

Timothee MICKUS

PERSONAL DATA

EMAIL: timothee.mickus@helsinki.fi GITHUB: [TimotheeMickus](https://github.com/TimotheeMickus) WEBSITE: timotheemickus.github.io
PHONE: +33 7 69 89 33 08 SOCIAL:     

WORK EXPERIENCE

APR 2022–Now Postdoctoral Researcher at UNIVERSITY OF HELSINKI LANGUAGE TECHNOLOGY GROUP, Helsinki
Machine Translation and explainable NLP, part of the FoTRAN and UNCERTAINTY-AWARE NEURAL LANGUAGE MODELS projects.
AUG 2018–SEP 2018 Web & NLP Engineer at ELDA, Paris; *Development—Web Annotation Interface*
NOV 2017–JUL 2018 NLP Data Scientist at ESPRITS COLLABORATIFS, Paris area; NLP—*R&D*

EDUCATION & SKILLS

INTERNSHIPS

FEB 2020–MAR 2020 NLP Intern at AIST AIRC, Tokyo
NOV 2017–JUL 2018 NLP Intern at UNIVERSITÉ PARIS DIDEROT, Paris
JAN 2016–JUN 2016 NLP Intern at UNIVERSITÉ PARIS DIDEROT, Paris
JUN 2015–SEP 2015 NLP Intern at LIPN, UNIVERSITÉ PARIS 13, Paris area

SCHOOL CURRICULUM

🏆 MAR 2022 PHD. in COMPUTER SCIENCE, ATILF, **Université de Lorraine**, Nancy
TITLE: “[On the Status of Word Embeddings as Implementations of the Distributional Hypothesis](#)”
Investigations on the linguistic value of word embeddings. Using NLG to study the relation between distributional semantic representations and traditional dictionary definitions. Awarded the *prix de thèse ATALA*.
JAN 2022 WINTER SCHOOL in MACHINE LEARNING, **ALPS 2022**, Online
JUL 2019 SUMMER SCHOOL in MACHINE LEARNING, **DeepLearn 2019**, Warsaw
JUL 2018 Master of Science in COMPUTATIONAL LINGUISTICS, **Université Paris Diderot**, Paris
summa cum laude, rank not available. Thesis subject: “[Grammatical Gender Contrasts in French: a distributional semantic approach](#)”.
JUL 2016 Undergraduate Degree in COMPUTATIONAL LINGUISTICS, **Université Paris Diderot**, Paris
cum laude, ranked 1st.
JUL 2015 Bachelor of Science in COMPUTATIONAL LINGUISTICS, **Université Paris Diderot**, Paris
summa cum laude, ranked 1st.

LANGUAGE SKILLS

NATIVE French & English
SOME PROFICIENCY German & Japanese

SCIENTIFIC OUTREACH

TEACHING

MAR 2024–JUN 2024, Lectures on **Natural Language Understanding**, at UNIVERSITY OF HELSINKI
MAR 2023–JUN 2023 as part of the Master’s Programme in LINGUISTIC DIVERSITY AND DIGITAL HUMANITIES.
AUG 2023 Lecture on **Multilingual Machine Translation**, within the 16TH MACHINE TRANSLATION MARATHON
Co-lecturer with J. TIEDEMANN and R. VÁZQUEZ.
NOV 2020–FEB 2021 Lectures on **Word Embeddings** and **Data Mining**, at UNIVERSITÉ DE LORRAINE
as part of the “Lexical Resources” class of the M.Sc. in NLP.
OCT 2020–JAN 2021, Practical Lectures on **Programming** (Python), at UNIVERSITÉ DE LORRAINE
OCT 2019–JAN 2020 as part of the “Algorithmics & Programming” class of the M.Sc. in NLP.
OCT 2020–DEC 2020 Practical Lectures on **Web Development**, at UNIVERSITÉ DE LORRAINE
HTML, CSS, JavaScript.
NOV 2020 Practical Lectures on **Shell** and **TeX**, at UNIVERSITÉ DE LORRAINE

NOV 2019–JAN 2020, NOV 2018–DEC 2018 Lectures on **Word Embeddings, Multilingual Resources and Data Mining**, at UNIVERSITÉ DE LORRAINE
as part of the “Lexical Resources” class of the M.Sc. in NLP.

AUG 2019 Tutor for PYTHON4NLP Summer School, Nancy
Providing help and guidance to attendees on presented python APIs & NLP libraries.

SUPERVISION

JAN 2024–NOW Supervised two NLP Masters 1 students
Topics: comparison of MT and LM objectives; uncertainty-aware definition modeling.

MAY 2023–NOW Co-supervision for a PhD student (Joseph Attieh)
Dissertation topic: Optimizing Multilingual Language Models for more Efficient Inference.

JUN 2020–AUG 2020 Supervised two NLP Masters 1 students
Topic: preliminary work for establishing comparable sets of word embeddings (work published).

FUNDING APPLICATIONS

JUN 2024 Successful application to the OTTO MALM FOUNDATION (5 000€)
For the development of a multilingual dataset of LLM hallucinations.

JUN 2023 Successful application to the OSKAR ÖFFLUND FOUNDATION (5 000€)
For the development of a dataset of hallucinations in sequence-to-sequence NLP models.

ORGANIZATION DUTIES

MAY 2024–JUL 2025 Co-organizer for the MU-SHROOM shared-task
A SemEval shared task on multilingual & token-level hallucination detection.

DEC 2023–MAY 2024 Co-organizer for the [AXOLOTL’24](#) shared-task
Lchange’24 shared task on detecting and explaining semantic changes.

OCT 2023–MAR 2024 Co-organizer for the [MOOMIN](#) workshop
ACL-affiliated workshop on modular NLP, collocated with EACL 2024.

MAY 2023–JUL 2024 Co-organizer for the [SHROOM](#) shared-task
SemEval 2024 Task 6, a shared task on the detection of hallucinations and overgeneration.

JUN 2021–JUL 2022 Co-organizer for the [CODWoE](#) shared-task
SemEval 2022 Task 1, a shared task on comparing embeddings & dictionaries.

PUBLIC OUTREACH

MAY 2023–NOW Blog posts on the [Language Technology Research Group’s blog](#).

NOV 2021 Demo for the BlankCrack game at the [FORUM DES SCIENCES COGNITIVES 2021 À NANCY](#).

JUL 2021–SEP 2022 Personal blog: “[My PhD, explained to my folks](#)”

APR 2021–DEC 2021 Member of the strategic committee “Université – Recherche – Santé”

JAN 2021–NOW Online game for linguistic data collection: “[BlankCrack](#)”

AWARDS

JUN 2024 Best task paper award at SEMEVAL 2024

JUN 2024 Best paper award at INSIGHTS FROM NEGATIVE RESULTS 2024

DEC 2023 Outstanding reviewer at EMNLP 2023

JUN 2023 Prix de thèse ATALA
Best PhD dissertation award from the French scientific association for NLP.

PEER-REVIEWED PUBLICATIONS

AUG 2024 [Domain-specific or Uncertainty-aware models: Does it *really* make a difference for biomedical text classification?](#)
A. SINHA, T. MICKUS, M. CLAUSEL, M. CONSTANT & X. COUBEZ. BionLP 2024 (ACL 2024).

AUG 2024 [AXOLOTL’24 Shared Task on Multilingual Explainable Semantic Change Modeling](#)
M. FEDOROVA, T. MICKUS, N. PARTANEN, J. SIEWERT, E. SPAZIANI & A. KUTUZOV. LChange’24 (ACL 2024).

- AUG 2024 [Isotropy, Clustering, and Classifiers](#)
T. MICKUS, S.-A. GRÖNROOS & J. ATTIEH. ACL 2024.
- JUN 2024 [Stranger than Paradigm: Word Embedding Benchmarks Don't Align With Morphology](#)
T. MICKUS & M. COPOT. SCiL 2024.
- JUN 2024 [Language Models and the Paradigmatic Axis](#)
T. MICKUS. SCiL 2024.
- JUN 2024 [The Emergence of High-Level Semantics in a Signaling Game](#)
T. BERNARD, T. MICKUS & H. TAKAMURA. *SEM 2024 (NAACL 2024).
- 🏆 JUN 2024 [I Have an Attention Bridge to Sell You: Generalization Capabilities of Modular Translation Architectures](#)
T. MICKUS, J. ATTIEH & R. VÁZQUEZ. Workshop on Insights from Negative Results in NLP 2024 (NAACL 2024).
- 🏆 JUN 2024 [Semeval-2024 Task 6: SHROOM – A Shared-task on Hallucinations and Related Observable Overgeneration Mistakes](#)
T. MICKUS, E. ZOSA, R. VÁZQUEZ, T. VAHTOLA, J. TIEDEMANN, V. SEGONNE, A. RAGANATO & M. APIDIANAKI. SEMEVAL 2024 (NAACL 2024).
- MAY 2024 [Can machine translation bridge multilingual pretraining and cross-lingual transfer learning?](#)
S. JI, T. MICKUS, V. SEGONNE & J. TIEDEMANN; LREC-COLING 2024.
- MAR 2024 [MAMMOTH 🐘 Massively Multilingual Modular Open Translation @ Helsinki](#)
T. MICKUS, S.-A. GRÖNROOS (equal contribution), J. ATTIEH, M. BOGGIA, O. DE GIBERT, S. JI, N. LOPPI, A. RAGANATO, R. VÁZQUEZ & J. TIEDEMANN. EAEL 2024.
- DEC 2023 [Why bother with geometry? On the relevance of linear decompositions of Transformer embeddings](#)
T. MICKUS & R. VÁZQUEZ. BLACKBOXNLP 2023 (EMNLP 2023).
- DEC 2023 [Grounded and Well-rounded: A Methodological Approach to the Study of Cross-modal and Cross-lingual Grounding](#)
T. MICKUS, E. ZOSA & D. PAPERNO. Findings of EMNLP 2023.
- JUL 2023 [„Mann” is to “Donna” as 「国王」 is to « Reine »: Adapting the Analogy Task for Multilingual and Contextual Embeddings.](#)
T. MICKUS, E. CALÓ, L. JACQMIN, D. PAPERNO & M. CONSTANT. *SEM 2023 (ACL 2023).
- JUL 2023 [So many design choices: Improving and interpreting neural agent communication in signaling games](#)
T. BERNARD & T. MICKUS. Findings of the ACL 2023.
- JUN 2023 [Definition Modeling: To Model Definitions. Generating Definitions With Little to No Semantics](#)
V. SEGONNE & T. MICKUS (equal contribution). IWCS 2023.
- MAY 2023 [Dozens of Translation Directions or Millions of Shared Parameters? Comparing Two Types of Multilinguality in Modular Machine Translation](#)
M. BOGGIA, S.-A. GRÖNROOS, N. A. LOPPI, T. MICKUS, A. RAGANATO, J. TIEDEMANN & R. VÁZQUEZ. NoDaLiDa 2023.
- DEC 2022 [Idiosyncratic Frequency as a Measure of Derivation vs Inflection](#)
M. COPOT, T. MICKUS & O. BONAMI. In *Journal of Language Modeling*.
- SEP 2022 [How to Dissect a Muppet: The Structure of Transformer Embedding Spaces.](#)
T. MICKUS, D. PAPERNO & M. CONSTANT. In *Transactions of the ACL*; EAEL 2023.
- JUL 2022 [Semeval-2022 Task 1: CODWOE – Comparing Dictionaries and Word Embeddings](#)
T. MICKUS, K. VAN DEEMTER, M. CONSTANT & D. PAPERNO. SemEval 2022 (NAACL 2022).
- DEC 2021 [About Neural Networks and Writing Definitions](#)
T. MICKUS, D. PAPERNO & M. CONSTANT; DSNA 2021.
- AUG 2021 [A Game Interface to Study Semantic Grounding in Text-Based Models](#)
T. MICKUS, M. CONSTANT & D. PAPERNO. Games & NLP (IEEE CoG).
- DEC 2020 [What Meaning-Form Correlation Has to Compose With](#)
T. MICKUS, T. BERNARD & D. PAPERNO. COLING 2020.
- JUN 2020 [Génération Automatique de Définitions pour le Français](#)
T. MICKUS, M. CONSTANT & D. PAPERNO. TALN 2020.
- JAN 2020 [What do you mean, BERT? Assessing BERT as a Distributional Semantics Model](#)
T. MICKUS, D. PAPERNO, M. CONSTANT & K. VAN DEEMTER. SCiL 2020.
- SEP 2019 [Mark my Word: a Sequence-to-sequence Approach to Definition Modeling](#)
T. MICKUS, D. PAPERNO & M. CONSTANT. DL4NLP workshop (NoDaLiDa 2019).
- JAN 2019 [Distributional Effects of Gender Contrasts Across Categories](#)
T. MICKUS, O. BONAMI & D. PAPERNO. SCiL 2019.