP infrastructure for quantitative text analysis in R (available on <https://quanteda.io> and <https://cran.r-project.org/web/packages/quanteda>).
- Used in research and teaching at leading universities that include LSE, Oxford University, New York University, Harvard University.

LSX

- Semi-supervised document scaling tool based on word-embeddings. (available on <https://github.com/koheiw/LSX> and <https://cran.r-project.org/web/packages/LSX>).

seededlda

- Semi-supervised LDA for document classification (available on <https://github.com/koheiw/seededlda> and <https://cran.r-project.org/web/packages/seededlda>).

newsmap

- Tool for multi-lingual geographical document classification (available on <https://github.com/koheiw/newsmap> and <https://cran.rproject.org/web/packages/newsmap>).

Updated on 2022/05/27