SPELLL 2023

December 7th, 2023
Kongu Engineering College
Tamil Nadu, India

Recent years on automatic poetry generation: my own thoughts

Manex Agirrezabal
Associate Professor
University of Copenhagen

Outline

- Me and myself
- Examples of generative machines
- Brief history of Automatic Poetry Generation
- The case of Basque, a European minority language
- Challenges with current paradigms
- Erato: Automatizing Poetry Evaluation

Outline

- **Me and myself**
- Examples of generative machines
- Brief history of Automatic Poetry Generation
- The case of Basque, a European minority language
- Challenges with current paradigms
- Erato: Automatizing Poetry Evaluation

About myself

- Computer Engineering at the University of the Basque Country (UPV/EHU) (2006-2011)
- M. Sc. and PhD in Analysis and Processing of Language at the same place (2011-2012, 2013-2017)
- PostDoc at the University of Copenhagen (2017-2019)
- Assistant Professor (2019-2022) at “
- Associate Professor (2023-) at “

About myself

- Natural Language Processing
- Computational creativity
- Multimodal communication
- Computational morphology and phonology
- Finite-State methods
- Poetry

About myself

I teach in a Master Program called IT & Cognition (M. of Sc.):

- Language Processing 1 (together with a colleague)
- Language Processing 2 (together with a colleague)

I am planning to offer a course soon:

- Computational Creativity (Starting fall 2024)

I taught in previous years:

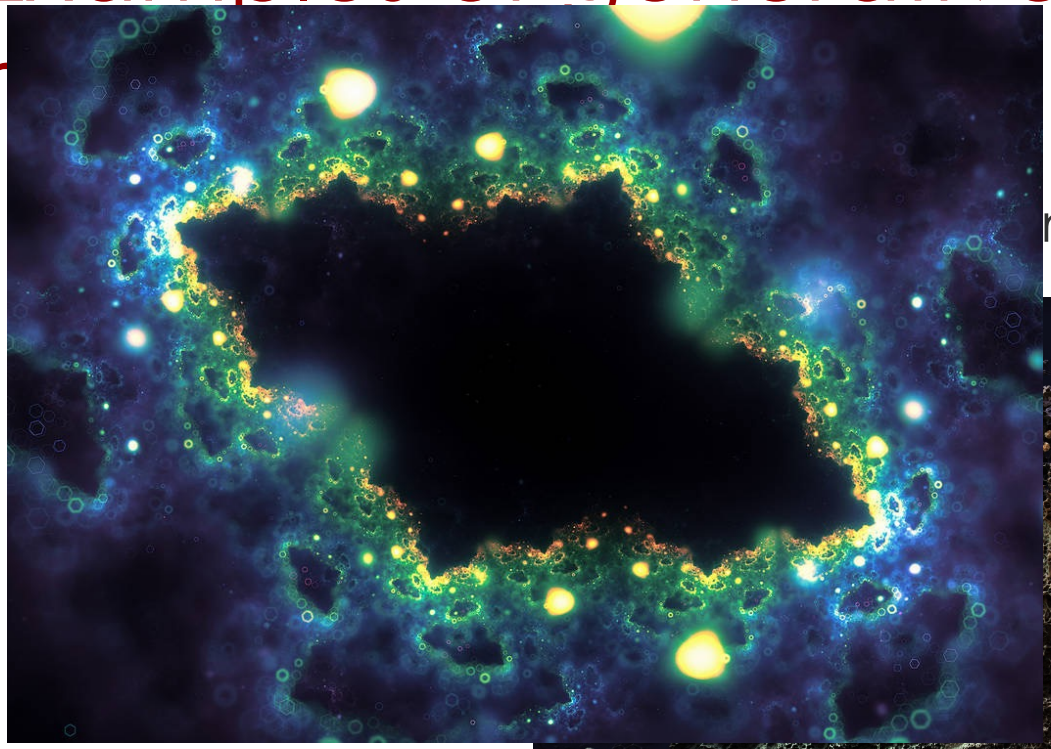
- Scientific Programming

Outline

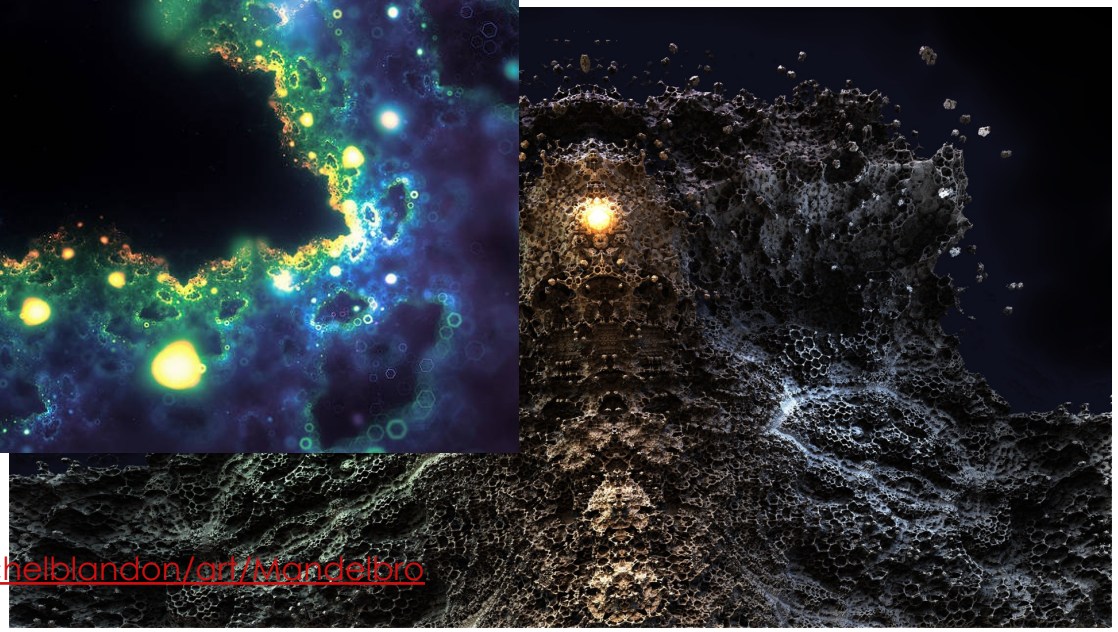
- Me and myself
- **Examples of generative machines**
- Brief history of Automatic Poetry Generation
- The case of Basque, a European minority language
- Challenges with current paradigms
- Erato: Automatizing Poetry Evaluation

Examples of generative r

- ntent



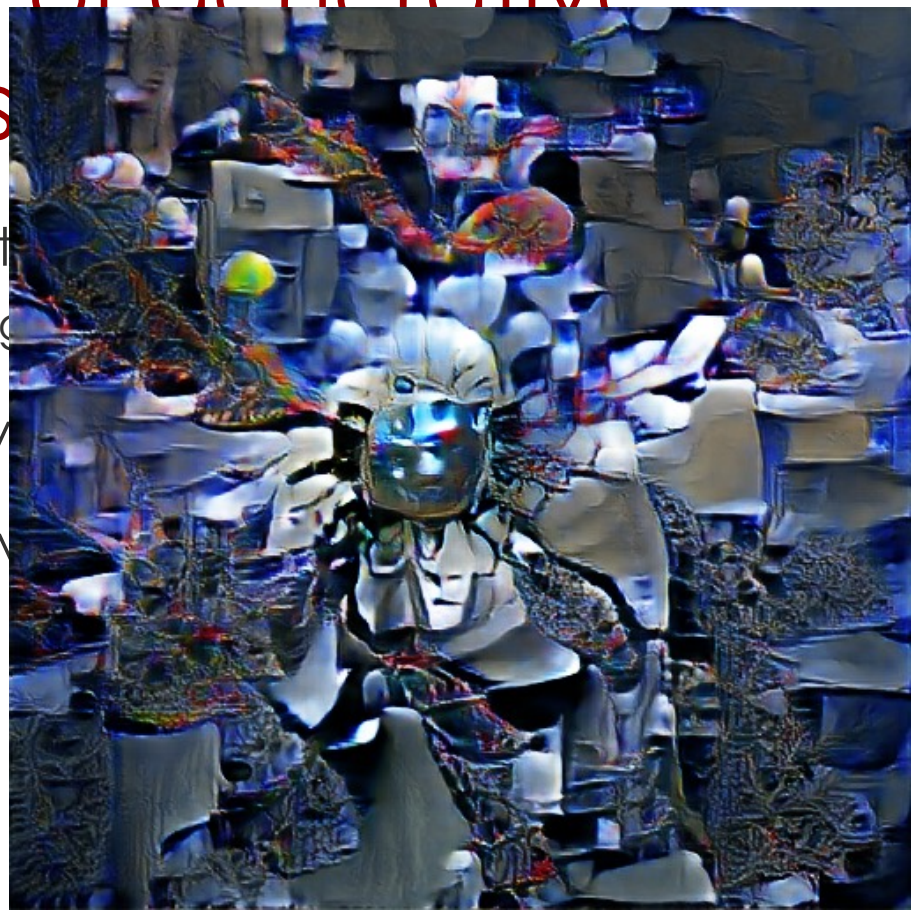
Mandelbrot Nebula
By rachelblandon
https://www.deviantart.com/rachelblandon/art/Mandelbrot-Nebula-845137047



AmazingFoldingTetrapack08 - Mandelbulb 3D fractal
By schizo604
https://www.deviantart.com/schizo604/art/AmazingFoldingTetrapack08-Mandelbulb-3D-fractal-449419101

Examples of generative machines

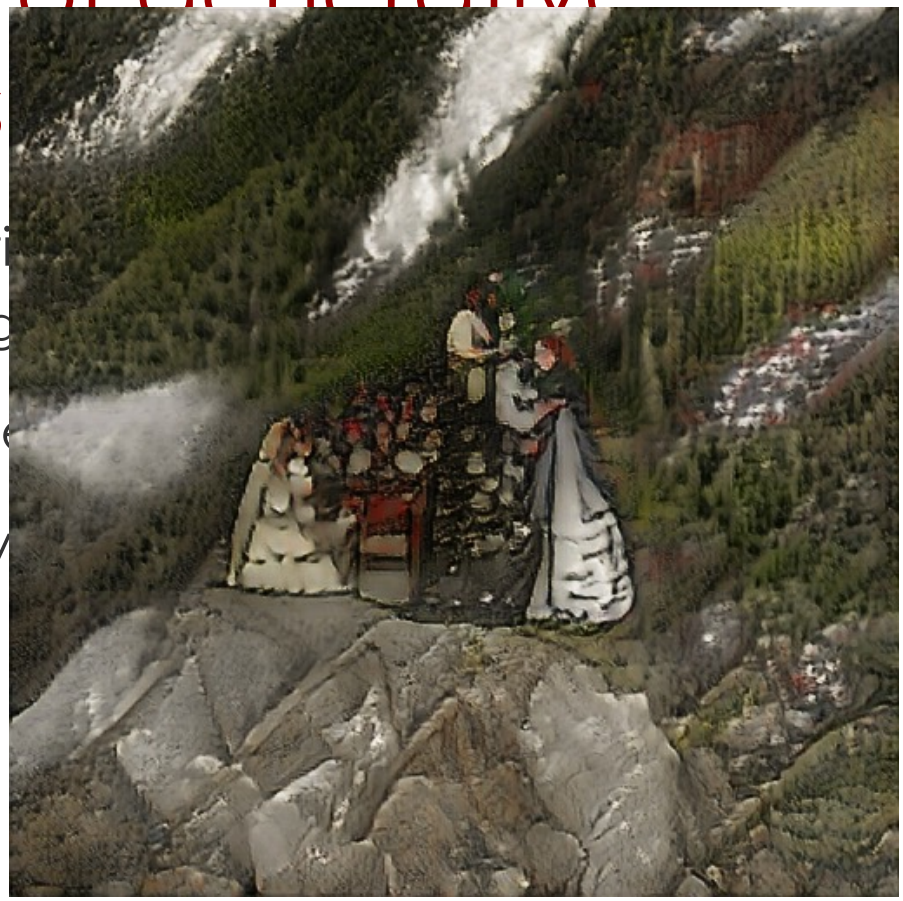
- Image Generat
  - CLIP + GAN: bi
- CLIP: Contrastiv
- GAN: Generati



Prompt: *artificial intelligence*
Source: https://github.com/lucidrains/big-sleep

Examples of generative machines

- Image Generati
  - CLIP + GAN: big
- CLIP: Contrastive
- GAN: Generativ



Prompt: *marriage in the mountains*
Source: https://github.com/lucidrains/big-sleep

Examples of generative machine

- Image Gene
  - CLIP + GAN:
  - Dall-E

An astronaut riding a horse in photorealistic style.
OpenAI
https://openai.com/dall-e-2

Examples of generative machines

- Image Generation
- Music Generation
  - APOPCALEAPS (Sneyers et al., 2010)
    - Generates Pop music based on high-level probabilistic rules
    - These rules are learned automatically from examples

https://sneyers.info/jon_old/apopcaleaps/index.html

Examples of generative machines

- Image Generation
- Music Generation
  - APOPCALEAPS (Sneyers et al., 2010)
  - DeepBach (Hadjeres et al., 2017)
    - Graphical model that models polyphonic music and specially hymn like music
    - Trained on chorale harmonizations by J.S. Bach

https://sites.google.com/site/deepbachexamples/

Examples of generative machines

- Image Generation
- Music Generation
  - APOPCALEAPS (Sneyers et al., 2010)
  - DeepBach (Hadjeres et al., 2017)
  - JazzML (Chow, 2014) and DeepJazz (Kim, 2016)
    - Jazz chord and music generation based on ML or DL

Examples of generative machines

- Image Generation
- Music Generation
- Jokes / Humor Generation
  - The Joking Computer (Ritchie and Masthoff, 2009)
    - Lexical database
    - Simple linguistic rules (sound)

•What do you get when you cross a frog with a road?
•A main toad

•What kind of temperature is a son?
•A boy-ling point

•What do you call a shout with a window?
•A computer scream

•What do you call a washing machine with a september?
•An autumn-atic washer

Brought by The Joking Computer
https://phys.org/news/2009-12-glasgow.html

Examples of generative machines

- Image Generation
- Music Generation
- Jokes / Humor Generation
  - The Joking Computer (Ritchie and Masthoff, 2009)
  - Hahacronym (Stock and Strapparava, 2005)
    - A computational humor system
    - Generated funny acronyms
      - Change words with very unrelated words

- Association for Computing Machinery
  - Association for Confusing Machinery
- Federal Bureau of Investigation
  - Federal Bureau of Intimidation
- Personal Digital Assistant
  - Penitential Demoniacal Assistant

Examples of generative machines

- Image Generation
- Music Generation
- Jokes / Humor Generation

Outline

- Me and myself
- Examples of generative machines
- **Brief history of Automatic Poetry Generation**
- The case of Basque, a European minority language
- Challenges with current paradigms
- Erato: Automatizing Poetry Evaluation

Brief history of Automatic Poetry Generation

- Classic rules and randomness
- Case-Based Reasoning
- Evolutionary models
- Markovian Language Models (with constraints)
- RNN-LSTM Language Models
- Transformers (LLM + Fine-tuning)
- Prompting

Classic rules and randomness

- Early attempts to automatic poetry generation
- Theo Lutz (1959)
  - Stochastic texts
- GUNICHT JEDER BLICK IST NAH .KEIN DORF IST SPAET
  EIN SCHLOS IST FREI UND JEDER BAUER IST FERN
  JEDER FREMDE IST FERN .EIN TAG IST SPAET
  JEDES HAUS IST DUNKEL .EIN AUGE IST TIEF
  NICHT JEDES SCHLOS IST ALT .JEDER TAG IST ALT –
  - 3002 poems (63 meters sheet)
  - IBM 7070 mainframe computer
  - https://narratingcomputing.com/n
    tellation-nanni-balestrini/

Case-Based Reasoning

- Gervás (2001)
  - Automatic Spanish Poetry Expert and Rewriting Application (ASPERA)
  - Semiautomatic interactive poetry generation
  - Four steps for poetry generation
    - Retrieve (get from database)
    - Reuse (modify an existing case)
    - Show to user and revise/modify
    - Retain (save in the database)

Evolutionary models

- Poevolve (Levy, 2001)
  - Originally written for limerick generation
  - Neural Network
  - Fitness function and evolutionary computation
  - Objective: Poems get better
  - https://github.com/rplevy/poevolve
- Mcgonagall (Manurung, 2003)
  - Evolutionary algorithm for poetry generation
  - Poetry generation as state-space search problem
  - Goal:
    - Meaningfulness, Grammaticality and Poeticness

Markovian Language Models with constraints

- Barbieri et al. (2012)
- Constrained Markov Processes
- This allows them to control
  - Rhyme
  - Meter
  - Meaning
- Applied for lyric generation

RNN (LSTM) language Models

- Zhang and Lapata (2014)
  - Recurrent Networks for Chine
  - Generate line from keyword
  - Generate next lines
    - Assess next line generatior
  - Human evaluation
    - Fluency, coherence, meai rank
  - Humans were the best

相 思

Missing You

红 豆 生 南 国， (*Z P P Z)

Red berries born in the warm southland.

春 来 发 几 枝？ (P P Z Z P)

How many branches flush in the spring?

愿 君 多 采 撷， (* P P Z Z)

Take home an armful, for my sake,

此 物 最 相 思。 (* Z Z P P)

As a symbol of our love.

RNN (LSTM) language Models

- Hopkins and Kiela (2017)
  - Two different models:
    - Neural language model trained on a phonetic encoding
    - Constraint satisfaction problem with a generative model and a WFST discriminator
  - Phoneme-level LSTM encoder and orthographic decoder
  - Intrinsic and extrinsic evaluation
    - Extrinsic: Turing-test (distinguishability test)

RNN (LSTM) language Models

- Lau et al. (2018)

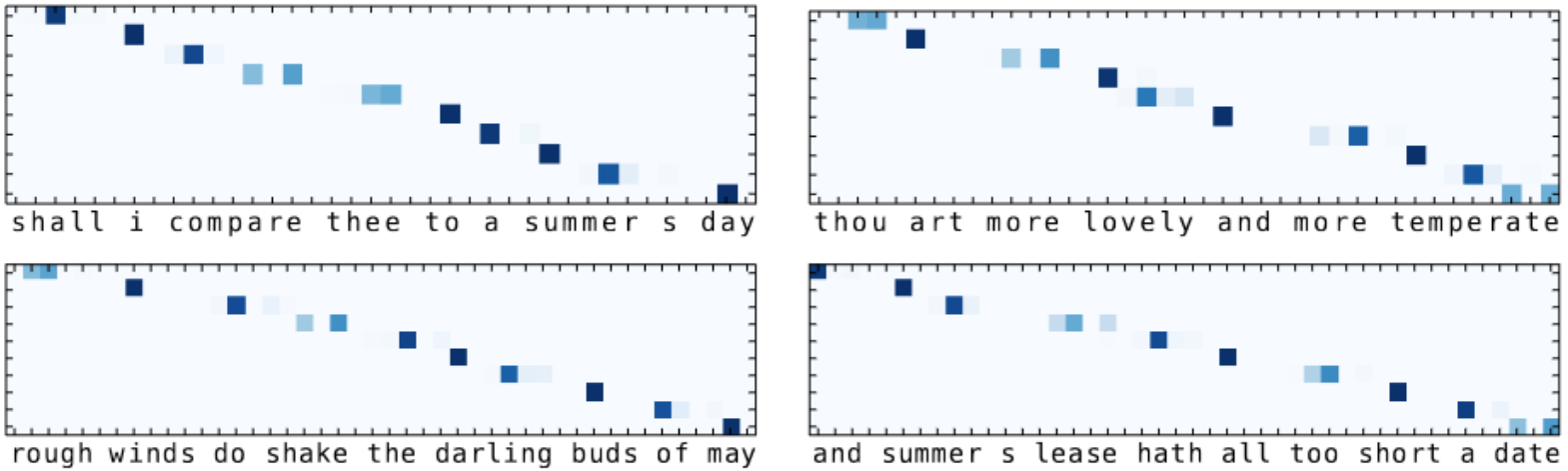


Figure 3: Character attention weights for the first quatrain of Shakespeare's *Sonnet 18*.

Transformers (LLM + Finetune)

- Bena and Kalita (2019)
  - They generate “dream poetry”
  - Finetune GPT2 on two datasets:
    - Poem dataset
    - Dreams dataset
  - Emotion model, 8 basic emotions
    - Joy, trust, fear, surprise, sadness, anticipation, anger, and disgust
  - Evaluation
    - Human judgements (likert scale)
    - Readability indices

Transformers (LLM + Finetune)

- Beheitt et al. (2022)
  - Arabic poetry generation by finetuning GPT-2
- Popescu-Bells et al. (2023)
  - Finetuned GPT-2 for English poetry generation
  - To be integrated in an interactive poetry generation system (Popescu-Bells et al., 2022)

Prompting poetry generation

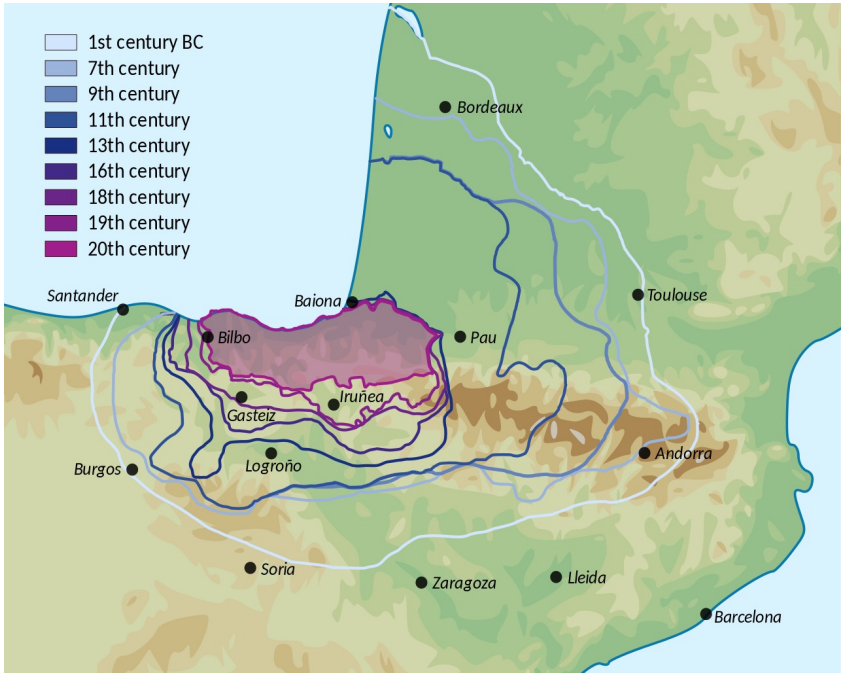
- Recent work on poetry generation:
  - Ask ChatGPT: Please write me a poem about second hand cars
- Is prompting really making art?
  - McCormack et al., 2023
  - At this point it is difficult to see prompt writing as a significant art practice.
- Collaborative writing
  - CoPoet (Chakrabarty et al., 2022)
  - Very thorough analysis of results

Outline

- Me and myself
- Examples of generative machines
- Brief history of Automatic Poetry Generation
- **The case of Basque, a European minority language**
- Challenges with current paradigms
- Erato: Automatizing Poetry Evaluation

The case of Basque, a European minority language

- Basque language
  - Last surviving paleo-European language
  - Language isolate
  - Spoken in the Basque region (between Spain and France)
  - ~751,700 speakers
  - Ergative-absolutive language
  - Agglutinative morphology



The case of Basque, a European minority language

- PoeLM (Ormazabal et al. 2022)
  - Basque and Spanish poetry
  - Finetuned LMs on raw text that was automatically tagged
    - Syllable count
    - Rhyme-marking
  - Used this information to sample the LM
  - Evaluation
    - Filtering rate: among N generated samples, how many of them pass meter filter?
    - Perplexity
    - Human evaluation

Outline

- Me and myself
- Examples of generative machines
- Brief history of Automatic Poetry Generation
- The case of Basque, a European minority language
- **Challenges with current paradigms**
- Erato: Automatizing Poetry Evaluation

Challenges with the current paradigms

- Evaluation is done in a rather arbitrary way
  - Turing tests
  - Crowdsourcing
  - Asking colleagues
  - Asking experts
- No standard practice (accepted by community)

Outline

- Me and myself
- Examples of generative machines
- Brief history of Automatic Poetry Generation
- The case of Basque, a European minority language
- Challenges with current paradigms
- **Erato: Automatizing Poetry Evaluation**

Motivation

- Evaluating poetry is hard, expensive
- It poses a challenge for Automatic Poetry Generation
- We commonly resort to humans

- Given the subjective nature of the task, this is fair
- Besides, low correlation between human assessments and automatic metrics was noted (Hämäläinen and Alnajjar, 2021)

Hämäläinen, Mika, and Khalid Alnajjar. "**Human Evaluation of Creative NLG Systems: An Interdisciplinary Survey on Recent Papers.**" In *Proceedings of the 1st Workshop on Natural Language Generation, Evaluation, and Metrics (GEM 2021)*, pp. 84-95. 2021.

Motivation

Still, we argue that automatic metrics can at least support the creation of automatic poetry generation models.

How it all started

- This is not the first attempt to automatize poetry evaluation
- Several automatic methods in Gonçalo Oliveira et al., (2017)

Gonçalo Oliveira, Hugo, Raquel Hervás, Alberto Díaz, and Pablo Gervás. "**Multilingual extension and evaluation of a poetry generator.**" *Natural Language Engineering* 23, no. 6 (2017): 929-967.

Erato (Agirrezabal et al., 2023)

- A framework that aims to make the evaluation of poetry easier
- Poetry evaluation cannot rely on a single metric
- A number of Python scripts for the assessment of complementary aspects
  - Language specific vs. language independent
  - Single poem vs. poem collection
- Distinction of analysis vs. evaluation

Aspects of study

- Manurung (2004)
  - Poeticness, meaningfulness and grammaticality

- Poetic features
- Lexico-semantic features
- Fluency-related features
- Novelty-related features

Manurung, Hisar. "**An evolutionary algorithm approach to poetry generation.**" (2003), University of Edinburgh

Poetic features

Related to poeticness

- Poetic arrangement
  - Syllables, lines, stanzas
- Rhythm
- Rhymes

- Remember, though, that different cultures have different aspects of interest

Poetic features

Which of them is a haiku?
Any special rhymes somewhere?

The way a crow
Shook down on me
The dust of snow
From a hemlock tree

Has given my heart
A change of mood
And saved some part
Of a day I had rued.

Dust of Snow. By Robert Frost

Centre of equal daughters, equal sons,
All, all alike endear'd, grown, ungrown, young or old,
Strong, ample, fair, enduring, capable, rich,
Perennial with the Earth, with Freedom, Law and Love,
A grand, sane, towering, seated Mother,
Chair'd in the adamant of Time.

America,. By Walt Whitman

a party
where everyone says goodbye
then stays

A party. By John Brandi

Poetic features

Which of them is a haiku?
Any special rhymes somewhere?

The way a **crow**
Shook down on **me**
The dust of **snow**
From a hemlock **tree**

Has given my **heart**
A change of **mood**
And saved some **part**
Of a day I had **rued**.

Dust of Snow. By Robert Frost

Centre of equal daughters, equal sons,
All, all alike endear'd, grown, ungrown, young or old,
Strong, ample, fair, enduring, capable, rich,
Perennial with the Earth, with Freedom, Law and Love,
A grand, sane, towering, seated Mother,
Chair'd in the adamant of Time.

America,. By Walt Whitman

a party
where everyone says goodbye
then stays

A party. By John Brandi

Lexico-Semantic features

Related to meaningfulness

- Poems should convey certain message
- If we randomly assemble poem lines, low chance to get a coherent message
- Word choice: deviation from common language could be a measure of quality

Fluency-related features

Related to grammaticality

- If we do not use proper morphology or syntax, the conveyed message might be affected
- Poetic licenses are common, though

- We suggest to ask: Does this sound like poetry?

Upon a verdant lea, I chanced to stray,
Beside an aqueous stream that gurgled free,
Its waters limpid, sparkling in the day,
Reflecting skies and arboreal glee.

I ambled, lost in thought, so far and wide,
My mind a labyrinth of cogitation,
Seeking elusive truths, a mental tide,
In sesquipedalian contemplation.

The zephyrs whispered secrets through the leaves,
A dulcet breeze, a susurrating tale,
Inviting me to ponder and perceive,
The mystic truths that language can unveil.

With lexicon obscure, my thoughts take flight,
In sesquipedalian depths of night.

Sesquepedalian Sonnet, By ChatGPT (GPT3), 2023/08/15

I walked by a green field the other day,
Next to a stream that bubbled and flowed free,
Its water clear, sparkling in the sunlight's ray,
Showing the sky and the happy, leafy tree.

I walked and thought, my mind wandered so wide,
I was lost in thinking about many things,
Trying to find truths that often hide,
In deep thoughts that complexity brings.

The gentle winds whispered secrets softly,
A sweet breeze telling stories through the trees,
Inviting me to think and see closely,
The hidden truths that words can sometimes seize.

With complex words, my thoughts take their flight,
In deep thinking through the complex night.

A Long Words Sonnet, By ChatGPT (GPT3), 2023/08/15

Novelty-related features

Related to imagination (Colton, 2008)

- Is the poem new as itself, or does it repeat lines/words?
- Are the poems –written by somebody or something– new, or do they(he/she/it) write always the same or similar poems?

We check novelty as it was done in Gonçalo Oliveira et al. (2017)

Colton, Simon. "**Creativity Versus the Perception of Creativity in Computational Systems**." In *AAAI spring symposium: creative intelligent systems*, vol. 8, p. 7. 2008

Gonçalo Oliveira, Hugo, Raquel Hervás, Alberto Díaz, and Pablo Gervás. "**Multilingual extension and evaluation of a poetry generator**." *Natural Language Engineering* 23, no. 6 (2017): 929-967.

Round and round it goes,
The story that life knows.
Endless echoes play,
In night and light of day.

Round and round it goes,
A dance life only shows.
Endless echoes ring,
Moments they always bring.

Round and round it goes,
Through highs and all the lows.
Endless echoes spin,
Life's melody within.

Round and round it goes,
A tale that forever flows.
Endless echoes sound,
In hearts, they can be found.

Endless Echoes, By ChatGPT (GPT3), 2023/08/15

Implemented modules

Poetic features	Novelty features	Lexico/semantic features	Fluency features
-Stanza counter -Line counter -*Syllable counter -*Rhythm (stresses) -*Rhyme count -*Rhyme richness	-Rouge	-+Topic classification -Type token ratio -Stylometry	-+P-Language Model (ALMOST)

*Language resource specific (Ling knowledge)
+Data resource specific

Implemented modules

Example:

https://github.com/manexagirrezabal/erato/blob/master/models/lindep/stanzaCounter.py

Case study

- Human and machine generated poetry
- English language
- Humans:
  - William Shakespeare, Emily Dickinson, Edgar Allan Poe
- Machine generated poetry
  - PoeTryMe API (Gonçalo Oliveira, 2012)
  - OpenAI GPT3

Gonçalo Oliveira, Hugo. "**PoeTryMe: a versatile platform for poetry generation.**" *Computational Creativity, Concept Invention, and General Intelligence* 1 (2012): 21.

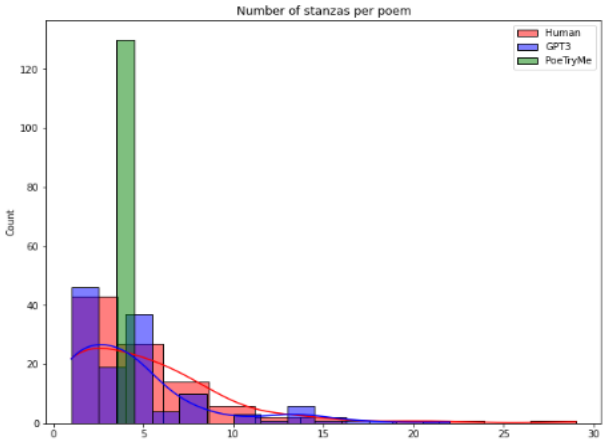
Case study

- We created poems using the following seed words (Same as in Gonçalo Oliveira (2017)):
  - Love
  - Artificial
  - Blue
  - Sing
  - Computer
  - Build
  - Football
  - New
  - Poetry
- 3 extra current words
  - Virus
  - Pandemic
  - Facemask

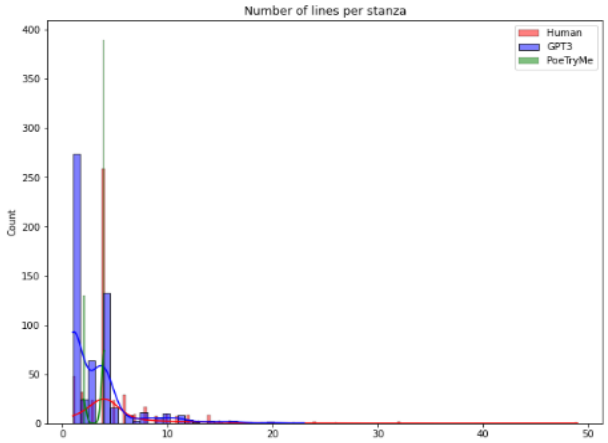
Studied aspects

- Poetic aspects
  - Stanza count
  - Lines per stanza
  - Syllables per stanza
  - Different rhyme patterns
  - Rhyme richness
- Novelty
- Lexico semantic aspects
  - Topic classification
  - Type-Token Ratio

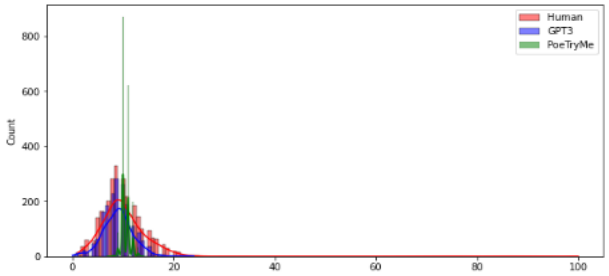
Poetic aspects



(a) Stanza count

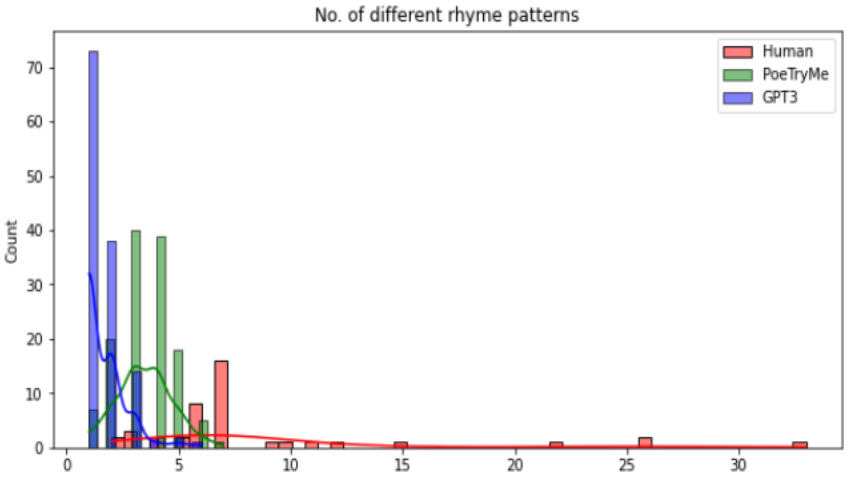


(b) Lines per stanza

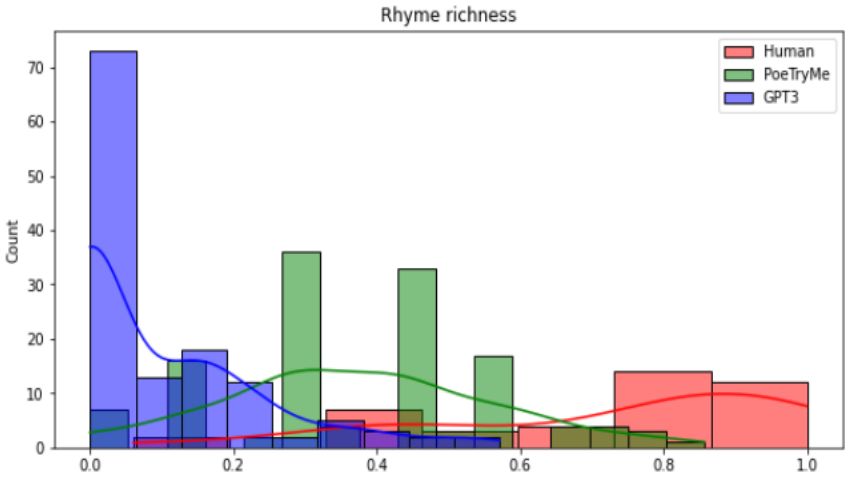


(c) Syllables per line

Poetic aspects



(a) Different rhyme patterns



(b) Rhyme richness

Novelty

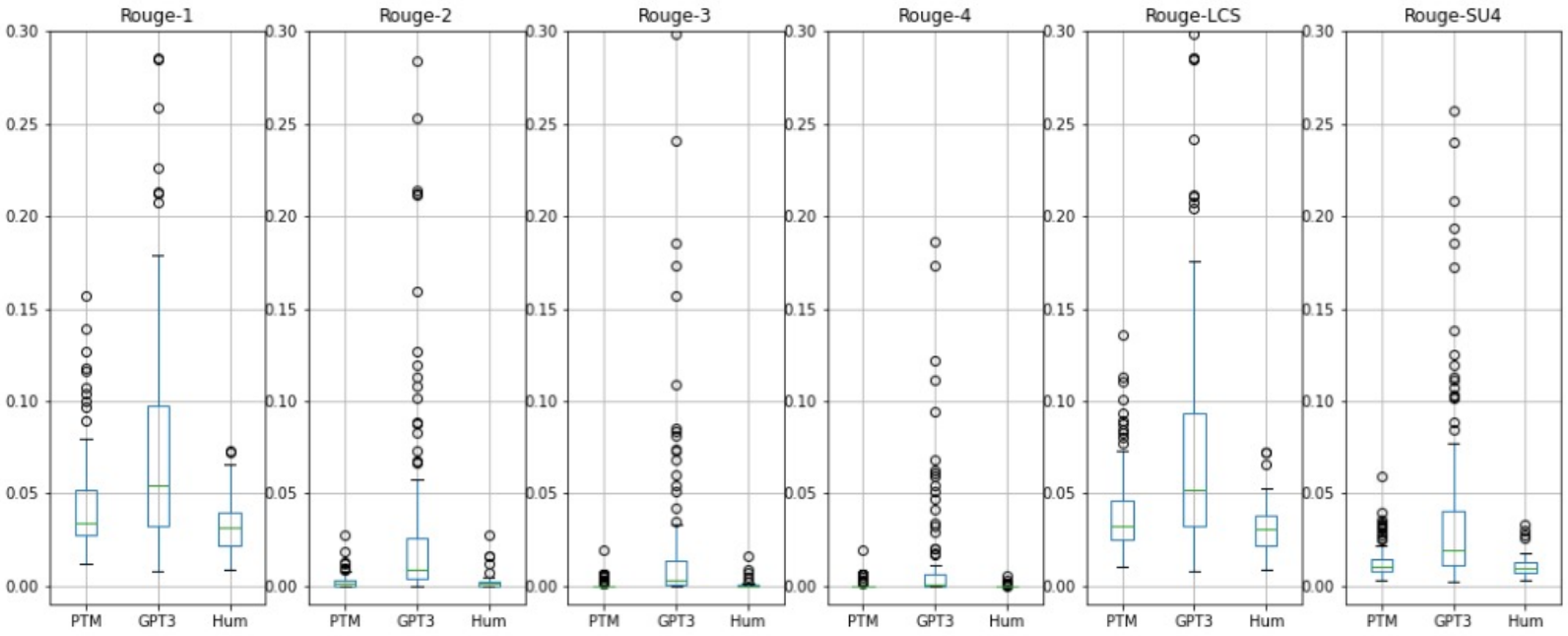


Fig. 3: Boxplot of Intra ROUGE scores for each of the three different authors.

Lexico-semantic aspects

- We retrieved poems for each seed word
- Macro F1-scores of 0.7 and 0.8 for both GPT3 and PoeTryMe
- GPT3
  - Very good at distinguishing poems about “blue”
- PoeTryMe in Spanish
  - Very bad at distinguishing poems about “blue”

Lexico-semantic aspects

- PoeTryMe produces generally well-sounding poems, but this can be done at the expense of less accurate semantics.
- Type-Token Ratio
  - GPT3: 0.130
  - PoeTryMe: 0.257
  - Human: 0.237
- It seems like GPT3 tends to repeat words more than PoeTryMe and humans

Conclusion

- History of computational models of poetry
- The case of Basque
- The challenge of evaluation
- An idea that hopefullly will help

References (1 of 2):

Agirrezabal, Manex, Hugo Gonçalo Oliveira, and Aitor Ormazabal. "Erato: Automatizing Poetry Evaluation.", EPIA Portuguese Conference on Artificial Intelligence, Faial island, Portugal (2023).

Barbieri, Gabriele, et al. "Markov Constraints for Generating Lyrics with Style." *Ecai*. Vol. 242. 2012.

Beheitt, Mohamed El Ghaly, and Moez Ben Haj Hmida. "Automatic Arabic Poem Generation with GPT-2." *ICAART (2)*. 2022.

Bena, Brendan, and Jugal Kalita. "Introducing Aspects of Creativity in Automatic Poetry Generation." *Proceedings of the 16th International Conference on Natural Language Processing*. 2019.

Chakrabarty, Tuhin, Vishakh Padmakumar, and He He. "Help me write a Poem-Instruction Tuning as a Vehicle for Collaborative Poetry Writing." In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pp. 6848-6863. 2022.

Chow, Evan. "jazzml repository", Github, https://github.com/evancchow/jazzml, 2014

Gervás, Pablo. "Generating poetry from a prose text: Creativity versus faithfulness." *Proceedings of the AISB'01 Symposium on Artificial Intelligence and Creativity in Arts and Science*. 2001.

Gillick, Jon, Kevin Tang, and Robert M. Keller. "Machine learning of jazz grammars." Computer Music Journal 34, no. 3 (2010): 56-66.

Hadjeres, Gaëtan, François Pachet, and Frank Nielsen. "Deepbach: a steerable model for bach chorales generation." In International conference on machine learning, pp. 1362-1371. PMLR, 2017.

Hopkins, Jack, and Douwe Kiela. "Automatically generating rhythmic verse with neural networks." *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2017.

Kim, Ji-Sung. "deepjazz repository", Github, https://github.com/jisungk/deepjazz, 2016

Lau, Jey Han, et al. "Deep-speare: A joint neural model of poetic language, meter and rhyme." *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2018.

References (2 of 2):

Levy, Robert P. "A computational model of poetic creativity with neural network as measure of adaptive fitness." *Proceedings of the ICCBR-01 Workshop on Creative Systems*. 2001.
Manurung, Hisar. "An evolutionary algorithm approach to poetry generation." (2004), PhD thesis
McCormack, J., Cruz Gambardella, C., Rajcic, N., Krol, S.J., Llano, M.T. and Yang, M., 2023, April. Is Writing Prompts Really Making Art?. In *International Conference on Computational Intelligence in Music, Sound, Art and Design (Part of EvoStar)* (pp. 196-211). Cham: Springer Nature Switzerland.
Ormazabal, Aitor, Mikel Artetxe, Manex Agirrezabal, Aitor Soroa, and Eneko Agirre. "PoeLM: A Meter-and Rhyme-Controllable Language Model for Unsupervised Poetry Generation." In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pp. 3655-3670. 2022.
Popescu-Belis, Andrei, Alex R. Atrio, Valentin Minder, Aris Xanthos, Gabriel Luthier, Simon Mattei, and Antonio Rodriguez. "Constrained language models for interactive poem generation." In *Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022)*. 20-25 June 2022, 2022.
Popescu-Belis, Andrei, Àlex R. Atrio, Bastien Bernath, Étienne Boisson, Teo Ferrari, Xavier Theimer-Lienhardt, and Giorgos Vernikos. "GPoeT: a language model trained for rhyme generation on synthetic data." In *Proceedings of the 7th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*. Association for Computational Linguistics, 2023.
Ritchie, Graeme, and Judith Masthoff. "The STANDUP 2 interactive riddle builder." In Proceedings of the Second International Conference on Computational Creativity. Universidad Autónoma Metropolitana, Mexico City, 2011.
Sneyers, Jon, and Danny De Schreye. "APOPCALEAPS: Automatic music generation with CHRiSM." In Proceedings of 22nd Benelux Conference on Artificial Intelligence (BNAIC'10), pp. 1-8. 2010.
Stock, Oliviero, and Carlo Strapparava. "Hahacronym: A computational humor system." In Proceedings of the ACL Interactive Poster and Demonstration Sessions, pp. 113-116. 2005.
Zhang, Xingxing, and Mirella Lapata. "Chinese poetry generation with recurrent neural networks." *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2014.

Spanish	Muchas gracias
Basque	Mila esker
தமிழ்	**மிக்க நன்றி**
English	Thanks a lot
Danish	Tusind tak

Questions?

SPELLL 2023

December 7th, 2023
Kongu Engineering College
Tamil Nadu, India

Recent years on automatic poetry generation: my own thoughts

Manex Agirrezabal
Associate Professor
University of Copenhagen