



Automatic Recognition & Standardisation of French Verlan: Native Verification of Dataset Entries

INFORMATION SHEET FOR PARTICIPANTS

Thank you for your interest! Merci beaucoup de votre intérêt !

You are invited to take part in a research project that seeks to improve the automatic recognition and standardisation of verlan: a form of French back-slang widely used in contemporary youth culture. Your help will allow us to release an openly available, high-quality dataset for future linguistic and natural-language-processing research. Participation is **entirely voluntary**. Please read this sheet carefully before deciding whether to participate. If you agree, thank you; if not, we equally appreciate your consideration.

What is the Aim of the Project?

The project builds a Verlan Detection Lexicon and a gold-standard corpus of annotated sentences. By having native speakers verify candidate entries, we will: confirm the spelling and meaning of modern verlan forms, and link each entry to its standard French equivalent. The verified dataset will underpin machine-learning models and sociolinguistic analyses of verlan usage.

What Type of Participants are being sought?

- Adults (18+) native speakers of French.
- Comfortable with contemporary slang; no specialist training required.
- Recruited via the researcher's network, online francophone communities, and word-of-mouth. We seek about 20 participants. No prior involvement with computational linguistics is necessary.

What will Participants be asked to do?

Should you agree to take part in this project, you will be asked to

You will receive a Google Sheet containing ≈ 50 verlan tokens (or 30–40 sentences).
For each item, you will:

1. **Tick “correct”** if the verlan form and its proposed standard equivalent are accurate or select “needs correction” and provide the right form.
2. (Sentence sheets only) **Mark any additional verlan words** the system may have missed.

Most participants complete a sheet in about 30 minutes. You may volunteer for additional sheets if you wish, but this is not required.

Are there any risks or inconveniences?

The task poses **no known risks** beyond the time commitment. You may withdraw at any time before 30 September 2025 without disadvantage.

Will you be paid?

Participation is voluntary and **unpaid**. This project has Category B ethics approval, which does not permit payment or other material rewards.

What Data or Information will be Collected and What Use will be Made of it?

- **Raw data:** your annotations in Google Sheets and (optionally) an email address so we can send follow-up questions.
- **Storage:** raw data and de-identified metadata will be stored on password-protected University of Otago servers for at least five years, accessible only to the research team.
- **Publication:** the final dataset will be released under a CC-BY licence without any personal identifiers. Research findings may appear in conference papers, journals, or a public GitHub repository.

Can you change your mind?

Yes. You may withdraw yourself and your data up to 30 September 2025 by emailing the researcher. After that date, the dataset will be finalised and de-linked from personal identifiers.

What if you have questions?

Student Researcher

Eden Li
liyi5784@student.otago.ac.nz
School of Computing

Academic Supervisor

Dr Lech Szymanski
lech.szymanski@otago.ac.nz
School of Computing
Dr Veronica Liesaputra
veronica.liesaputra@otago.ac.nz
School of Computing

This study has been approved by the School of Computing as stated above. If you have any concerns about the ethical conduct of the research, you may contact the University of Otago Human Ethics Committee through the Human Ethics Committee Administrator (ph +64-3-479-8256 or humanethics@otago.ac.nz). Any issues you raise will be treated in confidence and investigated and you will be informed of the outcome.