

Harish Tayyar Madabushi

Bath, BA2 | www.HarishTayyarMadabushi.com

scholar.google.com/citations?user=EHOS_5QAAAAJ | <https://www.linkedin.com/in/harishtayyarmadabushi/>

Research Interests: Natural Language Processing, Deep Contextual Meaning Representations, Question Answering, Integrating Cognitive and Psycholinguistic Information into Deep Learning

RELEVANT EXPERIENCE

Lecturer in Artificial Intelligence

Sept 2022 - present

School of Computer Science, University of Birmingham

- Research focused on explainable methods in NLP; Teach NLP and Machine Learning.

Research Associate in Machine Learning for NLP

March 2021 - Aug 2022

Department of Computer Science, University of Sheffield

- Research on transparent models to generate effective cross-lingual representations of idiomatic language with Professor Aline Villavicencio and Dr Carolina Scarton.
- Principal organiser of [SemEval 2022 Task 2](#).
- Led the writing of two successful proposals for tutorials to be given at [LREC and COLING](#).
- Awarded the [Alan Turing Institute Postdoctoral Enrichment Award](#).
- Co-Chair for events at Women+ @DCS (part of Sheffield EDI initiative).
- Co-supervise one PhD student.
- Nominated as [Think Ahead Outstanding Postdoc](#).
- Organised the [first Workshop of the MWE Special Interest Group](#) at Sheffield.

Honorary Research Fellow

March 2021 - present

School of Computer Science, University of Birmingham

- Currently co-supervise two PhD students. Research collaboration and publications.

Lecturer (Teaching Focused), MSc Director

Oct 2018 - March 2021

School of Computer Science, University of Birmingham

- Fellow of the Higher Education Academy (HEA) and nominated for Outstanding Teaching Award across the College of Engineering and Physical Sciences three years in a row.
- Responsible for the design, delivery and assessment of three MSc courses to a cohort of over 400 students and an undergraduate course to close to 300 students.
- Taught Artificial Intelligence, Introduction to Computer Science and Databases.
- Responsible for the smooth functioning of the MSc Computer Science programme consisting of four hundred students.
- Sole supervisor to 19 final year and 22 masters' student projects that led to seven publications. Co-supervisor to two PhDs students
- [Talk on autonomous cars](#) to the general public as part of the Pint of Science initiative.

Research Associate on Diachronic Embeddings (Bloomberg funded)

Oct 2017- Sept 2018

School of Mathematical Sciences, University of Nottingham

PhD [Thesis Link](#) (EPSRC funded)

Sept 2014 - Jul 2019

School of Computer Science, University of Birmingham

Co-Founder and Chief Technology Officer

2008 - 2014

- Built a Singapore based research-focused product company, *Webfluenz*, that provided brand insights to clients around the world. The system used NLP methods for identifying influencers and sentiment associated with brands. Acquired by *Mu Sigma Business Solutions*.

SELECTED PUBLICATIONS

Please see below for a complete list of publications.

Harish Tayyar Madabushi, Dagmar Divjak, Petar Milin, 2022., Abstraction not Memory: BERT and the English Article System. To appear In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* (<https://arxiv.org/abs/2206.04184>).

Harish Tayyar Madabushi, Edward Gow-Smith, Marcos Garcia, Carolina Scarton, Marco Idiart, and Aline Villavicencio. 2022. SemEval-2022 Task 2: Multilingual Idiomaticity Detection and Sentence Embedding. To appear In *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval2022)*. Association for Computational Linguistics (<https://arxiv.org/abs/2204.10050>).

Harish Tayyar Madabushi, Edward Gow-Smith, Carolina Scarton and Aline Villavicencio, 2021, November. [AStitchInLanguageModels: Dataset and Methods for the Exploration of Idiomaticity in Pre-Trained Language Models](#). In *Findings of the Association for Computational Linguistics: EMNLP 2021* (pp. 3464-3477).

Harish Tayyar Madabushi, Laurence Romain, Dagmar Divjak and Petar Milin, 2020, December. [CxGBERT: BERT meets Construction Grammar](#). In *Proceedings of COLING 2020, the 28th International Conference on Computational Linguistics: Technical Papers* (pp. 4020-4032).

Harish Tayyar Madabushi, Mark Lee, and John Barnden, 2018 August. [Integrating Question Classification and Deep Learning for Improved Answer Selection](#). In *Proceedings of COLING 2018, the 27th International Conference on Computational Linguistics: Technical Papers* (pp. 3283-3294).

[[Achieved state-of-the-art results in answer selection](#)]

RESEARCH FUNDING AND INDUSTRY COLLABORATION

- [2022, Sheffield] [Awarded](#) **The Alan Turing Institute Postdoctoral Enrichment Award** for research visit to Technical University Darmstadt (£2,000).
- [2021, Sheffield] Was Involved in the development of an **Innovate UK grant proposal** with ZOO Digital.
- [2019 to present] Regularly involved in conversations with wyona.com on chat bots.
- [2019, Birmingham] [Awarded](#) the **Ramsay Research Support Fund** of £3,300 to study native speakers' ability to use correct articles in speech.
- [2017, Birmingham] Worked on a Project on Exploring Public Health Through **Social Media Conversations** for the **Birmingham City Council**, which involved tracking and analysing social media conversations from Birmingham over a three-month period.
- [2016, Birmingham] Worked with the **CEO of IdeaLabz Investments**, U.K. for six months exploring ways in which **NLP can be used to better and more quickly understand quarterly reports and earnings call transcripts**.
- [2014, Birmingham] **EPSRC funded PhD** at the University of Birmingham.

TEACHING AND SUPERVISION

Fellow of the Higher Education Academy (HEA).

University of Sheffield, 2021 - 2022

- Second supervisor to one PhD student at Sheffield.

- Continue to be second supervisor to two PhD students at the University of Birmingham.
- Second supervisor to five MSc students.

University of Birmingham, 2019 - 2021

- Taught **Artificial Intelligence**, including **Machine Learning** and **Deep Learning**.
- Design, delivery and assessment of **two** MSc courses to a cohort of over **four hundred students** and an undergraduate course to close to **three hundred students**.
- Was nominated for the **Outstanding Teaching Award** across the College of Engineering and Physical Sciences **three years in a row**.
- Sole Supervisor to **19 final year and 22 masters' student projects** which led to **seven publications**.
- **Four MSc students** who did their projects with me **have gone on to do PhDs** in fields related to Natural Language Processing.
- Second supervisor to **two PhD students** working on sentiment analysis and text summarisation which led to **three publications**.
- Updated modules to be delivered and evaluated during **COVID-19 pandemic**.

Selected List of Modules Taught

- Artificial Intelligence (BSc and MSc)
 - **Module Lead for MSc**. Took over the machine learning section of a **module which required the design and development of lectures and assignments** from scratch..
 - Years taught: 2018, 2019, 2020 (Average cohort Size: 300)
- Introduction to Computer Science (MSc)
 - **Module lead and solely responsible** for delivery and evaluation of the module. Heavily updated the module and added new lecture materials and assignments.
 - Years taught: 2018, 2019, 2020 (Average cohort Size: 250, peaked at ~400)
- Databases (MSc)
 - **Module lead** and responsible for delivery and evaluation of the module.
 - Years taught: 2019, 2020 (Average cohort Size: 250, peaked at ~400)

TALKS AND PUBLIC ENGAGEMENT

- [2021, Dominican Republic] Talk presenting paper at EMNLP 2021.
- [2021, Zurich] **Invited talk** at Modelling Constructional Variation and Change, University of Zurich: [BERT "knows" Construction Grammar: Evidence and Implications](#).
- [2021, Tübingen] Talk at International Conference on Error-Driven Learning in Language: Less is More.
- [2020, London] Talk at Alan Turing Institute, Natural Language Processing Interest Group: [Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data](#).
- [2019, London] Talk at Alan Turing Institute, Natural Language Processing Interest Group: [Cost-Sensitive BERT for Generalisable Sentence Classification with Imbalanced Data](#).
- [2019, Birmingham] [Talk on autonomous cars](#) to the general public as part of the Pint of Science initiative.
- [2018, United States] Talk presenting paper at COLING 2018, Santa Fe, U.S.A
- [2016, Japan] Presented a poster at COLING 2016, Osaka, Japan.

ORGANISING COMMITTEES

- [2021 - 2022] **Principal organiser** of [SemEval 2022 Task 2 "Multilingual Idiomaticity Detection and Sentence Embedding"](#)

- Central to the conception and design of the task, including the formulation of a novel task aimed at **evaluating models' ability to capture idiomaticity in sentence embeddings**.
- Actively involved in creating the **task proposal**.
- Responsible for the creation of the dataset, setup of the Codalab competition, promotion of the task and establishing baselines.
- [2019] Collaborated with researchers at the University of Nottingham in organising a panel discussion on NLP for mental health at [HealTAC-2019](#).

LEADERSHIP AND COLLABORATION

MSc Director, University of Birmingham (Oct 2019 - March 2021):

- Responsible for the smooth functioning of the MSc Computer Science programme consisting of **four hundred students**.
- Coordinating the course across multiple modules leads to ensure reasonable workload, avoidance of timetabling conflicts and availability of labs and lecture theatres for the large student group, so as to maintain high student satisfaction.
- Guiding students through the selection of optional modules and providing mentorship in the form of progress tutorials where required.
- Was tutor for both MSc and BSc students, including during the COVID-19 pandemic.
- Attended **UG and PG open days** with potential students.

Research Collaboration

- Collaborating with researchers from the Universidade Federal do Rio Grande do Sul (Brazil), University of Santiago de Compostela (Spain), Aix-Marseille University (France).
- Collaborated with a researcher from the University of Canterbury, New Zealand, on the convergence of learned Construction Grammars.
- Led an interdisciplinary team consisting of researchers, including a sociologist from the University of Warwick, in creating methods for the detection of propaganda in the news. This work was ranked second place out of forty participating teams.
- Collaborated with a team from the University of Arizona on research into 4th grade question answering based on question classification.

Professional activities

- Co-Chair of events at [women+ @DCS](#) (Part of Sheffield EDI Initiative).
- Organise two reading groups at the University of Sheffield: the NLP Reading Group and the reading group of the MWE Special Interest Group.
- Review papers for Frontiers of Artificial Intelligence and IEEE Transactions on Neural Networks and Learning Systems.
- **Senior Area Chair** for SemEval 2022
- Chaired session at COLING 2020.

INDUSTRY EXPERIENCE

Mar 2014-
Aug 2014

Associate Director
Mu Sigma Business Solutions

Bangalore, India

Following the successful acquisition of Webfluenz PTE by Mu Sigma Business Solutions, the latter continued to thrive and expand as a data analytics innovator with offices based in Chicago and Bangalore, employing over 3,000 decision scientists with experience across 10 verticals.

- Leading the shift in focus of the company from a service-based approach to creating big data products, which would be accessible to a broader and more varied range of customers, ensuring the smooth transition of the Webfluenz technology team into Mu Sigma.
- Managing the integration of Webfluenz technology offerings with those of Mu Sigma and identifying opportunities for the use of Webfluenz technologies in existing Mu Sigma projects.

Nov 2009- **Co-Founder & Chief Technology Officer** Bangalore, India
Feb 2014 *Webfluenz PTE*

Co-Founder of this product-based **social media marketing company**, initially established as “4am Design & Technology Labs”, a three-person operation in India. Rebranded as Webfluenz, the company expanded to nearly 50 people, shifted its headquarters to Singapore and was subsequently acquired, in early 2014, by Mu Sigma Business Solutions. Webfluenz was listed among the top fifteen tools in Goldbach Interactive’s Social Media Monitoring Tool Report during 2012 and was a Red Herring 100 Asia finalist for two consecutive years in 2012 and 2013.

- Responsible for growing the engineering team to approximately twenty-five diversely skilled personnel.
- Driving the creation of proprietary Webfluenz Intelligence and Data Gathering systems by combining established computer science concepts with continuously evolving technologies.
- Responsible for all technology matters coupled with the overall direction of the company in the capacity as Co-Founder.

Aug 2008- **Co-Founder** Bangalore, India
Sept 2009 *4am Design & Technology Labs*

- Capturing and understanding user needs and effectively translating requirements into workable solutions.
- Recruiting and training the engineering team.
- Providing guidance and mentoring to the team to enable optimum performances and to drive technical capabilities.

May 2007- **Team Leader** Bombay, India
Apr 2008 *Pinstorm Technologies*

- Responsible for establishing the R&D division effectively creating a technology platform for online advertising. Hiring and training appropriately qualified and skilled engineers to create such a platform.

EDUCATION & ADDITIONAL SKILLS

2005-2007	Master of Science (Grade 9.06/10.00) <i>Chennai Mathematical Institute</i>	<i>Chennai, India</i>
2002-2005	Bachelor of Science (Grade 92% Majors) <i>Loyola College</i>	<i>Chennai, India</i>

Languages: fluent English & Telugu; spoken Tamil, Kannada & Hindi

Tech Skills: proficient in Python, Perl, C, C++, PyTorch, HTML, CSS, JavaScript, MySQL, Emacs, LaTeX, Git, SVN, Linux & Amazon Web Services; familiar with Java R & Tensorflow

FULL LIST OF PUBLICATIONS (Google Scholar: scholar.google.com/citations?user=EHOS_5QAAAAAJ)

ACCEPTED (Refereed conference publications)

Harish Tayyar Madabushi, Dagmar Divjak, Petar Milin, 2022,. Abstraction not Memory: BERT and the English Article System. To appear In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* (<https://arxiv.org/abs/2206.04184>).

Harish Tayyar Madabushi, Edward Gow-Smith, Marcos Garcia, Carolina Scarton, Marco Idiart, and Aline Villavicencio. 2022. "SemEval-2022 Task 2: Multilingual Idiomaticity Detection and Sentence Embedding." Accepted for publication In *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*. (<https://arxiv.org/abs/2204.10050>)

Phelps, D., Fan, X.R., Gow-Smith, E., Tayyar Madabushi, H., Scarton, C. and Villavicencio, A., 2022. Sample Efficient Approaches for Idiomaticity Detection. Accepted for publication In *Proceedings of the 17th Workshop on Multiword Expressions (MWE 2021)*. (<https://arxiv.org/abs/2205.11306>)

PUBLISHED (Refereed conference publications)

Jonathan Dunn and Harish Tayyar Madabushi. 2021. [Learned Construction Grammars Converge Across Registers Given Increased Exposure](#). In *Proceedings of the 25th Conference on Computational Natural Language Learning*, pages 268–278, Online. Association for Computational Linguistics.

Harish Tayyar Madabushi, Edward Gow-Smith, Carolina Scarton, and Aline Villavicencio. 2021. [AStitchInLanguageModels: Dataset and Methods for the Exploration of Idiomaticity in Pre-Trained Language Models](#). In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 3464–3477, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Peter Vickers, Rosa Wainwright, Harish Tayyar Madabushi, and Aline Villavicencio. 2021. [CogNLP-Sheffield at CMCL 2021 Shared Task: Blending Cognitively Inspired Features with Transformer-based Language Models for Predicting Eye Tracking Patterns](#). In *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics*, pages 125–133, Online. Association for Computational Linguistics.

Wei Li, Harish Tayyar Madabushi, and Mark Lee. 2021. [UoB_UK at SemEval 2021 Task 2: Zero-Shot and Few-Shot Learning for Multi-lingual and Cross-lingual Word Sense Disambiguation..](#) In *Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021)*, pages 738–742, Online. Association for Computational Linguistics.

Erik Yan and Harish Tayyar Madabushi. 2021. [UoB at SemEval-2021 Task 5: Extending Pre-Trained Language Models to Include Task and Domain-Specific Information for Toxic Span Prediction](#). In *Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021)*, pages 243–248, Online. Association for Computational Linguistics.

Frances Adriana Laureano De Leon, Harish Tayyar Madabushi, and Mark Lee. 2021. [UoB at ProfNER 2021: Data Augmentation for Classification Using Machine Translation](#). In *Proceedings of the Sixth Social Media Mining for Health (#SMM4H) Workshop and Shared Task*, pages 115–117, Mexico City, Mexico. Association for Computational Linguistics.

Harish Tayyar Madabushi, Laurence Romain, Dagmar Divjak, and Petar Milin. 2020. [CxGBERT: BERT meets Construction Grammar](#). In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 4020–4032, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Anushka Prakash and Harish Tayyar Madabushi. 2020. [Incorporating Count-Based Features into Pre-Trained Models for Improved Stance Detection](#). In *Proceedings of the 3rd NLP4IF Workshop on NLP for Internet Freedom: Censorship, Disinformation, and Propaganda*, pages 22–32, Barcelona, Spain (Online). International Committee on Computational Linguistics (ICCL).

Kyle Reed and Harish Tayyar Madabushi. 2020. [Faster BERT-based re-ranking through Candidate Passage Extraction](#). In *Proceedings of The Twenty-Ninth Text REtrieval Conference (TREC 2020)*

Calum Perrio and Harish Tayyar Madabushi. 2020. [CXP949 at WNUT-2020 Task 2: Extracting Informative COVID-19 Tweets - RoBERTa Ensembles and The Continued Relevance of Handcrafted Features](#). In *Proceedings of the Sixth Workshop on Noisy User-generated Text (W-NUT 2020)*, pages 352–358, Online. Association for Computational Linguistics.

Wah Meng Lim and Harish Tayyar Madabushi. 2020. [UoB at SemEval-2020 Task 12: Boosting BERT with Corpus Level Information](#). In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pages 2216–2221, Barcelona (online). International Committee for Computational Linguistics.

Eleri Sarsfield and Harish Tayyar Madabushi. 2020. [UoB at SemEval-2020 Task 1: Automatic Identification of Novel Word Senses](#). In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pages 239–245, Barcelona (online). International Committee for Computational Linguistics.

Ghadi Alnafesah, Harish Tayyar Madabushi, and Mark Lee. 2020. [Augmenting Neural Metaphor Detection with Concreteness](#). In *Proceedings of the Second Workshop on Figurative Language Processing*, pages 204–210, Online. Association for Computational Linguistics.

Frances Adriana Laureano De Leon, Florimond Guéniat, and Harish Tayyar Madabushi. 2020. [CS-Embed at SemEval-2020 Task 9: The Effectiveness of Code-switched Word Embeddings for](#)

[Sentiment Analysis](#). In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pages 922–927, Barcelona (online). International Committee for Computational Linguistics.

Alexander Crosby and Harish Tayyar Madabushi, 2020. [UoB at AI-SOCO 2020: Approaches to Source Code Classification and the Surprising Power of n-grams](#). In *Working Notes of FIRE 2020 - Forum for Information Retrieval Evaluation*, Proceedings of CEUR, pages. 677–693, Hyderabad, India, December 16-20, 2020

Shah, B. and Tayyar Madabushi, H., 2020, October. [Efficient Brain Tumour Segmentation Using Co-registered Data and Ensembles of Specialised Learners](#). In International MICCAI Brainlesion Workshop (pp. 15-29). Springer, Cham.

Harish Tayyar Madabushi, Elena Kochkina, and Michael Castelle. 2019. [Cost-Sensitive BERT for Generalisable Sentence Classification on Imbalanced Data](#). In *Proceedings of the Second Workshop on Natural Language Processing for Internet Freedom: Censorship, Disinformation, and Propaganda*, pages 125–134, Hong Kong, China. Association for Computational Linguistics.

Dongfang Xu, Peter Jansen, Jaycie Martin, Zhengnan Xie, Vikas Yadav, Harish Tayyar Madabushi, Oyvind Tafjord, and Peter Clark. 2020. [Multi-class Hierarchical Question Classification for Multiple Choice Science Exams](#). In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 5370–5382, Marseille, France. European Language Resources Association.

Harish Tayyar Madabushi, Mark Lee, and John Barnden. 2018. [Integrating Question Classification and Deep Learning for Improved Answer Selection](#). In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 3283–3294, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Harish Tayyar Madabushi and Mark Lee. 2016. [High Accuracy Rule-based Question Classification using Question Syntax and Semantics](#). In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 1220–1230, Osaka, Japan. The COLING 2016 Organizing Committee.

Harish Tayyar Madabushi, Mark Buhagiar, and Mark Lee. 2016. [UoB-UK at SemEval-2016 Task 1: A Flexible and Extendable System for Semantic Text Similarity using Types, Surprise and Phrase Linking](#). In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 680–685, San Diego, California. Association for Computational Linguistics.

[Preprints \(Not refereed\)](#)

Petar Milin, Harish Tayyar Madabushi, Michael Croucher, Dagmar Divjak, 2020. [Keeping it simple: Implementation and performance of the proto-principle of adaptation and learning in the language sciences](#). arXiv preprint arXiv:2003.03813.