

Wessel Gauke Poelman

Email	contact@wesselpoelman.nl
Website	wesselpoelman.nl
GitHub	github.com/WPoelman
Scholar	Google Scholar
Research profile	Multilingual natural language processing (NLP), language model evaluation, linguistic typology, morphology, experimental methods for cross-lingual and low-resource language technology.

Education

2023–present	PhD in Computer Science, KU Leuven. Advisor: Miryam de Lhoneux. Experimental Design in Multilingual NLP. Expected: 2027.
2022	MSc in Information Science, Rijksuniversiteit Groningen, <i>cum laude</i> . Advisor: Johan Bos. Thesis results published at COLING 2022.
2021	BSc in Information Science, Rijksuniversiteit Groningen
2021	BA in Communication Sciences, Rijksuniversiteit Groningen Summer Schools: ESSLLI 2023, LxMLS 2024

Positions

2026	Visiting Researcher, Ludwig-Maximilians-Universität München, München, DE
2023	PhD Researcher, Technische Universität München, München, DE
2021–2023	Machine Learning Engineer, Web-IQ, Groningen, NL
2019–2021	Software Engineer, Beeproger, Groningen, NL
2018–2021	Teaching, Research, and Student Assistant, Rijksuniversiteit Groningen, Groningen, NL

Grants, honors & awards

2026	Cost Action Short Term Scientific Mission (€3k). Three months at LMU. Host: Barbara Plank.
2025	Danish Data Science Academy Travel Grant (€800). Aalborg University. Host: Johannes Bjerva.
2025	Best Paper Award Nominee at EMNLP 2025.
2025	Outstanding Reviewer (2×) ACL Rolling Review.
2024	Outstanding Paper Award, NLP for Conversational AI Workshop at ACL 2024.
2024	Cost Action Travel Grant (€700). Attended UniDive General Meeting in Budapest.

Publications

* denotes shared first authorship, ◇ denotes mentored students.

Preprints

2026	Wessel Poelman , Yiyi Chen & Miryam de Lhoneux. QQ: A Toolkit for Language Identifiers and Metadata.
------	---

Conference Proceedings

- 2026 Kushal Tatariya*, Artur Kulmizev*, **Wessel Poelman**, Esther Ploeger, Marcel Bollmann, Johannes Bjerva, Jiaming Luo, Heather Lent & Miryam de Lhoneux. How Good Is Your Wikipedia? Auditing Data Quality for Low-resource and Multilingual NLP. In *Proceedings of ACL*.
- 2026 Stef Accou[◊] & **Wessel Poelman**. TIPA: Typologically Informed Parameter Aggregation. In *Proceedings of EACL: Findings*.
- 2026 **Wessel Poelman** & Miryam de Lhoneux. Form and Meaning in Intrinsic Multilingual Evaluations. In *Proceedings of EACL*. Oral Presentation.
- 2025 Kushal Tatariya*, **Wessel Poelman*** & Miryam de Lhoneux. On the Interplay Between Positional Encodings, Morphological Complexity, and Word Order Flexibility. In *Proceedings of IJCNLP-AACL*.
- 2025 **Wessel Poelman***, Thomas Bauwens* & Miryam de Lhoneux. Confounding Factors in Relating Model Performance to Morphology. In *Proceedings of EMNLP*. Oral Presentation.
- 2025 **Wessel Poelman** & Miryam de Lhoneux. The Roles of English in Evaluating Multilingual Language Models. In *Proceedings of NoDaLiDa*.
- 2024 Esther Ploeger*, **Wessel Poelman***, Miryam de Lhoneux & Johannes Bjerva. What is “Typological Diversity” in NLP? In *Proceedings of EMNLP*. Oral Presentation.
- 2022 **Wessel Poelman**, Rik van Noord & Johan Bos. Transparent Semantic Parsing with Universal Dependencies Using Graph Transformations. In *Proceedings of COLING*.

Journals

- 2025 Esther Ploeger, **Wessel Poelman**, Andreas Holck Høeg-Petersen, Anders Schlichtkrull, Miryam de Lhoneux & Johannes Bjerva. A Principled Framework for Evaluating on Typologically Diverse Languages. *Computational Linguistics*.
- 2024 Pieter Fivez, Walter Daelemans, Tim Van Cruys, Yury Kashnitsky, Savvas Chamezopoulos, Hadi Mohammadi, Anastasia Giachanou, Ayoub Bagheri, **Wessel Poelman**, Juraj Vladika, Esther Ploeger, Johannes Bjerva, Florian Matthes & Hans Halteren. The CLIN33 Shared Task on the Detection of Text Generated by Large Language Models. *Computational Linguistics in the Netherlands*.

Workshops

- 2025 Robin Kokot[◊] & **Wessel Poelman**. Type and Complexity Signals in Multilingual Question Representations. In *Proceedings of the 5th Workshop on Multilingual Representation Learning at EMNLP*.
- 2024 Phillip Schneider, **Wessel Poelman**, Michael Rovatsos & Florian Matthes. Engineering Conversational Search Systems: A Review of Applications, Architectures, and Functional Components. In *Proceedings of the 6th Workshop on NLP for Conversational AI at ACL*. Outstanding Paper Award.
- 2023 Mahdi Dhaini, **Wessel Poelman** & Ege Erdogan. Detecting ChatGPT: A Survey of the State of Detecting ChatGPT-Generated Text. In *Proceedings of the Recent Advances in Natural Language Processing Student Research Workshop*.
- 2022 Frank van den Berg*, Gijs Danoe*, Esther Ploeger*, **Wessel Poelman***, Lukas Edman & Tommaso Caselli. RUG-1-Pegasussers at SemEval-2022 Task 3: Data Generation Methods to Improve Recognizing Appropriate Taxonomic Word Relations. In *Proceedings of SemEval at NAACL*.

Research Dissemination

Research Talks

- 2026 *Experimental Design in Multilingual Language Modeling*
MaiNLP, LMU, DE (planned for June 2026)
- 2026 NLP Research Seminar, KU Leuven, BE
- 2025 NLPnorth, IT University of Copenhagen, DK
- 2025 AAU-NLP, Aalborg University, DK
- 2025 CENTAL, UCLouvain, BE

Oral Presentations

- 2026 Form and Meaning in Intrinsic Multilingual Evaluations. *EACL*, Rabat
- 2025 Confounding Factors in Relating Model Performance to Morphology. *EMNLP*, online
- 2024 What is “Typological Diversity” in NLP? *EMNLP*, Miami

Poster Presentations

- 2026 Form and Meaning in Intrinsic Multilingual Evaluations. *Human Perspective on Cross-Lingual AI Models*, Georg-August-Universität Göttingen
- 2025 Multilingual Language Models: What do I need to keep in mind? *Leuven.AI Conference*, KU Leuven
- 2025 Relating Language Modeling Performance to Morphology. *Computational Linguistics in the Netherlands*, KU Leuven
- 2025 The Roles of English in Evaluating Multilingual Language Models. *NoDaLiDa*, Tallinn
- 2025 Quantifying Claims of Typologically Diverse Language Samples. *UniDive 3rd General Meeting*, Budapest
- 2024 Leveraging Linguistic Typology for Fairer Multilingual Language Technology. *Computational Linguistics in the Netherlands*, Universiteit Leiden
- 2024 A Call For Consistency in Reporting Typological Diversity. *Leuven.AI: Science of AI Workshop*, KU Leuven
- 2024 A Call For Consistency in Reporting Typological Diversity. *Human Perspective on Cross-Lingual AI Models*, VU Amsterdam
- 2023 Detecting Machine-Generated Text with Purely Linguistic Features. *Computational Linguistics in the Netherlands*, Universiteit Antwerpen

Teaching

- 2024–present Teaching Assistant, Natural Language Processing, MSc KU Leuven.
Designed assignments, taught weekly seminars, and graded exams. 150–180 students/year.
- 2024–present Teaching Assistant, Data Structures and Algorithms, BSc KU Leuven.
Taught weekly seminars and graded online assignments. 120–140 students/year.
- 2023 Co-organizer, Natural Language Processing – Methods and Applications, MSc TUM.
Organized course and supervised students with small research projects. ~40 students.
- 2022 Instructor, Workshop at LCT Summer School, Saarland University. With Esther Ploeger.

2019 Teaching Assistant, Introduction to Programming I, BSc RUG.
Taught weekly seminars and assisted with exam creation.

Research Advising

2026–present Sing Sing Ngai, MSc AI at KU Leuven
2026–present Ran Zhang, MSc AI at KU Leuven
2025 Stef Accou, MSc AI at KU Leuven, *TIPA: Typologically Informed Parameter Aggregation*. Published at EACL 2026: Findings. → PhD student at KU Leuven.
2025 Robin Kokot, MSc AI at KU Leuven, *Type and Complexity Signals in Multilingual Question Representations*. Published at MRL 2025. → PhD student at KU Leuven.
2023 Berkay Ugur Senocak, MSc Informatics at TUM, co-supervised with Mahdi Dhaini. *Implementing and Evaluating Methods for Detecting Machine-Generated Text*. → Researcher at Siemens.

Service & Volunteering

2024–present Reviewer, ACL Rolling Review (ACL, EACL, NAACL-HLT, EMNLP, AACL).
2024–present Working Group Sub-leader, *UniDive: Universality, Diversity, and Idiosyncrasy in Language Technology*, EU COST Action.
2023–present Software Engineer, Walter Schottky Institute at TUM. Writing drivers and parameter sweeps for measurement instrumentation in experimental physics lab.
2025 Program Chair, 2nd UniDive Training School on *Quantifying Linguistic Diversity*.
2025–2026 Guest Editor, *Computational Linguistics in the Netherlands Journal*.
2025 Co-organizer, 35th edition of *Computational Linguistics in the Netherlands Conference*.

Software

QQ. Toolkit for language identifiers and metadata (GitHub).
Planit. Tooling for SLURM-based research workflows (GitHub).
Typdiv Sampling. Toolkit for language sampling and analysis (GitHub).

Languages

Dutch (native), West-Frisian (native speaker), English (C2), German (A2, following B1 course)