

Marc Allasonnière-Tang, Ph.D.

I

✉ marc.allasonniere-tang@mnhn.fr

🐙 github.com/marctang

🆔 orcid 0000-0002-9057-642X

🌐 https://marctang.github.io/

> Digital Humanities

> Linguistic Typology

> Computational Linguistics

> Phylogenetics

Employment History

- 2021 – present ■ **CNRS Researcher.** EA (Ecological Anthropology, UMR 7206) lab at the Muséum National d'Histoire Naturelle (MNHN) in Paris, France. (Also associated to the Lab Dynamics of Language, UMR 5596, Lyon)
- 2019 – 2021 ■ **Postdoctoral Researcher.** DDL (Dynamics of Language, UMR 5596) lab of University Lumière Lyon 2, France. (80% research – 20% Teaching/Supervision)
- 2016 – 2019 ■ **Ph.D. Student.** Department of Linguistics and Philology, Uppsala University, Sweden. Defence completed on the 9th of March 2019. (80% research – 20% teaching, full-time funded working contract)
- 2013 – 2015 ■ **French Instructor.** Chinese Institute of European Languages, Taiwan. (Group courses 5–20 members, 30 hours/week, level A1–C2)
- **Research Assistant.** Syntax/Phonology lab, National ChengChi University, Taiwan. (Experiment design, database maintenance, and data analysis)
- 2011 – 2012 ■ **Product Manager.** North Africa division, Asustek Computer, Taiwan. (Product planning, sales and marketing for notebooks and tablets)



Education

- 2016 – 2019 ■ **Ph.D., Uppsala University, Sweden/ National Institute of Oriental Languages and Civilizations, France** in Linguistics 📄. Dir: M. Dunn and C. Lamarre. Thesis: *A typology of classifiers and gender: From description to computation.*
- 2013 – 2015 ■ **M.A., National ChengChi University, Taiwan** in Linguistics. 📄. Dir: O-S. Her. Thesis: *A GIS typological analysis of the convergence and divergence among numeral classifier, genders and plural markers in the world's languages.*
- 2006 – 2011 ■ **B.A., National ChengChi University, Taiwan** in Diplomacy/Arabic Language and Literature (double major).




Skills

- | | |
|---------------------------------|---|
| Languages | ■ French (Native), Mandarin Chinese (Native), English (TOEIC 990/990), Swedish (CEFR B1), Arabic (CEFR B1). |
| Data visualization and analysis | ■ The R tidyverse and its extensions ↗, classification and regression (Generalized Linear Mixed Models, Random Forests, Neural Networks, Support Vector Machines), clustering methods. |
| Natural Language Processing | ■ Stylometry, topic modelling, word embeddings (GloVe, word2vec, fastText), text processing (Word segmentation, POS tagging, dependency parsing), web data harvesting (Docker, Selenium). |
| Computational methods | ■ Bayesian phylogenetic inference (BayesTraits, MrBayes, BEAST), Bayesian agent-based modelling with network analysis. |
| Linguistics | ■ CLAN (Computerized Language Analysis), ELAN (EUDICO Linguistic Annotator), Praat, Toolbox, VTL (VocalTractLab). |
| Computer | ■ Programming in R/Python, OS in Linux/ Mac/ Windows, \LaTeX . |





Certification (France)

- **Assistant Professor in Language Sciences** [Maître de conférences]: General linguistics and phonetics, Section 07 (Sciences du langage : linguistique et phonétique générales), Conseil national des universités (CNU), France. 
- **Assistant Professor in Language and Literature** [Maître de conférences], Section 15 (Langues et littératures arabes, chinoises, japonaises, hébraïques, d'autres domaines linguistiques), Conseil national des universités (CNU), France. 




Certification (Sweden)




- **Academic Teacher Training Course**, Unit for Academic Teaching and Learning, Uppsala University, Sweden. 
- **Supervising Students for Degree Project**, Unit for Academic Teaching and Learning, Uppsala University, Sweden. 
- **Assessment, Grading, and Feedback**, Unit for Academic Teaching and Learning, Uppsala University, Sweden. 

Courses



- **Current Research in Linguistics**: Undergraduate course (30h). Spring 2019 (Department of Linguistics and Philology, Uppsala University). Instructor in charge of the full course. The students learn qualitative and quantitative methods to develop and test linguistic hypotheses. 
- **Cognitive Linguistics**: Undergraduate course (30h). Spring 2019 (Department of Linguistics and Philology, Uppsala University). Instructor in charge of the full course. This course provides basic theoretical and methodological knowledge in the area of cognitive linguistics. 
- **Visualisation and Statistics**: Postgraduate course (30h). Autumn 2018 (Faculty of Languages, Uppsala University). Instructor for weekly two-hour lab sessions of R programming in data visualization and statistical analysis. 
- **French**: High school/University group courses (2500h). 2013 – 2015 (Chinese Institute of European Languages). Instructor for group courses (A1–C2 levels). The teaching involved conversation, writing, grammar courses, and preparation for the DELF diplomas. 

Teaching modules

- **Linguistic categories and language change**: Linguistic summer school (Linguistics, sociolinguistics, and ethnolinguistics) teaching module (3h). Spring 2022, 2023, 2024 (National Museum of Natural History, Paris). Instructor for two teaching sessions about conducting linguistic analyses on corpus data. 
- **Language and technology**: General education course (Language and World Civilization) teaching module (4h). Spring 2023 (National ChengChi University, Taipei). Instructor for two teaching sessions about computational linguistics.
- **An introduction to mixed models**: Postgraduate teaching module (3h). Autumn 2021 (Lab Dynamics of Language, CNRS/ University of Lyon). Instructor for a 3-week introduction to (generalized) linear mixed models. 
- **Quantitative typology from linguistic corpora**: Postgraduate teaching module (3h). Spring/Autumn 2021 (Department of Linguistics, University of Lyon). Instructor for 90-minute sessions to the use of corpora in linguistic research. 

- **A statistical analysis of the morphology-syntax distinction:** Postgraduate teaching module (3h). Spring 2020 (Lab Dynamics of Language, CNRS/ University of Lyon). Instructor for a 3-hour session introduction to generalized linear mixed models and tree-based computational classifiers. 
- **Fieldwork practice session with native speakers:** Linguistic summer school teaching module (15h). Autumn 2019 (International School in Linguistic Fieldwork, Paris). Instructor for 2-hour daily sessions on planning and conducting fieldwork. 
- **An Introduction to Random Forests in R:** Postgraduate teaching module (5h). Spring 2018 (Department of Linguistics and Philology, Uppsala University). Instructor for 90-minute sessions to the computational classifier of random forests. 

Awards and Grants

- 2024 ■ **CNRS bronze medal, 2024.** French National Center for Scientific Research, France. 
- 2023 ■ **Strategic Projects Grant, 2024–2025** (est. 60 000 euros). Quantifying the interaction between grammatical gender and social gender roles at a worldwide scale. Data Intelligence Institute of Paris, France. 
- **Constructive Advanced Thinking (CAT) Grant, 2023–2025** (est. 60 000 euros). Unraveling the interactions between culture and language: Does grammatical gender foster gender inequality and vice versa? Network of European Institutes for Advanced Study. 
- 2021 ■ **Labex ASLAN project, 2021–2024** (229 916 euros). MACDIT: Multi-agent models and social media data: Collective dynamics and individual trajectories in linguistic populations. French National Centre for Scientific Research, France. 
- 2020 ■ **ANR Junior Researcher Grant, 2021–2024** (166 936 euros). EVOGRAM: The role of linguistic and non-linguistic factors in the evolution of nominal classification systems. French National Research Agency, France. 
- **Labex ASLAN project, 2020–2021** (8 000 euros). RELI: Recherche En Linguistique Illustrée [Illustrating research with comics]. French National Centre for Scientific Research, France. 
- 2019 ■ **Development grant, fully-funded participation** (2500 euros) at the Oxford Digital Humanities Summer School. Riksbankens Jubileumsfond, Sweden.
- 2018 ■ **Research grant, funding** (1300 euros) to invite foreign researchers. Uppsala Forum Funding Development of Research, Sweden.
- 2018 – 2019 ■ **Invited collaboration, funding** (3000 euros) covering travel and accommodation for a two-week trip with one seminar and three work meetings. Graduate Institute of Linguistics, National ChengChi University, Taiwan.
- 2016 – 2019 ■ **Ph.D. full-time funded contract, funding** (monthly salary 2000 euros) of four years with one candidate chosen among (est.) 20. Uppsala University, Sweden.
- 2015 ■ **Best M.A. Thesis Award, Linguistic Society of Taiwan.**

- 2024 – 2025 ■ **Quantifying the interaction between grammatical gender and social gender roles at a worldwide scale** (Principal Investigator, with O Bonami and H Burnett), Strategic Projects Grant from the Data Intelligence Institute of Paris (est. 60 000 euros). The goal of this project is to gain new insight on the relation between grammatical and social gender roles through exploration of correlations between linguistic and social variables in thousands of languages and cultures. [!\[\]\(13b6bdd0ca077c333d50231f1443cb1d_img.jpg\)](#)

- 2023 – 2025 ■ **Unraveling the interactions between culture and language: Does grammatical gender foster gender inequality and vice versa** (Principal Investigator, with N Rochant, P-C Li, O Shcherbakova and C Cathcart), Constructive Advanced Thinking (CAT) Grant from the Network of European Institutes for Advanced Study (est. 60 000 euros). This project studies the interaction between grammatical gender and sociocultural gender by cooperating with various Institutes for Advanced Study in Europe. [!\[\]\(5dbedd4e1e8871e3a0e67053ad2f9701_img.jpg\)](#)

- 2021 – 2024 ■ **MACDIT: Multi-agent models and social media data: Collective dynamics and individual trajectories in linguistic populations** (Principal Investigator, with J-P Magué), Labex ASLAN Grant (229 916 euros). The goal of this project is to study the interaction between individual and collective levels of language variation and change in Twitter and Wikipedia data using Bayesian agent-based models. [!\[\]\(d4749465acb9b53e115af1f9ce82539c_img.jpg\)](#)
- **EVOGRAM: The role of linguistic and non-linguistic factors in the evolution of nominal classification systems** (Principal Investigator), ANR Junior Researcher Grant: ANR-20-CE27-0021 (166 936 euros). This project aims at building a database on nominal classification systems to identify the factors affecting their evolution. [!\[\]\(3e3001313d495ec87b5a6a5de6205728_img.jpg\)](#)

- 2018 – 2021 ■ **Variation, Change and Complexity in Linguistic and Health-related Behaviours (V2C)** (Postdoctoral researcher), IDEXLyon Fellowship Grant: 16-IDEX-0005. This project of Dan Dediu hosted at the lab Dynamique Du Langage (DDL) in Lyon investigates the variant patterns of language and speech between individuals and groups and how does it affect linguistic diversity. [!\[\]\(e26df985e6d3e053d2593dc7b93b41cf_img.jpg\)](#)




- 2020 – 2022 ■ **What underlies linguistic classification?** (Analyst), Project funded by the Swedish Research Council (VR-2019-02967). This project of Gerd Carling (Lund University) is hosted at the *Language Typology and Evolution* research group of Lund University. It aims at investigating gender and classifier assignment, stability, and change from a cross-linguistic, historical, and cultural evolutionary perspective. [!\[\]\(2d8989e35a5d1c61f2b9b0307dee0da4_img.jpg\)](#)





- 2016 – 2020 ■ **Les Parlers du Croissant** (Analyst and Fieldwork researcher), ANR funded research project: ANR-17-CE27-0001-01. This project of Nicolas Quint (LLACAN) aims at providing a multidisciplinary approach (linguistics, sociolinguistics, psycholinguistics, computational linguistics, among others) to endangered local varieties displaying mixed Occitan-Oilic features on the Northern Fringe of the Massif Central in France. In this project, I conducted fieldwork on the local varieties of Luchapt (Region of Vienne) [!\[\]\(10225a66c9f99322b84a7fc32767a3b8_img.jpg\)](#)

- 2014 – 2017 ■ **The Cognitive Mechanism and Neural Correlates Underlying the Convergence and Divergence of Chinese Numeral Classifiers and Measure Words** (Analyst), MOST funded research project: MOST103-2410-H004-136-MY. This project of One-Soon Her (National ChengChi University) developed hypotheses from formal syntax and investigated the cognitive processing of classifiers. The experiments involved the use of fMRI, priming tasks, and numerical stroop tasks. [!\[\]\(e628152647eaf2a5b8bfa082bbb75081_img.jpg\)](#)




- 2019 – 2024 ■ **International School in Linguistic Fieldwork – FieldLing**, with Nicolas Quint (LLACAN), Katharina Haude (SEDYL), and Lameen Souag (LACITO). 80 international participants. French National Centre for Scientific Research (LLACAN, LACITO, SEDYL, DDL), France. 
- 2018 ■ **Two-day Introductory Workshop on Geographic Information System**, with Alexandra Petrulovich (Uppsala University). 50 international participants. Uppsala University, Sweden. 
- 2017 ■ **Kick-off Seminar of the Research Environment ‘GIS for Language Study’**, with Alexandra Petrulovich (Uppsala University). 50 international participants. Uppsala University, Sweden. 
- 2016 ■ **PLM 2016 Workshop Session on Nominal Classification in South-East Asia**, with Marcin Kilarski (Adam Mickiewicz University). 20 international participants. Adam Mickiewicz University, Poland. 
- **29th Paris Meeting on East Asian Linguistics**, with Odile Roth (INALCO) and Adeline Tan (INALCO). 100 international participants. Research Centre for East Asian Linguistics, France. 

Supervision

























- 2023 – 2024 ■ Supervisory panel: Hung-Hsin Romain Hsu, PhD thesis. Thesis: *A corpus-based contrastive study of questions in French and Mandarin*, Université catholique de Louvain.
- 2023 ■ Evaluating committee: Saizhu Hu, PhD thesis defence. Thesis: *Lost and added in translation: A corpus-based study on Chinese numeral classifiers in translation between Chinese and English*, Adam Mickiewicz University.
- 2020 – 2023 ■ Collaborating expert: Neige Rochant, PhD thesis. Thesis: *A Grammar of Baga Pukur, an Atlantic language of Guinea*, Paris Sorbonne Nouvelle University.
- Collaborating expert: Mathilde Josserand, PhD thesis. Thesis: *Impact of inter-individual and inter-group variations on language evolution*, Lyon 2 University.
- 2023 ■ Evaluating committee: A-Han Lin, PhD thesis defence. Thesis: *The sortal concept in language: An investigation based on the distribution of numeral classifiers and plural markers*, National ChengChi University.
- Evaluating committee: Yung-Ping Liang, Master 2 thesis defence. Thesis: *A Theoretical Investigation of the Word Order of Early Multiplicative Numerals*, National Cheng-Chi University.
- Evaluating committee: Zi-Yun Cao, Master 2 thesis defence. Thesis: *On the Putative [Classifier Noun Numeral] Construction in Tai-Kadai Languages*, National ChengChi University.
- 2019 – 2022 ■ Co-supervision (with François Pellegrino): Natalja Ulrich, 3-year funded PhD. Thesis: *Linguistic and speaker variation in Russian fricatives*, University of Lyon. 
- 2021 – 2022 ■ Supervision: Jayden Macklin-Cordes, 2-year Postdoctoral researcher within the ANR grant EVOGRAM on nominal classification, Lab Dynamics of Language UMR 5596. 
- 2022 ■ Supervision: Fella Bennadji, Master 2 6-month internship about web-scrapping Indeed data and NLP analyses for detecting gender biases in online job offers, Lab Dynamics of Language UMR 5596. 
- Supervision: Fatma Bellili, Master 2 6-month internship about web-scrapping Twitter data and NLP analyses for detecting grammatical errors, Lab Dynamics of Language UMR 5596.

- 2021  Supervision: two middle school students during one-week internship at the National Museum of Natural History, Paris.
- 2021  Supervision: six high school students during one-week internship from the association Un Peu de Bon Science, University of Lyon. 
- 2019 – 2020  Supervision: Jean-Baptiste Chaudron, Master 1 internship in cognitive sciences, University of Lyon.

Academic reviews (non-exhaustive list)

- Journals  Linguistic Typology, Studies in Language, Language and Linguistics, Corpus Linguistics and Linguistic Theory, Linguistic Variation, International Journal of American Linguistics, Linguistics Vanguard, Language Documentation and Description, Scientific Data, Journal of Chinese Linguistics, Language and Cognition.
- Publishers  De Gruyter Mouton, John Benjamins, Language Science Press.
- Conferences  Empirical Methods in Natural Language Processing (EMNLP), Conference of the European Chapter of the Association for Computational Linguistics (EACL), Annual meeting of the Societas Linguistica Europaea (SLE).

Community Work

- 2024  **Interactive round table with Museum visitors**, ‘Le genre sous toutes ses formes’, Musée de l’Homme. 
-  **Interactive speech with Museum visitors**, ‘More than 7000 languages worldwide’, Musée de l’Homme. 
-  **Interview on the paper ‘Language trees with sampled ancestors support a hybrid model for the origin of Indo-European languages’**, radiofrance. 
- 2022  **Presentation on French dialects in the Croissant region**, Canal-U. 
-  **Podcast interview on natural language processing**, Canal-U. 
-  **Podcast interview on popularisation of science**, L’Onde – la webradio du Haut Limousin. 
- 2022 – 2027  **Advisory board for the Edinburgh Historical Linguistics series**, Edinburgh University Press. 
- 2020 – 2022  **Technical assistance for livestreams dedicated to the popularisation of linguistics**, The invitations of the Institute of Rare Languages (ILARA). Ecole Pratique des Hautes Etudes, Université Paris Sciences et Lettres, France. 
- 2021  **Chair and mentor for PhD candidates in typology**, Emerging Topics in Typology (ETT conference). University of Helsinki, Finland. 
- 2019 – 2020  **Instructor for High School Students**, Fête de la science: Promote research to high school students, DDL lab, University Lumière Lyon 2, France. 
- 2018 – 2019  **Instructor for High School Students**, Ph.D. project: Promote research to high school students, Faculty of Languages, Uppsala University, Sweden. 
- 2017 – 2018  **Organizer of Team-Building Sessions**, Department of Linguistics and Philology, Uppsala University.
- 2013 – 2015  **Instructor Volunteer**, Introducing Linguistics to students and public audience, Linguistic Society of Taiwan/National ChengChi University, Taiwan.