Talking in French Like an Academia

Machine Learning Powered Verlan Identification

Author:
Yitian LI (4556502)

Supervisor(s):
Dr. Lech SZYMANSKI
Dr. Veronica LIESAPUTRA

September 27, 2025

Abstract

something.

1   Introduction

1.1   Context and Motivation

Since the early 19th century, the French people have started to talk using verlan. Just like Pig Latin[1] exists in English culture, verlan is an unusual and creative form of *argot* (slang) that is formed by flipping the syllables around in a word.[2][1, 2] Time flies, verlan has become more and more popular, and it is now widely used amongst teens and young people in francophone societies[3][3]. Examples of verlan can be as follows:

- bite = bi + te → te + bi → tebie (penis)

- shit = shi + t → t + shi → teuchi[3]

- bonjour = bon + jour → jour + bon → jourbon (greetings)

In real-life conversations, such can be used as in the example sentences below:

- *Le graff géant représente une tebie pixel art.*
  (The giant graffiti depicts a pixel art penis.)

- *Il a du bon teuchi du bled.*
  (He's got some good shit from the countryside.)

- *Un p'tit[4]jourbon et tout le monde sourit.*
  (A quick hello and everyone smiles.)

Indeed, verlan can be formed with different original languages, not only French, but also English and other languages. However, it always follows the same rule of flipping syllables, although, for better pronunciation reasons, certain minor amendments such as dropping unnecessary letters and applying accents (e.g., é, è) can be used from time to time[1]. Besides, due to the

---

[1]en.wikipedia.org/wiki/Pig_Latin
[2]In fact, the word *verlan* is a verlan from the word *l'inver* (the inversion).
[3]Such as France, Belgium, Switzerland, Luxembourg, and Canada.
[4]Standard spelling: petit.

universal trait of slang being used more often phonetically instead of written, verlan users tend to spell them differently when writing them down. As technology develops, this has been occurring more frequently than ever in daily texting[4].

Thinking internationally, when people are communicating with translators, it is possible that slang in their mother language can be brought to the conversation, which could be tricky for translators to translate[5]. Using translators such as DeepL[5] and Google Translate[6] to translate sentences that contain verlan from French to English can be a specific example to prove this. Furthermore, although both of the translators above are using Machine Learning (ML) for translation, their results of translating verlans are not ideal[6, 7]. For example, when attempting to translate the sentence above, *Le graff géant représente une tebie pixel art.*, both Google Translate1 and DeepL2 cannot translate the word *tebie* correctly. Specifically, for DeepL, there is no desired translation as *penis* in its alternative word list for *tebie*3.
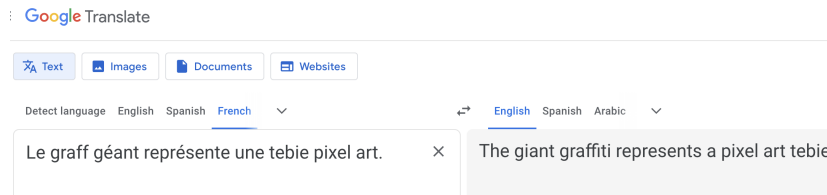
Figure 1: Google Translate cannot translate the verlan *tebie* correctly.
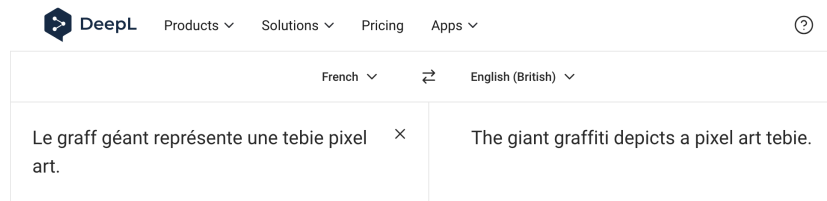
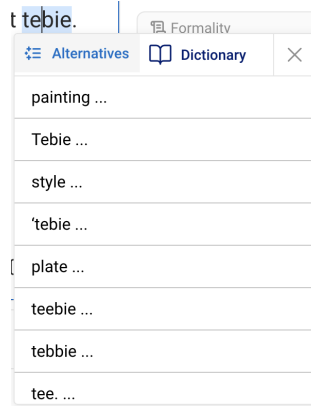Figure 2: DeepL cannot translate the verlan *tebie* correctly.

Figure 3: No desired translation for verlan *tebie* in DeepL's alternative word list.

Thus, a question shall naturally arise: Can we improve translators' performance in translating slang by improving the ML model? The answer is undoubtedly 'yes' in an era where artificial intelligence research is expanding rapidly. Researchers have been making progress in identifying slang using ML[9] and, moreover, in translating noisy text, of which slang is a part[8].

But what about verlan? There is no known ongoing or completed research on identifying such slang or their translations[7], nor does a proper dataset exist. The only work similar to this is an assignment published at the University of Toronto[8], asking students to train a Neural Machine Translation (NMT) model to transform standard English into Pig Latin. It is not only the other way around; instead of identifying Pig Latin and transforming it back to standard English, it is also more of an example for students to practice using NMT than a discussion on its identification and translation. Shouldn't we do something?

This report aims to change that.

1.2   Objective

The purpose of the project is to create two verlan datasets: one functioning as a dictionary, containing the verlan words and their normalised standard French equivalents; the other a dataset of sentences that contain verlan,

---

[7]Until September 2025.
[8]https://uoft-csc413.github.io/2022/assets/assignments/PA03.pdf

paired with the same sentences containing normalised words, with labels indicating whether a sentence contains verlan. After that, the project embeds and classifies verlan using Large Language Models (LLMs) and analyses the results.

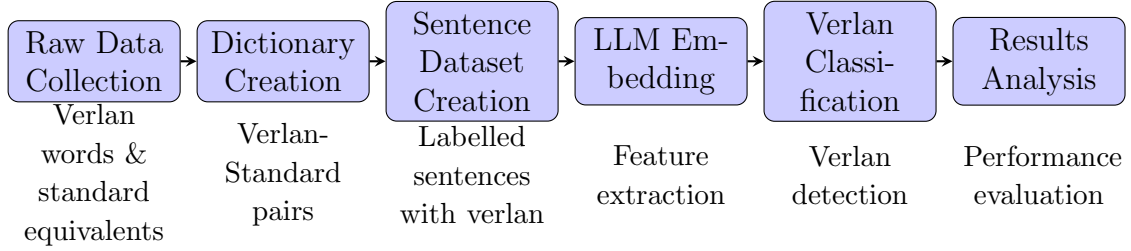| Raw Data Collection	Dictionary Creation	Sentence Dataset Creation	LLM Embedding	Verlan Classification	Results Analysis
Verlan words & standard equivalents	Verlan-Standard pairs	Labelled sentences with verlan	Feature extraction	Verlan detection	Performance evaluation

Figure 4: A visulisation of the objectives.

With the purpose above, the report contributes to the linguistics and the AI researchers two verlan datasets, for dictionary making or LLMs training. The report also evaluates how good we can achieve the identification of verlan with ML, to benefit machine translation in the future.

The code and the unannotated, un peer-reviewed dataset developed as part of the project are released under openlicences and aligns with open science best practices, with the usage of a version controlled software development platform (GitHub)[9]. The annotated, peer-reviewed dataset will be published shortly after this report, aiming by the end of 2025.

2   Background

Typically, introduction will be followed with a section providing the background for the project. This often means a bit of literature review or explanation of the fundamental concepts/terminology. This section has a dual purpose:

- to provide enough information for the readers, who might not be experts in your field, so that they can understand everything that follows - methods, results, etc.;

---

[9]github.com/greateden/Verlan-Identification-Normalisation

- to demonstrate that you have gained a sufficient level of competency in the topic.

Don't forget to cite the sources of your information. For instance, a reference for LaTeX can be found here [**?**]. If you were, say, interested in Computer Vision you might want to take a look at this article [**?**]. If you want to cite two related document at once, you can do it like so [**?**, **?**].

3 How to compile LaTeX

The remaining sections should present what you have done in your project including the results and analysis. For this document, this will be a demonstration of how to use LaTeX in order to create your report.

LaTeX documents are prepared using markup language and need to be compiled to produce pdfs.

The easiest way is to use the Overleaf service - it's free for private projects. Create an account and once you login you can create different projects (essentially different documents). If you joined the "COSC4x0 Report Template project" by following the link from Blackboard, you can *Copy* and rename it to create a new project from your Overleaf *Home* page. Alternately, you can click the "New Project" button, select the "Upload project" option and upload the *cosc4x0report.zip* file where this pdf came from. The template will open on the website and you'll be able to edit the report in your browser, compile LaTeX to pdf online and have it saved on a remote server. Later you just download the final pdf and submit as your report.

Another option is to install a LaTeX compiler on your machine and compile the .tex file yourself. On macOS you need to download and install MacTeX. Then, using programs like TexShop (free), or VSCode (with appropriate extensions) (also free), or Texifier (awesome, but not free) you can edit and compile a .tex file into a .pdf.

4 LaTeX markup examples

4.1 Sections

Use `\section{}` and `subsection{}` commands to organise your document. LaTeX handles all the formatting and numbering automatically. Use `\label{}`

Figure 5: This is a figure caption.

Item	Quantity
Widgets	42
Gadgets	13

Table 1: An example table.

and \ref{} commands for cross-references.

4.2 Comments

Comments might be useful during the writing process, as reminders or questions to your supervisor (who should get a chance to comment on your report). Comments can be added to the margins of the document using the \todo{} command, as shown in the example on the right. You can also add inline comments:

This is an inline comment.

Here's a comment in the margin!

4.3 Tables and Figures

Use the table{} and \tabular{} commands for basic tables — see Table 1, for example. You can include a figure (JPEG, PNG or PDF) with the \includegraphics{} command as in the code for Figure 5 below.

4.4 Mathematics

LaTeX is great at typesetting mathematics. Let $X_1, X_2, \ldots, X_n$ be a sequence of independent and identically distributed random variables with $E[X_i] = \mu$ and $\text{Var}[X_i] = \sigma^2 < \infty$, and let

$$S_n = \frac{X_1 + X_2 + \cdots + X_n}{n} = \frac{1}{n}\sum_i^n X_i \tag{1}$$

denote their mean. Then as n approaches infinity, the random variables $\sqrt{n}(S_n - \mu)$ converge in distribution to a normal $\mathcal{N}(0, \sigma^2)$. You can also reference labeled equations, such as Equation 1.

4.5 Lists

You can make lists with automatic numbering ...

1. Like this,

2. and like this.

...or bullet points ...

- Like this,

- and like this.

5 Conclusion

Concluding remarks. Send the pdf (not the `*.tex` file) to your supervisor for comments (as early as possible). Don't forget to change the `\usepackage[draft]{cosc4x0style}` setting to `\usepackage{cosc4x0style}` to produce the pdf in the format for the final submission.

References

[1] Radjabov, Ruslan Rajabmurodovich. *Understanding "Verlan" in the French Language.* Web of Scientist: International Scientific Research Journal, vol. 6, no. 3, 2025, pp. 368–372. Available at: https://webofjournals.com/index.php/3/article/view/3264.

[2] Bach, Xavier. *Tracing the origins of verlan in an early nineteenth century text.* Journal of French Language Studies, vol. 28, no. 1, 2018, pp. 1–18. Cambridge University Press. doi:10.1017/S0959269516000221.

[3] Olivier Sécardin. *Évolution du verlan, marqueur social et identitaire, comme reflet de la langue et de la société françaises.* Synergies Europe, no. 3, 2008, pp. 223–232. Available at: https://journal.lib.uoguelph.ca/index.php/synergies/article/download/1037/1859?inline=1.

[4] Rúa, Paula López. "Shortening Devices in Text Messaging." *Journal of Computer-Mediated Communication*, vol. 10, no. 4, July 2005. Wiley. doi:10.1111/j.1083-6101.2005.tb00268.x.

[5] Hajiyeva, Bulbul. "Translating Idioms and Slang: Problems, Strategies, and Cultural Implications." Acta Globalis Humanitatis et Linguarum, vol. 2, no. 2, 2025, pp. 284–293. doi:10.69760/aghel.025002123.

[6] DeepL. "DeepL Translator translates texts using artificial neural networks. These networks are trained on many millions of translated texts." *DeepL Blog*, 2020. Available at: https://www.deepl.com/en/blog/how-does-deepl-work.

[7] Wu, Yonghui, et al. "Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation." arXiv preprint arXiv:1609.08144, 2016. Available at: https://arxiv.org/abs/1609.08144.

[8] Michel, Paul, and Graham Neubig. "MTNT: A Testbed for Machine Translation of Noisy Text." *Proceedings of EMNLP*, 2018. Available at: https://aclanthology.org/D18-1050/.

[9] Pei, Wenzheng, Lin Sun, and Baoxun Xu. "Slang Detection and Identification." *Proceedings of the 23rd Conference on Computational Natural*

Language Learning (CoNLL), 2019, pp. 881-889. Association for Computational Linguistics. doi:10.18653/v1/K19-1082. Available at: https://aclanthology.org/K19-1082

Appendix A   Some extra things

If you have anything more to add such as:

- not essential details - things that might be too much for first time reading, or could be distracting from the main points...but are still important for reproducibility or deeper understanding

- work that was done in the project but doesn't go with the main work, or detracts/is not essential for the main narrative.

Appendix B   Aims and Objectives

Interim report only! – you do not need to include this appendix in the final report. However, in your interim the last appendix should include your original Aims and Objectives, and, if the things have changed, the revised Aims and Objectives. If you used the LaTeX template provided for your Aims and objectives document, just copy the `\paragraph{Aims}` and `\paragraph{Objectives}` sections and paste them here.

Original

Aims   Here you are describing the term goal of the project. What do you want to achieve by the end? What is the ultimate goal of this work? For example, the primary aim of this document is to have students produce suitable aims and objectives for their COSC480/490 project. While the aims and objectives document is not an assessed deliverable, a clear definition of what is to be done, and a bit of planning of how it is to be accomplished is paramount to the project's success. It is important to establish the scope of the project.

Objectives   Objectives list the milestones that you need to achieve in order to achieve the projects aim(s). It's a rough plan for what needs to happen

in what order. It's best to list the objectives in bullet point form. For many projects the structure to these objectives might follow the following pattern (objective names are just examples – you can have different objective names):

- background reading; going through the literature; learning about the research field;
- setting up of some kind of system for the project; getting the environment for experiments working;
- conducting preliminary experiments; implementation of a basic/simple approach; producing base case results;
- trying method 1; recording the results;
- trying method 2; recording the results.

Revised

Aims   Here you are describing the term goal of the project. What do you want to achieve by the end? What is the ultimate goal of this work? For example, the primary aim of this document is to have students produce suitable aims and objectives for their COSC480/490 project. While the aims and objectives document is not an assessed deliverable, a clear definition of what is to be done, and a bit of planning of how it is to be accomplished is paramount to the project's success. It is important to establish the scope of the project.

Objectives   Objectives list the milestones that you need to achieve in order to achieve the projects aim(s). It's a rough plan for what needs to happen in what order. It's best to list the objectives in bullet point form. For many projects the structure to these objectives might follow the following pattern (objective names are just examples – you can have different objective names):

- background reading; going through the literature; learning about the research field;
- setting up of some kind of system for the project; getting the environment for experiments working;
- conducting preliminary experiments; implementation of a basic/simple approach; producing base case results;
- trying method 1; recording the results;
- trying method 2; recording the results.