




# Euan McGill

✉ [euan.mcgill@gmail.com](mailto:euan.mcgill@gmail.com)  <https://www.linkedin.com/in/euan-mcgill>  
 <https://scholar.google.com/citations?user=IPtwnwYAAAAJ>  <https://euan-mcgill.github.io/>

## EXPERIENCE

---

### SignON (at Universitat Pompeu Fabra)

Mar. 2021-present

*Research Associate*

*Barcelona*

SignON is a 17-institution consortium and community-driven project which is building an app to translate between spoken and signed languages. My work includes leading the NLP pipeline module, app development, writing and contributing to EC deliverables, and machine translation research. I have also been involved in data augmentation for Spanish Sign Language from Spanish language text and video data, and parameter tuning for low-resource translation models.

### Apple Inc. (via Globetech Ltd., Appen)

Oct. 2017-Dec. 2018 & Dec. 2019-Oct. 2020

*Computational Linguist*

*Cupertino, Paris, Cork, Barcelona*

- ASR/TTS - (Inverse) text normalisation, maintenance, and innovation in language data for a voice assistant. Involved localisation of English varieties, primarily British. Additional responsibility of managing local team's repository of resources and questions relating to version control.
- TTS - Curated, localised, and supervised language data in the voice-building pipeline for British English speech synthesis. Implemented and modified tools to automate and test these processes, as well as testing work from other English locales. Regularly built new versions, and created tests to compare them. Assisted in delivery ahead of schedule.

### Samsung Electronics (Independent Contractor)

Dec. 2018-Nov. 2019

*Language Engineer*

*London*

- NLU - Creating, testing, and optimising training data for third-party apps' voice assistant requests. Involved analysing user data and required close work with software developers, business management, and other stakeholders.
- TTS - Established a pipeline for patching pronunciations in the front end and for text normalisation. Built tools to verify in-house resources, and reviewed the work of other engineers.
- ASR - Made improvements to the language model and inverse text normalisation, creating datasets for testing via voice recording.

## EDUCATION

---

### PhD, Information and Communication Technologies

Oct. 2021-present

*Universitat Pompeu Fabra*

- Within the Large Scale Text Understanding Laboratory of the NLP department.
- Developed skills in multilingual NLP, text mining, experimental methods, web crawling, text simplification, building and working with parallel corpora, and participating in the academic community.
- Teaching assistantship: Databases (Autumn 2021, 2022), Data Structures and Algorithms seminars (Spring 2022), Intelligent Web Applications (Autumn 2022)

### MSc Speech and Language Processing

Sep. 2016-Aug. 2017

*The University of Edinburgh*

- with Merit, Thesis: Investigating articulatory variation in the DoubleTalk corpus: Electromagnetic Articulography evidence of phonetic and speech-style variation
- Courses: Accelerated Natural Language Processing, Automatic Speech Recognition, Computer Programming, Human Factors, Instrumental Phonetics, Speech Processing, Speech Synthesis, Univariate Statistics with R

### MA(Hons) Linguistics

Sep. 2012-Jun. 2016

*The University of Edinburgh*

- 2:1, Thesis: "The thing about [hʌ]: Production, perception, and qualitative data from (th):[h] in Edinburgh speech". Courses including phonology, psycholinguistics, sociolinguistics and pre-Hons. French and Russian.

## RESEARCH & CERTIFICATIONS

---

### • Published work:

- Saggion, H., McGill, E., Chiruzzo, L. and Egea Gómez, S. (2023) *Linguistic Processing for Sign Language Translation*. Manuscript in preparation.
- McGill, E. (2023) *BSL-Hansard: A parallel, multimodal corpus of English and interpreted British Sign Language data from parliamentary proceedings*. Manuscript in preparation.
- Chiruzzo, L., McGill, E., Egea Gómez, S. and Saggion, H. (2022) Translating Spanish into Spanish Sign Language: Combining Rules and Data-driven Approaches. *Proceedings of the Fifth Workshop on Technologies for Machine Translation of Low-Resource Languages (LoResMT 2022)*. Gyeongju, Republic of Korea (Hybrid).
- Egea Gómez, S., Chiruzzo, L., McGill, E. and Saggion, H. (2022) Linguistically Enhanced Text to Sign Gloss Machine Translation. *Proceedings of the 27th International Conference on Natural Language & Information Systems*. València, Spain.
- McGill, E. (2021) Tailoring Automatic Text Simplification Output for Deaf and Hard of Hearing Adults. *Proceedings of the Doctoral Symposium on Natural Language Processing from the PLN.net network 2021*, Jaén, Spain.
- Egea Gómez, S., McGill, E. & Saggion, H. (2021) Syntax-aware Transformers for Neural Machine Translation. *Proceedings of the 14th Workshop on Building and Using Comparable Corpora*, Online.
- McGill, E. (2017, Jan.). The thing about [hʉ]: Production, perception, and qualitative data from (th):[h] in Edinburgh speech. Poster presented at **Scots@Ed**, Edinburgh, Scotland.
- Purse, R. & McGill, E. (2016). An intraspeaker variation study of Scottish /r/ pharyngealisation. *Lifespans and Styles*, 2(2), 36-44.

### • Training and Further Education:

- Mediterranean Machine Learning school (M2L) (Università degli Studi Milano-Bicocca) (2022, Sep.)
- Research Ethics and Personal Data Protection (Institutional Committee for Ethical Review of Projects, Universitat Pompeu Fabra) (2022, Nov.)
- Deep Learning for NLP (DL4NLP) (HiTZ Zentroa, Universidad del País Vasco) (2021, Jul.)
- Bayesian Statistics: From Concepts to Data Analysis (University of Santa Cruz) (2020, Dec.)

## SKILLS & ACHIEVEMENTS

---

### • Technical skills:

- *Programming languages*: Python, Bash, some exposure to C++, Perl
- *Version Control*: Git, P4 (Perforce)
- *Text Editing*: RegEx, L<sup>A</sup>T<sub>E</sub>X, Markup languages (MD, YAML, HTML), BNF Grammars
- *Statistics*: R, SQL, matplotlib, univariate analysis, linear regression, big data, data representation
- *Software*: Acoustic (Praat, Wavesurfer), Docker, NLP toolkits (e.g. nltk, spaCy, stanza), speech tech toolkits (HTK, Kaldi, Merlin, TensorFlow), experimental design, project-specific GUIs/VUIs

### • Research skills:

Phonetic-pragmatic level transcription, creating tools, QA and peer review, experimental design, confidential and sensitive information, debugging

### • Project management:

Teamwork, Agile, JIRA, Radar, establishing and implementing pipelines and workflows, working with external stakeholders, presentations, onboarding, training and mentoring

### • Natural languages:

English (*native*), French (*intermediate*), Catalan (*intermediate*), Spanish (*basic*), Russian (*beginner*)

### • Selected achievements:

Chair, 10th EITIC Doctoral Workshop @ Universitat Pompeu Fabra - 2022, Librarian @ Edinburgh University Wind Band committee - 2015-17, Reserve member - Edinburgh 'University Challenge' team - 2016, Student Volunteer at Lothian Equal Access Programme for Schools - 2014-17

Employment and academic references available upon request