

Modèle pour la génération automatique de phrases littéraires

La rime comme artefact littéraire

Luis MORENO

Juan-Manuel TORRES

Roseli WEDEMANN

Séminaire

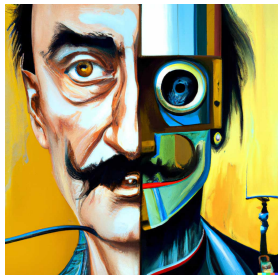
November 22, 2023

Presentation Overview

- 1 Introduction
Créativité computationnelle
- 2 Ressources linguistiques
Corpus littéraires
- 3 Modèles génératifs
Description
Exemples
- 4 Évaluation
Résultats
Perspectives

Créativité computationnelle

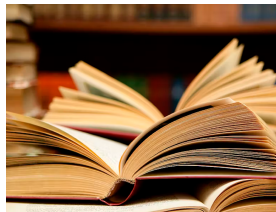
- Domaine de l'Intelligence Artificielle
- Étude du processus créatif ¹
 - Créativité exploratoire
 - Créativité transformationnelle
 - **Créativité combinatoire**



(a) Peinture



(b) Sculpture



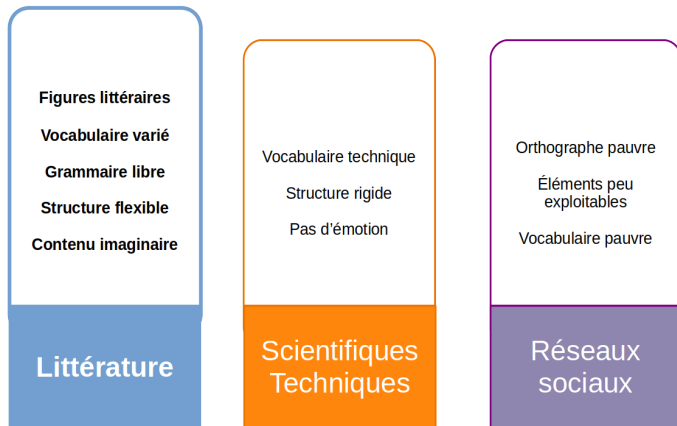
(c) Littérature

¹Boden. *The creative mind : Myths and Mechanism*. 2004

- Construction de corpus littéraires trilingues (**FR, ES, PT**)
- Production de phrases littéraires par **homosyntaxe**
- Incorporation de **rimes sémantiques**
- Création des **protocoles d'évaluation** pertinents
- **Définition** opératoire:

Une phrase artificielle littéraire est un objet linguistique respectant une structure grammaticale appartenant à un corpus littéraire. Elle possède un vocabulaire sélectionné par des contraintes lexicales et sémantiques ²

²Moreno-Jiménez et al. *Generación automática de frases literarias*. 2020



- Constitué **uniquement** de **documents littéraires (FR, ES, PT)**
3 , 4

³Moreno et al. *An Extended Bilingual Comparative Literary Corpus*. 2022

⁴Morgado et al. *A Corpus of Literature in Portuguese for NLP*. 2022

Caractéristiques de MegaLite

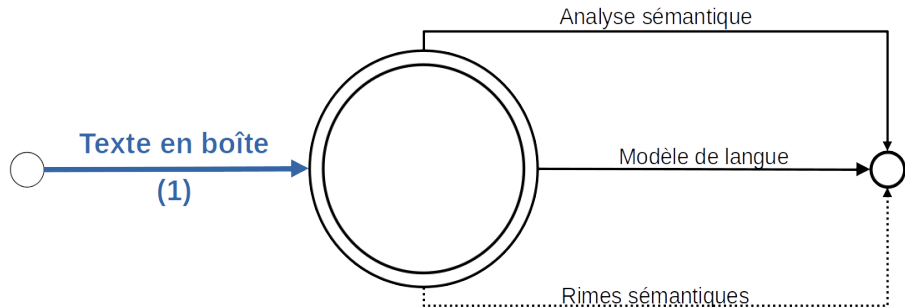
	Documents	Auteurs	Phrases	Mots
MegaLite^{ES}	5 075	1 329	14,7M	211M
MegaLite^{FR}	2 690	620	9,7M	182M
MegaLite^{PT}	4 311	1419	21M	252M
Total	12 076	3 368	45,4M	646M

Adéquat pour apprentissage automatique⁵

⁵Moreno et al. *MegaLite-2 : An Extended Bilingual Comparative Literary Corpus*. 2022; Morgado et al. *MegaLite-PT : A Corpus of Literature in Portuguese for NLP*. 2022

- Basée sur des **corpus littéraires**
- Guidée par **homosyntaxe**
- Guidée par un **contexte**
- Production des phrases artificielles **littéraires**
 - français
 - espagnol
 - portugais
- Production de **rimes sémantiques**
 - espagnol du Mexique

Première étape



- Méthode employée dans la génération de texte
 - Dialogues
 - Poèmes
 - Bulletin météo
 - Nouvelles
 - ...
- **Focaliser sur la sémantique**, pas sur la structure
- Principe de **l'homosyntaxe**

Génération de Structures Grammaticales partiellement-Vides (SGV)

f = La maison reste fraîche



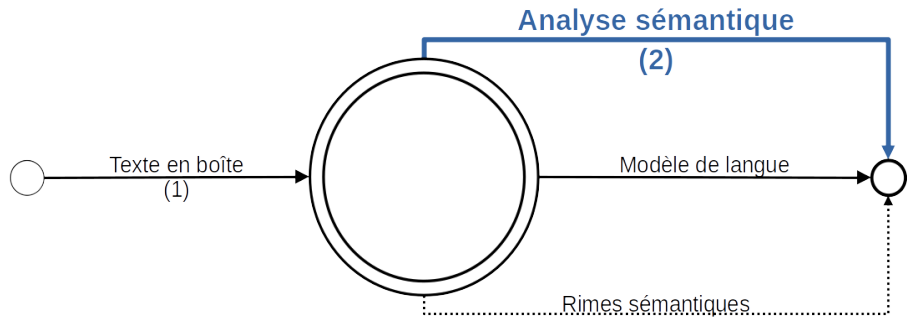
Remplacement de **mots lexicaux** par des **étiquettes grammaticales**

SGV = La NCFS VMIP AQ0FS

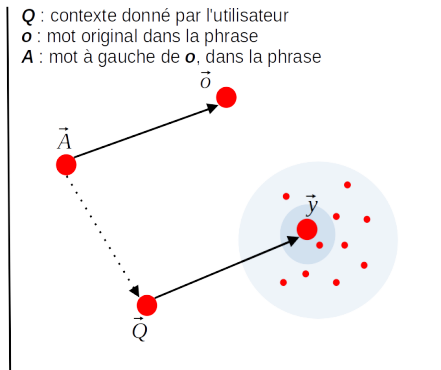
Analyse morphosyntaxique⁶

⁶Padró and Stanilovsky. *Freeling 3.0: Towards wider multilinguality*. 2012

Deuxième étape



Analyse sémantique (plongements lexicaux)

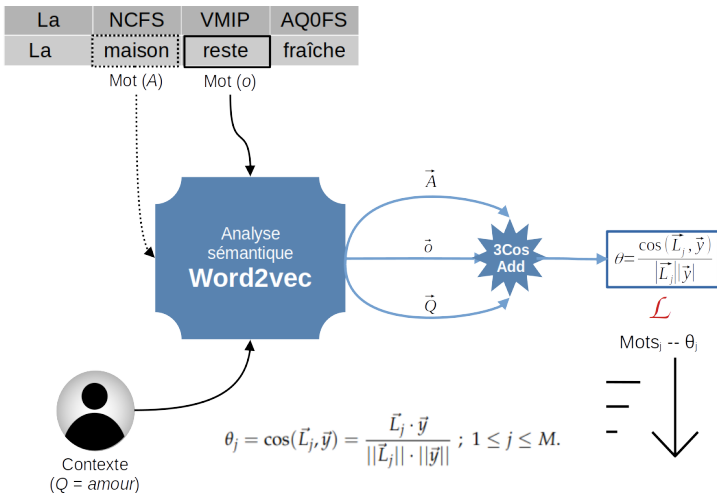


Quel est le mot qui est similaire à \vec{Q} dans le même sens que \vec{A} est similaire à \vec{o} ?

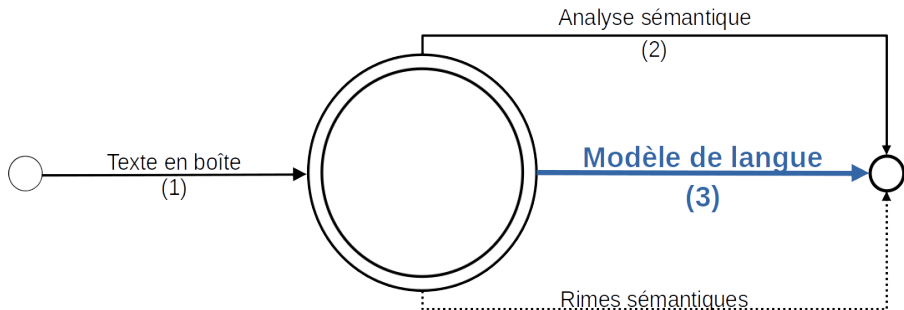
Analogie de type 3CosAdd ⁷

⁷Drozd, Gladkova, and Matsuoka. *Word embeddings, analogies, and machine learning : Beyond king-man+woman= queen* . 2016

Analyse sémantique

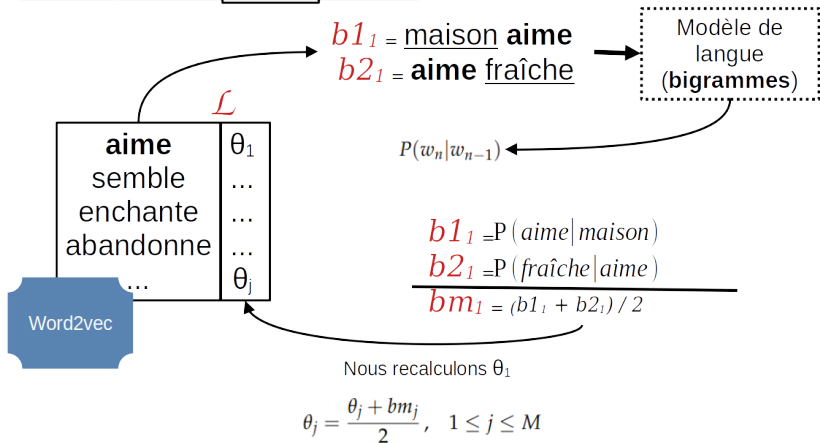


Troisième étape



Analyse syntaxique (Modèle de langue)

La	NCFS	VMIP	AQ0FS
La	<u>maison</u>	<u>reste</u>	<u>fraîche</u>

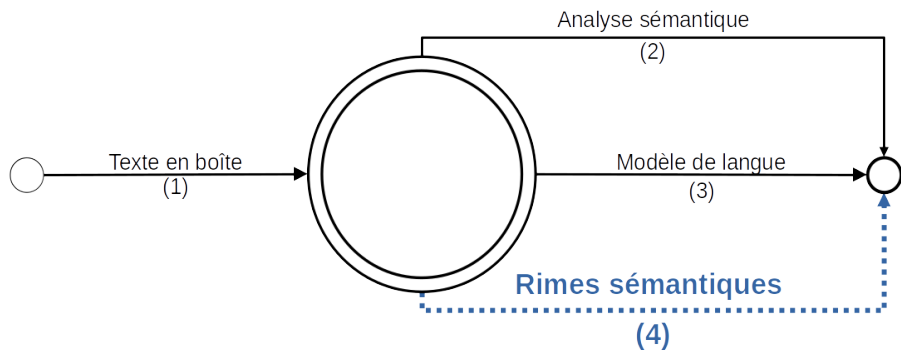


Même processus pour tous les mots ... **sauf la rime**

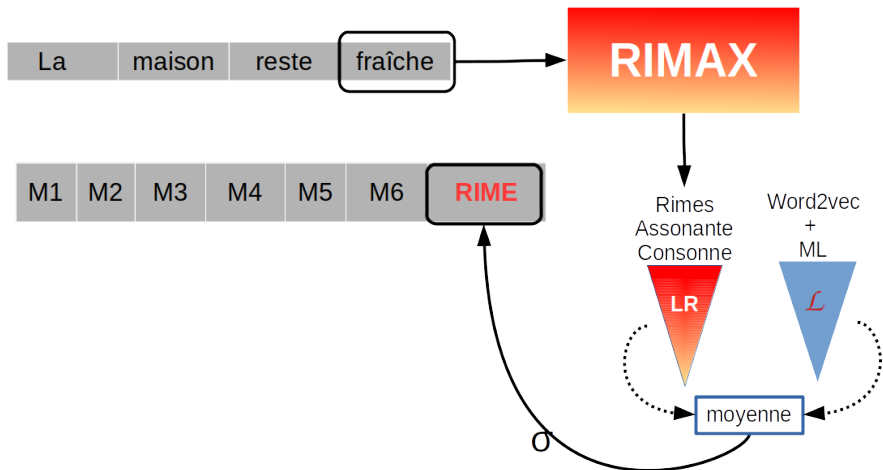
Exemples de phrases en français

- ① $f(\text{AMOUR}) = \text{Il n'y a pas de passion sans impulsif.}$
 - ② $f(\text{AMOUR}) = \text{Il n'y a pas d'affection sans estime.}$
 - ③ $f(\text{AMOUR}) = \text{Il n'y a ni aimante ni mélancolie en sérénité.}$
-
- ④ $f(\text{TRISTESSE}) = \text{En solitude, la première tendresse est la plus forte.}$
 - ⑤ $f(\text{TRISTESSE}) = \text{Il n'y a pas de confusion sans amour.}$
 - ⑥ $f(\text{TRISTESSE}) = \text{Il n'y a pas de liaison sans ordre.}$
-
- ⑦ $f(\text{AMITIÉ}) = \text{En union, la première sollicitude est la plus belle.}$
 - ⑧ $f(\text{AMITIÉ}) = \text{Il n'y a ni fraternelle ni faiblesse en impuissance.}$
 - ⑨ $f(\text{AMITIÉ}) = \text{Il n'y a pas de sympathie sans émoi.}$
-

Quatrième étape



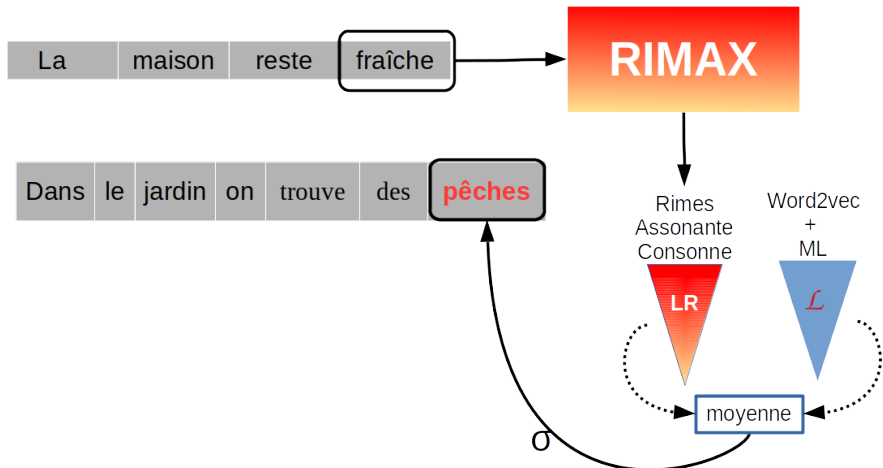
Rime sémantique



Algorithme de rime sémantique (Rimax)⁸

⁸Medina et Torres-Moreno. *RIMAX: Ranking Semantic Rhymes by calculating Definition Similarity*. 2019.

Rime sémantique



Algorithme de rime sémantique (Rimax)⁹

⁹Medina et Torres-Moreno. *RIMAX: Ranking Semantic Rhymes by calculating Definition Similarity*. 2019.

Exemple de phrases rimées en espagnol

- Q =amour

① **Jamás he sido más afectuoso; yo era ya un ofrecimiento.**

Je n'ai jamais été aussi affectueux ; j'étais déjà une offrande.

② **Corría el sol rápidamente hacia las cumbres que anteponían el firmamento.**

Le soleil se dirigeait vers les sommets qui surplombaient le firmament.

-
- Q =tristesse

① **El sol de mediodía encapota sobre la inexpresable niebla de mi bosque. — Le soleil de midi recouvre la brume inexprimable de ma forêt.**

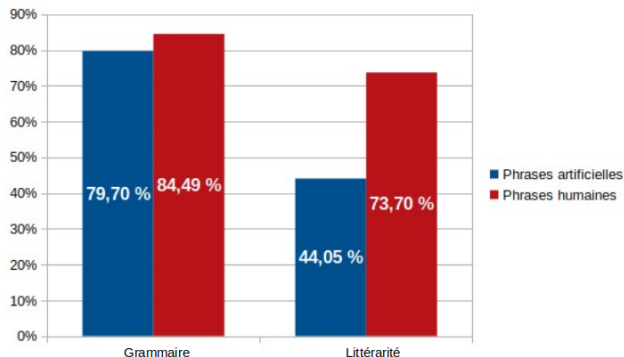
② **Subía el sol rápidamente hacia las desolaciones que limitaban el zopilote. — Le soleil montait rapidement vers les désolations qui bordaient le zopilote.**

- Q =lune
 - 1 **El sol de mediodía brilla sobre la impenetrable aurora de mi bosque.** — *Le soleil de midi brille sur l'aube impénétrable de ma forêt.*
 - 2 **Ascendía el sol rápidamente hacia las estrellas que limitaban el sol.** — *Le soleil s'est élevé rapidement vers les étoiles qui le bordent.*

Protocole d'évaluation (phrases en français)

- Phrases: **70 artificielles, 230 manuelles**
- **3 contextes** → {tristesse, amitié, amour}
- **18 évaluateurs** francophones, étudiants M2
- **Critères:**
 - **Grammaire** → {0 = mauvaise, 1 = bonne}
 - **Littérarité** → {0 = mauvaise, 1 = bonne}
 - **Relation contextuelle** → {0 = mauvaise, 1 = bonne, 2 = très bonne}
 - **Test de Turing** → {0 = machine, 1 = humaine}

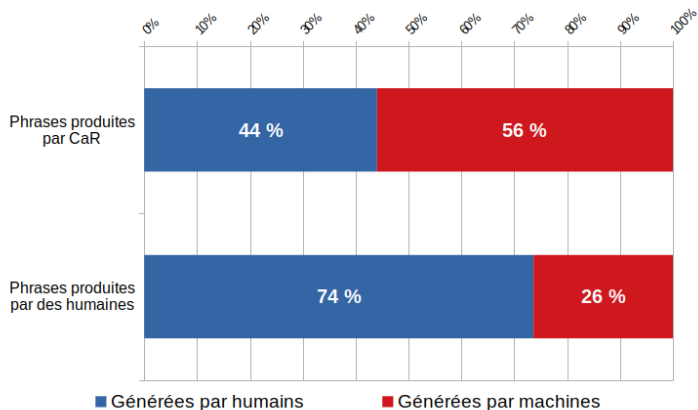
Grammaire et littérature :



Relation contextuelle :

- 24% mauvaise; 36% bonne; 40% très bonne

Test de Turing :



Rouge : phrases perçues comme artificielles

Bleu : phrases perçues comme écrites par des personnes

Protocole d'évaluation (rimes en espagnol)

- **44 paires** de phrases
- **11 contextes** → {amour, haine, tristesse, joie, soleil, lune, homme, femme, forêt, désert, mer}
- **6 évaluateurs** hispanophones, diplômés en littérature
- **2 critères**
 - **Rime ?** → {0 = non, 1 = oui}
 - **Relation sémantique** → {aucune, faible, acceptable, bonne, forte}

- **Perception de rimes**
 - 61% des couples de phrases
- **Relation sémantique**
 - *faible*

- **Analyse Conceptuelle**
 - Énergie textuelle[6] ¹⁰
 - Apprentissage profond (transformers + charge émotionnelle)
 - Apprentissage de styles
- **Utiliser autres modèles**
 - *BERT, CamemBERT*
 - *GPT*
 - Llama...
- **Production de long textes**
 - Vers un écrivain artificiel (?)

¹⁰Torres-Moreno, Molina, and Sierra. *Textual Energy of Associative Memories*. 2011

The End

Questions? Comments?