Thomas H. Kober

tttthomasssss.github.io tkober@inf.ed.ac.uk github.com/tttthomasssss +44 7427 512313

Research Interest

My research is focused on building better natural language understanding systems. My work is based on the combination of distributional semantics, distributional composition and logical semantics to understand and model causation and consequence of actions in text. I furthermore have a strong interest in leveraging unsupervised, semi-supervised and active learning techniques to improve and debug natural language understanding models.

Current Role

2018–today **Postdoctoral Researcher**, *Institute for Language, Cognition and Computation, University of Edinburgh, UK*.
Working with Prof. Mark Steedman

Education

2014–2017 **PhD**, *Department of Informatics, University of Sussex, UK*, Supervised by Prof. David Weir & Dr. Julie Weeds.
PhD thesis: *Inferring Unobserved Co-occurrence Events in Anchored Packed Trees*, available from: `http://sro.sussex.ac.uk/75718/`

2011–2014 **BSc**, *Computer Science & Artificial Intelligence, University of Sussex, UK, First Class with Honours*.
BSc thesis: *Scaling Semi-Supervised Multinomial Naïve Bayes*, available from: `http://www.sussex.ac.uk/informatics/documents/kober-proj.pdf`

Awards & Prizes

2014–2017 Fully Funded Informatics PhD Award, University of Sussex

2014 Brandwatch Prize for the Best Computer Science & Artificial Intelligence BSc Final Year Project.

2013 American Express Award for the Best second year Undergraduate Informatics Student.

Teaching & Supervision

2018–today **MSc Project Supervision**, University of Edinburgh.
Qiwei Peng, *Can compositional distributional semantics capture the aspectual class of clauses?*, supervised together with Prof. Bonnie Webber; Qiwei is an incoming PhD student at the University of Sussex this year.
Yulong Chen, *Identifying causal "until" relations in instructions*, supervised together with Prof. Bonnie Webber; Yulong is currently a PhD student at Westlake University.

2014–2017 **Teaching Assistant for Natural Language Engineering (G5119)**, University of Sussex.

Research Projects

2018–today **Semantax**, University of Edinburgh.
Our goal is to build a robust form-independent semantics from scratch. The backbone of the new semantics are entailment graphs as inference mechanism on the basis of distributional and logical semantic techniques. One of my major tasks is building better entailment graphs by modelling fine-grained semantic properties of verbs such as tense and aspect, that govern many causal and consequential relations in a discourse.

2015–2017 **A Theory of Composition for Distributional Semantics**, University of Sussex.
We introduced *Anchored Packed Trees*, a novel framework for modelling composition in distributional semantic models, leveraging dependency trees to bridge the syntax-semantics gap. Our model is fully interpretable and allows the study of the distributional semantics of phrases and sentences in the same space as individual lexemes.

2013–2015 **Method51 for Mining Insight from Social Media Datasets**, University of Sussex.
Method51 is an application that combines active learning methods together with a Naïve Bayes classifier and a range of semi-supervised learning algorithms (e.g. Expectation-Maximization, Semi-Supervised Frequency Estimate and Feature Marginals) to allow a user to perform agile analyses on large and diverse social media datasets.

Professional Experience

2014–2015 **Co-Founder of paprr.co (RIP)**, Brighton, UK.
paprr.co was a platform about collectively annotating and discussing scientific publications. Our vision was to be a robust tool for active academic researchers to manage the vast amount of papers released every day, as well as a tool for undergraduate and graduate students to more easily find relevant papers for their research and projects.

2013–2014 **Co-Founder of napse.co (RIP)**, London, UK.
napse.co was a platform aiming to connect startups in the life sciences and related areas with new university graduates, coming from both, science and business/marketing backgrounds, as well as experienced industry experts (e.g. senior scientists, patent attorneys, VC's, etc.), and providing a central platform for startups to find lab space and other technical and non-technical equipment.

2010–2011 **Software Engineer at mquadr.at**, Vienna, Austria.
My responsibilities at mquadr.at were to design, develop and maintain several iOS and Mac OS applications, aimed at setting up and managing the connection between a router and the web as well as between the router and the end-user devices (computers, smartphones, tablets, TV's, etc.).

2007–2010 **Software Engineer at Comit AG (now part of Swisscom)**, Zurich, Switzerland.
At Comit, I was part of the internet banking team which supplied Swiss private banks with internet banking software. My responsibilities included the design development and maintenance of the actual banking functionality in the backend, e.g. I wrote the foreign exchange and money market modules and contributed to several others, as well as the implementation and customisation of these modules for clients in numerous projects on site.

References

Prof. Mark Steedman — Professor of Cognitive Science, ILCC, University of Edinburgh, steedman@ed.ac.uk

Prof. David Weir — Professor of Computer Science, University of Sussex, d.j.weir@sussex.ac.uk

Dr. Julie Weeds — Lecturer in Informatics, University of Sussex, j.e.weeds@sussex.ac.uk

Service

2018–today **Co-Organiser:** PyData Edinburgh Meetup Group (`https://www.meetup.com/PyData-Edinburgh/`)

2017–today **Program Committee:** *ACL, EMNLP, COLING, CoNLL, various SRWs

2017–today **ML/NLP Advisor:** Teebly (`https://teebly.co/`)

Publications

Thomas Kober, Sander Bijl de Vroe, Mark Steedman. Temporal and Aspectual Entailment. *In Proceedings of IWCS*, pages 103–119, 2019.

Thomas Kober. Inferring Unobserved Co-occurrence Events in Anchored Packed Trees, *PhD Thesis*, University of Sussex, 2018.

Thomas Kober, Julie Weeds, Jeremy Reffin and David Weir. Improving Semantic Composition with Offset Inference. *In Proceedings of ACL*, pages 433–440, 2017.

Thomas Kober, Julie Weeds, John Wilkie, Jeremy Reffin and David Weir. One Representation per Word — Does it make Sense for Composition? *In Proceedings of the 1st Workshop on Sense, Concept and Entity Representations and their Applications*, pages 79–90, 2017.

Julie Weeds, Thomas Kober, Jeremy Reffin and David Weir. When a Red Herring is Not a Red Herring: Using Compositional Methods to Detect Non-Compositional Phrases. *In Proceedings of EACL*, pages 529–534, 2017.

Thomas Kober, Julie Weeds, Jeremy Reffin and David Weir. Improving Sparse Word Representations with Distributional Inference for Semantic Composition. *In Proceedings of EMNLP*, pages 1691–1702, 2016.

Miroslav Batchkarov, Thomas Kober, Jeremy Reffin, Julie Weeds and David Weir. A critique of word similarity as a method for evaluating distributional semantic models. *In Proceedings of the 1st Workshop on Evaluating Vector Space Representations for NLP*, pages 7–12, 2016.

Thomas Kober and David Weir. Optimising Agile Social Media Analysis. In *Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis (WASSA 2015)*, pages 31–40, 2015.