

Constrained Language Models for Interactive Poem Generation

Andrei Popescu-Belis^{1,2}, Alex R. Atrio^{1,2}, Valentin Minder¹, Aris Xanthos³,
Gabriel Luthier¹, Simon Mattei¹, Antonio Rodriguez³

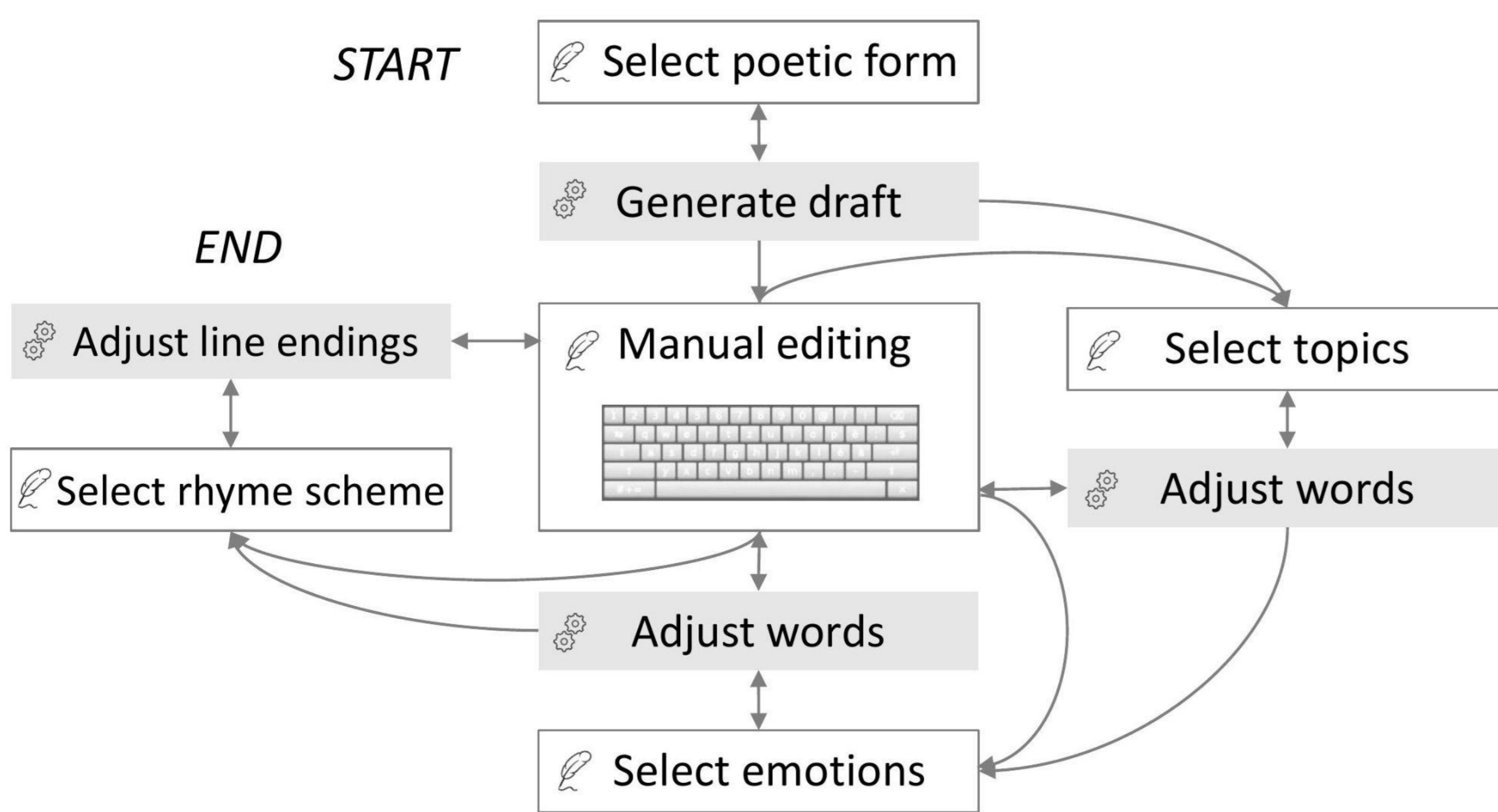
¹ Haute École d'Ingénierie et Gestion du Canton de Vaud (HEIG-VD), Switzerland

² École Polytechnique Fédérale de Lausanne (EPFL), Switzerland

³ Université de Lausanne, Switzerland

Summary

- **CR-PO (Création Poétique Assistée par Ordinateur):** A system for interactive French poem generation, which combines neural language models (LMs) with explicit constraints that can be set by users on form, topic, emotion, and rhyming scheme.
- **Challenges:** LM on low-resource (genre and form); text tuning once the poem is already generated; autonomous system in public exhibition.
- **In stages:** **Form** - **Topic** - **Emotion** - **Rhyme** (+ manual editing in each).



Model, Data, and Implementation

- **Form** and **Rhyme:** Input (40)+Emb (100)+2 BiLSTM (256)+Attention (612)+Softmax (89) with character-based TextGenRNN.
- **Topic** and **Emotion:** CamemBERT fine-tuned.
- **15 MB (10k poems) general dataset.** For topic and emotion-specific datasets, each <1 MB manually labeled → Naive Bayes apply to general dataset.
- Implemented on small-form-factor PC with 32" touchscreen; stages logged in JSON for further analysis; final poem printed or uploaded.

Evaluation

- Digital Lyric exhibition
 - 100 poems in 13 days
- Workshop open to the public at HEIG-VD (Nov '21).
 - 42 poems from 25 visitors
- Average number of interactions per stage (>1 means some users started over).

	AVG	STD
1. Generation of 1 st draft	1.62	1.10
Manual editing	0.31	0.47
2. Topic adjustment	1.05	1.23
Manual editing	0.36	0.48
3. Emotion adjustment	0.95	0.91
Manual editing	0.26	0.45
4. Rhyming scheme	1.26	1.67
Manual editing	0.38	0.49

Table 4: Average number of interactions with the CR-PO system, for each stage, at the 2021 Open Doors of HEIG-VD (25 visitors, 42 poems).

Constrained Autoregressive Generation of Poems

I. Setting the Poetic Form

- LM generates a poem based on 4 fixed possible forms chosen by the user (although any number of stanzas, lines, and lengths are possible).
- Sample from distribution + forbidden characters (punctuation, etc.)
- Temperature augments the highest probability values in the distribution.
- **Cold start:** 4 first lines generated not shown to the user.
 - Length of Verses
 - Generate 85% of verse disallowing punctuation.
 - Loop through 15% with objective of ending with punctuation mark. If not, relax constraints.
- Post-processing: whitespaces, uppercase, and dictionary checking for closest match (not common).

2 and 3. Adjusting to Topics and Emotions

- 5 topics and 3 emotions. User selects proportion in each stage with sliders.
 - Word Selection
 - Independence Quotients: a corpus-based measure of the correlation between words and topics (resp. emotions).
 - Select about 8% of the words (exclusive of stopwords) for replacement.
 - Word Replacement
 - Word-level CamemBERT fine-tuned to the topic/emotion.

4. Setting the Rhyming Scheme

- User chooses out of a fixed number of schemas (i.e. AABB).
- Select ending words to be changed, and get list of candidates (same POS) from rhyming dictionary (regex from phonetic dictionary - 150k words).
- General LM scores them, and selects highest.

le rêve des couleurs de la forêt en silence.
La terre s'accroche aux premiers pas de l'homme.
Le soir, dans tout ce qu'il trouve dans ces tomes
Sous les champs d'or de la fleur de la naissance !

Être distrait pour la chair sourde de son âme
L'espoir d'un chant sous le cinname ;
La fille du ciel s'abat, le soir, le temps est éclatant,
Et le printemps, et le soleil se clapotant
Et ses doigts d'or s'en vont au fond du soir.

Ma tête dans les nuages, les arbres noirs, en haut,
Et les parfums pleurants s'en vont en lui rendant les cieux.
Les fleurs de la saison descend de la terre,
A l'entour des vagues du vent, les ombres de l'enfance.

Example

- A poem made up of 3 concatenated sample outputs of our system (out of 6), each prompted internally with the underlined strings.
- CR-PO's first participation to a poetry contest (at EPFL), on the same topic as the prompt strings.

Conclusions

- French poem generator as combination of neural LM + rule-based constraints on **form**, **topic**, **emotion**, and **rhyme**.
- Autonomous system left without supervision in public exposition.
- Lack a high-level meaning (such as narratives) - still very difficult to solve w/o very large LMs.
- Plans for the future:
 - To be extended to English.
 - Improving individual stages such as rhyme.
 - Users able to introduce seed words.

Acknowledgements

The design and implementation of the system were directly supported by the SNSF Agora project "Digital Lyric" (n.~184330), for which Antonio Rodriguez received the Agora Optimus prize. We also acknowledge the support of HES-SO through the PhD support fund (AGP n.~99864), of SNSF through the DOMAT project (n.~175693), and of the Institute for ICT at HEIG-VD.

Contact us at:

andrei.popescu-belis@heig-vd.ch
alejandroramirezatrio@heig-vd.ch

aris.xanthos@unil.ch
antonio.rodriguez@unil.ch