

TOMASZ KORBAK

Department of Informatics, University of Sussex, UK

tomekkorbak.com \diamond t.korbak@sussex.ac.uk

INTERESTS

reinforcement learning, natural language generation, control as inference, active inference, compositionality, AI alignment

FEATURED WORK

1. Kuciński, Ł., Miłoś, P., Korbak, T., Kołodziej, P. (2021). Emergence of compositionality in communication over a noisy channel. NeurIPS 2021.
2. Korbak, T. Elsahar, H., Kruszewski, G., Dymetman, M. (2021). Controlling Conditional Language Models with Distributional Policy Gradients. NeurIPS 2021 workshop on Controllable Generative Models.
3. Korbak, T. Elsahar, H., Dymetman, M. Kruszewski, G (2021). Energy-Based Models for Code Generation under Compilability Constraints. ACL 2021 workshop NLP4Prog.
4. Korbak, T., Zubek, J., Rączaszek-Leonardi, J. (2020). Measuring non-trivial compositionality in emergent communication. NeurIPS 2020 workshop “Talking to Strangers: Zero-Shot Emergent Communication”.
5. Główna, K., Niklewski, M., Korbak, T., Zubek, J., Rączaszek-Leonardi, J. (2020). Emergence of Action-grounded Compositional Communication. 42nd Annual Virtual Meeting of the Cognitive Science Society.
6. Korbak, T., Zubek, J., & Rączaszek-Leonardi, J. (2019). Developmentally motivated emergence of compositional communication via template transfer. NeurIPS 2019 workshop Emergent communication.

RESEARCH EXPERIENCE

**Department of Informatics,
University of Sussex**

PhD researcher
September 2020 – September 2023

Working on probabilistic approaches to control with Dr. Christopher Buckley and Prof. Anil Seth.

**Naver Labs Europe
Naver Corporation**

Research intern
February 2021 – August 2021

Working on energy-based models for controllable language generation with Marc Dymetman

**Human Interactivity and Language Lab,
Faculty of Psychology, University of Warsaw**

Research assistant
February 2019 – October 2020

Investigating compositional generalisation in neural networks with Prof. Joanna Rączaszek-Leonardi.

**Institute of Computer Science,
Polish Academy of Sciences**

Research intern
April 2017 – November 2017

Working on NLP tools for Polish as part of Clarin-PL project, funded by the European Commission.

INDUSTRIAL EXPERIENCE

Daftmobile , Senior Machine Learning Engineer	<i>August 2020 — February 2021</i>
Sigmoidal , Machine Learning Engineer	<i>June 2018 — August 2020</i>
Samsung R&D , Junior NLP Engineer	<i>April 2017 — December 2017</i>
Intelclinic , Python Developer	<i>December 2015 — March 2017</i>
Webinterpreter , Junior Python Developer Intern	<i>July 2015 — September 2015</i>

EDUCATION

PhD in Informatics , University of Sussex, UK	<i>2020 – 2023</i>
MSc in Cognitive Science , University of Warsaw, Poland	<i>2016 – 2019</i>
BSc in Cognitive Science , University of Warsaw, Poland	<i>2013 – 2016</i>
BAs in Philosophy , University of Warsaw, Poland	<i>2012 – 2015</i>

ADDITIONAL TRAINING

Diverse Intelligences Summer Institute , University of California, Los Angeles	<i>2020</i>
Bayesian Methods in Deep Learning , National Research University (Moscow)	<i>2018</i>
School of Pioneers (tech entrepreneurship workshops), University of Cambridge	<i>2018</i>
Computational Psychiatry Course , ETH (Zurich)	<i>2017</i>

JOURNAL PAPERS

1. Korbak, T. (2019). Computational enactivism under the free energy principle. *Synthese*.
2. Korbak, T. (2019). Unsupervised learning and the natural origins of content. *Avant*.
3. Korzeniowski, R., Rolczyński, R., Sadownik, P., Korbak, T. & Możejko, M. (2019). Exploiting Unsupervised Pre-training and Automated Feature Engineering for Low-resource Hate Speech Detection in Polish. *Proceedings of the PolEval 2019 Workshop*.
4. Korbak, T. & Żak, P. (2017). Fine-tuning Tree-LSTM for phrase-level sentiment classification on a Polish dependency treebank. *Proceedings of the 8th Language & Technology Conference (LTC 2017)*.
5. Korbak, T. (2015). Scaffolded Minds and the Evolution of Content in Signaling Pathways. *Studies in Logic, Grammar and Rhetoric*, 41 (54).

SKILLS

Python (web frameworks and data science ecosystem), C++, PyTorch, tensorflow, git, Docker, Kubernetes, slurm, cloud computing, GNU/Linux, L^AT_EX

AWARDS AND FELLOWSHIPS

Leverhulme Doctoral Scholarship (Leverhulme Trust)	<i>2020-2023</i>
Diverse Intelligences Summer Institute Fellowship (Templeton Foundation)	<i>2020</i>
Collegium Invisibile Fellowship	<i>2017 – present</i>
Minister of Science and Higher Education (Poland) scholarship for exceptional students	<i>2016</i>
Diamond grant award (Ministry of Science and Higher Education, Poland) (168 000 PLN)	<i>2016 – 2020</i>