WILKER FERREIRA AZIZ
November 15, 2021

Institute for Logic, Language and Computation
University of Amsterdam
Science Park 107, F2.11
Amsterdam, 1098 XG, Netherlands

w.aziz@uva.nl
https://wilkeraziz.github.io
https://probabll.github.io
Tel: +31 (0) 631991009

My research is concerned with enabling computers to understand and process human languages. To that end, I build upon results from probability theory, Bayesian statistics, optimisation, formal languages, and linguistics. I develop machine learning models as well as techniques to estimate such models efficiently from data. Because availability of supervision is a major bottleneck to statistical learning, my research has a strong component of unsupervised learning and other forms of latent variable modelling.

Education

2010–2014 **Ph.D. Computational Linguistics**
Research Institute in Information and Language Processing University of Wolverhampton
Thesis: Exact Sampling and Optimisation in Statistical Machine Translation
Supervisors: Prof. Dr. Lucia Specia, Dr. Marc Dymetman, Prof. Dr. Ruslan Mitkov
Summary: I introduce an approach to exact optimisation and sampling based on adaptive rejection sampling which addresses challenges in global optimisation and unbiased sampling in high-dimensional discrete spaces. In this view, an intractable goal distribution is upperbounded by a tractable proxy distribution which is then incrementally refined to be closer to the goal.

2005–2010 **B.Sc. Computer Engineering**
Escola de Engenharia de São Carlos - Universidade Estadual de São Paulo (USP)

Employment

01/2019–present **Assistant Professor, Institute for Logic, Language and Computation, Universiteit van Amsterdam, Netherlands**
Summary: My research group focuses on probabilistic models of language with a strong focus on unsupervised learning and latent variable modelling.

01/2015–12/2018 **Research Associate, Institute for Logic, Language and Computation, Universiteit van Amsterdam, Netherlands**
Summary: I joined the Statistical Language Processing and Learning Lab led by Professor Khalil Sima'an in January 2015 where I worked on several aspects of machine translation (e.g. word alignment, word reordering, and morphological analysis and generation) and paraphrasing employing log-linear, Bayesian, and deep generative models.

11/2013–12/2014 **Research Associate, Department of Computer Science, University of Sheffield, UK**
Summary: My work was funded by EPSRC under the MODIST (MOdelling DIscourse in Statistical Translation) project led by Prof. Dr. Lucia Specia. Discourse information typically requires nonlocal forms of parameterisation. I developed better decoding algorithms for SMT aiming at incorporating global features, particularly, I worked on a lazy incorporation of nonlocal parameterisation using a form of adaptive rejection sampling.

Services

I regularly serve on the programme committee of all *ACL events since around 2012, as well as WMT, CoNLL, and COLING, having served as area chair for *ML for NLP* (*SEM 2019, ACL 2020) and *MT* (ACL 2021). I am also active member of ML events and since 2019 I serve on the program committees of NeurIPS, ICML, and ICLR (including as area chair since 2020). I am a standing reviewer for Transactions of ACL (journal) and ACL rolling reviewing (ARR, for which I am also an action editor). Since 2020, I am a member of the ELLIS society (and since 2021, ELLIS faculty in the Amsterdam unit).

Selected Publications

I have dedicated most of the last 4 years to probabilistic approaches to NLP and learning with latent variables, which, together with my collaborators and students, I explore to build models of increased transparency [12], to expose knobs that control how predictions are formed in language generation [7], and to change the implicit knowledge of large language models [3]. Transparency and control are promoted via structured representations whose unsupervised induction poses challenges to deep learning, hence I have invested my expertise in overcoming those challenges [14, 12, 8]. I also contribute to core machine learning, in particular, density estimation [10], gradient estimation [8], and probabilistic deep learning [9, 2]. My interests in NLP include language generation [6]), decoding [1], learning from multiple modalities [13]), and knowledge intensive tasks [4, 11]. My most recent interest is automatic criticism of machine-generated text, which I approach through the lens of Bayesian data analysis—my recent work on machine translation [5] has been granted a best paper award.

The following open access papers reflect my expertise and wide range of research. I underline student authors who I directly supervised during the project.

1. Bryan Eikema, **Wilker Aziz**. Sampling-Based Minimum Bayes Risk Decoding for Neural Machine Translation. *Pre-print*, 2021.
2. António Farinhas, **Wilker Aziz**, Vlad Niculae, and André Martins. Sparse Communication via Mixed Distributions. *Pre-print*, 2021.
3. Nicola De Cao, **Wilker Aziz**, Ivan Titov. Editing Factual Knowledge in Language Models. In: *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2021.
4. Nicola De Cao, **Wilker Aziz**, Ivan Titov. Highly Parallel Autoregressive Entity Linking with Discriminative Correction. In: *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2021.
5. Bryan Eikema, **Wilker Aziz**. Is MAP Decoding All You Need? The Inadequacy of the Mode in Neural Machine Translation. In: *Proceedings of the 28th International Conference on Computational Linguistics (COLING)*, 2020.
6. Tom Pelsmaeker, **Wilker Aziz**. Effective Estimation of Deep Generative Language Models. In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2020.
7. Duygu Ataman, **W. Aziz**, Alexandra Birch. A Latent Morphology Model for Open-Vocabulary Neural Machine Translation. In: *International Conference on Learning Representations (ICLR)*, 2020.
8. Gonçalo Correia, Vlad Niculae, **Wilker Aziz**, André Martins. Efficient Marginalization of Discrete and Structured Latent Variables via Sparsity. In: *Advances in Neural Information Processing (NeurIPS)*, 2020.
9. Nicola De Cao, **Wilker Aziz**. The Power Spherical distribution. In: *ICML INNF+ Workshop*, 2020.
10. Nicola De Cao, **Wilker Aziz**, Ivan Titov. Block Neural Autoregressive Flows. In: *Proceedings of the Conference on Uncertainty in Artificial Intelligence (UAI)*, 2019.
11. Nicola De Cao, **Wilker Aziz** and Ivan Titov. Question Answering by Reasoning Across Documents with Graph Convolutional Networks. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, 2019.
12. Jasmijn Bastings, **Wilker Aziz**, Ivan Titov. Interpretable Neural Predictions with Differentiable Binary Variables. In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2019.
13. Iacer Calixto, Miguel Rios, **W. Aziz**. Latent Variable Model for Multi-modal Translation. In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2019.
14. Philip Schulz, **Wilker Aziz**, Trevor Cohn. A Stochastic Decoder for Neural Machine Translation. In: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2018.
15. **W. Aziz**, Philip Schulz. Variational Inference and Deep Generative Models. In: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL): Tutorial Abstracts*, 2018.

For a complete list dating back to my PhD, refer to `https://wilkeraziz.github.io/papers/`.

Teaching

I am highly committed to education. Since 2015, I offer courses at UvA covering topics such as probability theory, unsupervised learning, Bayesian methods for NLP, and deep learning. I contribute to the bachelor's in AI (BAI), master's in AI (MAI), and master's of logic (MoL):

- *natural language models and interfaces* (BAI 2018–now, coordinator),
- *basic probability theory and programming* (MoL 2019–2020, coordinator),
- *natural language processing 1* (MAI 2021, co-lecturer with Dr. Sandro Pezzelle),
- *natural language processing 2* (MAI 2015–2017, coordinator),
- *unsupervised language learning* (MAI 2018, co-lecturer with Dr. Ekaterina Shutova),
- *deep learning for natural language processing* (MAI 2020–2021, co-lecturer with Prof. Christof Monz),
- and, with a colleague (Dr. Efstratios Gavves), I am developing *deep learning 2* (MAI, 2022).

I have developed a tutorial on variational inference and deep generative models for NLP audiences (15). The tutorial has been hosted by universities (e.g. Edinburgh, Melbourne, Heidelberg, IST Lisbon, Alicante), companies (e.g. Amazon, Naver Labs, Yandex), schools (e.g., DTU-DIKU school on generative models, ProbAI2021), and international conferences (ACL 2018). See https://vitutorial.github.io for all of the material.

Supervision

At the PhD level, I am the main supervisor of Bryan Eikema (UvA) and co-supervisor of Nicola De Cao (UvA, jointly with Ivan Titov) and Gonçalo Correia (IST Lisbon, jointly with André Martins and Vlad Niculae). I have been co-supervisor of Jasmijn Bastings (UvA, now Google Brain) and main supervisor of Philip Schulz (UvA, now Amazon), who completed their PhDs successfully in 2020. My expertise contributed to the scope and methodology of their theses, which concern unsupervised learning (all), Bayesian modelling (Philip's, Jasmijn's and Bryan's), approximate inference via MCMC (Philip's and Nicola's) or variational inference (all), natural language generation (all but Nicola's), deep learning for NLP (all). Prior to that I was co-supervisor of Miloš Stanojević (UvA, now Edinburgh) and Joachim Daiber (UvA, now Apple) who completed their PhDs in 2017 and 2018. Since 2017, I have supervised 12 MSc theses. For more information and links to theses, please refer to https://wilkeraziz.github.io/supervision/.

Personal

I am a native Portuguese speaker who most of the time expresses himself in English (fluent) or Italian (intermediate), and who used to know some French. I play the guitar and try to keep Brazilian bossa alive. I have a thing for food, sun and mixology. I bike everywhere, under sun or rain, and I swim when I can (but not in the Dutch canals, they are just too murky). I love pets (all of them!), but I find dogs objectively superior to cats. If you know me, you probably know Pepper. If you do not know Pepper, you should meet her: https://wilkeraziz.github.io/pepper/.