

Article

# Automatic Generation of Literary Sentences in French

Luis-Gil Moreno-Jiménez <sup>1,\*</sup> , Juan-Manuel Torres-Moreno <sup>1,\*</sup>  and Roseli Suzi. Wedemann <sup>2</sup> 

<sup>1</sup> Laboratoire Informatique d'Avignon, Université d'Avignon, 339 Chemin des Meinajariès, CEDEX 9, 84911 Avignon, France

<sup>2</sup> Instituto de Matemática e Estatística, Universidade do Estado do Rio de Janeiro, Rua São Francisco Xavier 524, Rio de Janeiro 20550-900, Brazil

\* Correspondence: luis-gil.moreno-jimenez@univ-avignon.fr (L.-G.M.-J.); juan-manuel.torres@univ-avignon.fr (J.-M.T.-M.)

**Abstract:** In this paper, we describe a model for the automatic generation of literary sentences in French. Although there has been much recent effort directed towards automatic text generation in general, the generation of creative, literary sentences that is not restricted to a specific genre, which we approached in this work, is a difficult task that is not commonly treated in the scientific literature. In particular, our present model has not been previously applied to the generation of sentences in the French language. Our model was based on algorithms that we previously used to generate sentences in Spanish and Portuguese and on a new corpus, which we constructed and present here, consisting of literary texts in French, called MEGALITE<sup>FR</sup>. Our automatic text generation algorithm combines language models, shallow parsing, the canned text method, and deep learning artificial neural networks. We also present a manual evaluation protocol that we propose and implemented to assess the quality of the artificial sentences generated by our algorithm, by testing if they fulfil four simple criteria. We obtained encouraging results from the evaluators for most of the desired features of our artificially generated sentences.

**Keywords:** computational creativity; literary sentences; automatic text generation; shallow parsing and deep learning



**Citation:** Moreno-Jiménez, L.-G.; Torres-Moreno, J.-M.; Wedemann, R.S. Automatic Generation of Literary Sentences in French. *Algorithms* **2023**, *16*, 142. <https://doi.org/10.3390/a16030142>

Academic Editors: Christos Troussas, Cleo Sgouropoulou, Akrivi Krouska, Ioannis Voyiatzis, Athanasios Voulodimos

Received: 23 December 2022

Revised: 14 February 2023

Accepted: 15 February 2023

Published: 6 March 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Creativity is a phenomenon that has been addressed for the past few decades by the *Natural Language Processing* (NLP) scientific research community [1–3], and it is not easily defined. In *The Creative Mind: Myths and Mechanisms* [4], Margaret Boden argues that the creative process is an intuitive path followed by humans to generate new artifacts that are valued for their novelty, significance to society, and beauty. Although there have been major advances resulting from the research efforts that have been put forward to develop automatic procedures for producing creative objects, there are difficulties and limitations related to the inherent complexity of understanding the creative process in the human mind [5]. These difficulties hamper the task of modeling and reproducing the cognitive abilities employed in the creative process. The search for automated processes capable of creatively generating artifacts has recently given rise to a field of research called *Computational Creativity* (CC), which offers interesting perspectives in various fields of art such as the visual arts, music, and literature, among others.

The creation of literature is especially intriguing and difficult, when compared to other types of *Automatic Text Generation* (ATG) tasks, mainly due to the fact that literary texts are not perceived in the same way by different persons, and the reader's emotional state may have an influence on his/her perception of the text. Literary documents often refer to imaginary, allegorical, or metaphorical worlds or situations, unlike journalistic or encyclopedic genres, which mainly describe factual situations or events, and it is often difficult to establish a precise boundary between general language and literary language. The difficulty of ensuring that a text generated by an ATG algorithm is literary is then

largely due to this inherent, subjective nature of literary perception. To approach the issue of the subjective ambiguity in literary perception, we used, as a guide, the understanding that literature uses a vocabulary that may strongly vary when compared to the one that is commonly used in everyday language and that it employs different writing styles and figures of speech, such as rhymes, anaphora, metaphors, euphemisms, and many others, which may lead to a possibly artistic, complex, and emotional text [6].

In this paper, we describe our model for the creative production of literary sentences based on homosyntax, where semantically different sentences present the same syntactic structure. This approach is different than paraphrasing and seeks to generate a new sentence that preserves a given syntactic structure by strongly modifying the semantics. Our proposal is based on a combination of language models, grammatical analysis, and semantic analysis based on an implementation of *Word2vec* [7].

Some of the ATG models proposed in the literature that are based entirely on neural networks such as [8–10] are capable of producing texts that seem to have been completely written by humans. However, it has been observed that these algorithms may generate some errors that are very difficult to control, such as a phenomenon similar to human hallucinations [11], which can generate grammatically incorrect or simply incoherent text. The model we propose is composed of a modular structure, consisting of a module for grammatical analysis and another for semantic analysis. This structure allowed us to better observe and control each step in the process of textual production, increasing the capacity to prevent and eventually correct errors that may be generated. Each module of our model is in some degree independent, allowing modifications to be made in one without negatively affecting the other. Moreover, our model is capable of generating creative, literary sentences that are not restricted to a specific genre.

In Section 2, we discuss examples of important current work that regards the production of literary text, with a focus on methods related to our own work. In Section 3, we present the literary corpus that we constructed and used for training our algorithms. Literary corpora, such as the one we introduce here, are difficult to produce and are essential for the production of literary text in computational creativity. We explain the algorithms that we developed for the production of literary sentences in Section 4. Some examples of the artificial sentences that were generated are shown in Section 5, along with the results obtained with evaluations performed by humans. Finally, we present our conclusions and propose some ideas for future investigations in Section 6.

## 2. Related Work

In this section, we present some of the recent work with a focus on the production of literary and non-literary text. Early attempts to model the process of automatic text production were based on stochastic models, such as the one presented by Szymanski and Ciota [12], where text is generated in Polish from  $n$ -character grams combined with Markov chains. Other techniques have also been used for text generation, such as the one proposed in [13], by Molins and Lapalme, which is based on grammatical structures for text generation in French and English. This method is known as *canned text*, and it provides a stable and consistent grammatical basis for text production. As the grammatical structure is previously determined, the efforts of the ATG algorithm are concentrated on selecting an appropriate vocabulary considering meaning.

More recently, many of the algorithms that have been proposed have been based on artificial deep neural networks.

In [8], the authors proposed a model capable of generating rhymes by using a language model based on transformers. The experiments were conducted in Spanish, and the evaluation was performed manually by mixing the artificially generated rhymes with rhymes written by humans. Only approximately 30% of the rhymes met the required criteria due to the fact that, in many cases, the rhyming words were repeated, which the model was not directly trained to prevent. In 37.3% of the cases, the human evaluators preferred the poems that their system generated when compared to those written by renowned poets.

In [14], the algorithm that was proposed by Van de Cruys generates coherent text using a Recurrent Neural Network (RNN) and a set of keywords. The keywords are used by the RNN to determine the context of the text that is produced. Another RNN approach, proposed by Clark, Ji, and Smith in [15], is used to produce narrative texts, such as fiction or news articles. Here, the entities mentioned in the text are represented by vectors, which are updated as the text is produced. These vectors represent different contexts and guide the RNN in determining which vocabulary to retrieve to produce a narrative. Fan, Lewis, and Dauphin improved the efficiency of standard RNNs, by using a convolutional architecture, to build coherent and fluent passages of text about a topic in story generation [16].

In [9], the authors proposed a model based on an encoder–decoder neural network for the generation of descriptions in the form of short paragraphs. They performed experiments using the Wikipedia database and employing various metrics such as ROUGE, BLEU, and pre-trained models such as BERT [17–19]. Some proposals based on GPT models have been used for poetry generation, such as [10]. In this work, the authors used a system that fine-tunes a pre-trained GPT-2 language model, to generate poems that express emotions and also elicit them in the readers, with a focus on poems that use the language of dreams (called dream poetry). The authors thus used a corpus of dream stories to train their model and showed that 87.5% and 85% of the generated poems were able to elicit emotions of sadness and joy, respectively, in readers.

Some very interesting proposals combine both the neural-network-based procedures and canned text methods [13]. This is also the case of Oliveira [20], who made a thorough study of models for generating poems automatically and proposed his own method of generation based on the use of canned text [21]. In [22], we find another canned-text-based proposal for the production of stanzas of verse in Basque poetry. Finally, in [23], Zhang and Lapata proposed an RNN for text structure learning for the production of Chinese poetry.

### 3. The MEGALITE<sup>FR</sup> Corpus

We now present the properties of the corpus that we constructed for the training and validation of our literary ATG models. Although the constitution and the use of specifically literary corpora are very important for developing and evaluating algorithms for literary production, the need for such corpora has been systematically underestimated. Literary corpora are needed mainly due to the possible level of complexity of literary discourse and the subjectivity and ambiguity aspects normally found in literary texts. The increased difficulty involved in producing these literary corpora usually induces users to resort to the use of corpora consisting of encyclopaedic, journalistic, or technical documents for textual production. In order to have a substantial and appropriate resource for the generation of literary sentences in French, we concentrated our efforts on the construction of a corpus consisting solely of French literature, called MEGALITE<sup>FR</sup>.

Our MEGALITE<sup>FR</sup> corpus consists of 2690 literary documents in French, written by 620 authors [24]. Most of the documents were originally written in French; some were translated to French; an important part of this corpus comes from the Bibebook (Site available under the Creative Commons BY-SA license, <http://www.bibebook.com>, accessed on 1 April 2021) ebook library. Some relevant properties of MEGALITE<sup>FR</sup> are shown in Table 1, and the distribution of works by genre is shown in Table 2. These properties suggest that it is a corpus with an adequate size for training machine learning algorithms for ATG, and we thus considered that the MEGALITE<sup>FR</sup> corpus, containing only literary documents, is well adapted to our purpose of producing literary sentences in French.

**Table 1.** Properties of MEGALITE<sup>FR</sup>.  $M = 10^6$  and  $K = 10^3$ .

	Documents	Phrases	Words	Characters	Authors
MEGALITE <sup>FR</sup>	2690	10 M	182 M	1081 M	620
Mean per document	-	3.6 K	67.9 K	401 K	

**Table 2.** Distribution of genres in MEGALITE<sup>FR</sup>.

	Plays	Poems	Narratives
MEGALITE <sup>FR</sup>	97 (3.61%)	55 (2.04%)	2538 (94.35%)

#### 4. A Literary ATG Model

We developed algorithms for the production of literary sentences in Spanish and Portuguese [6,25–28], and in this paper, we review some of their main features, which we adapted here for the production of literary sentences in French. The algorithms of our ATG model use keywords (queries) provided by the user as a semantic guide that determines the semantic context of the phrases that are produced. The model involves two basic steps:

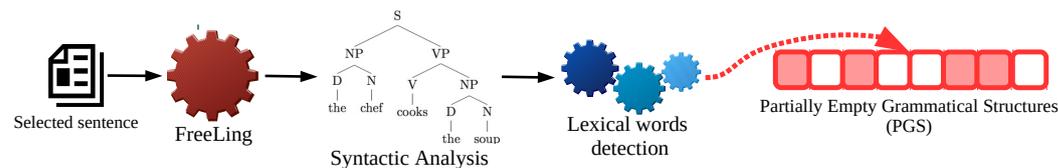
- The first step consists of generating Partially empty Grammatical Structures (PGSs) corresponding to the words of a chosen sentence. For this purpose, we implemented a procedure based on the *canned text* method, which is efficient for parsing in ATG tasks [29].
- In the second step, each tag (morpho-syntactic label) of the PGS generated in the first step is substituted by an alternative word, which is selected by means of a semantic analysis carried out with the aid of a procedure based on Word2vec [7].

##### 4.1. Canned-Text-Based Procedure

We prepared a set composed of sentences that were selected from MEGALITE<sup>FR</sup> manually, respecting the following criteria:

- Sentences must convey a specific, clear message that does not require a prior context to be understood.
- The length  $N$  of each sentence must be in the interval  $5 \leq N \leq 10$ .
- Sentences must have three or more lexical words.

We chose a literary sentence in French,  $f$ , that was parsed with FreeLing [30]. The lexical words (verbs, adjectives, and nouns) in  $f$  were then replaced by their morpho-syntactic labels (POS tags), and in this way, we generated a PGS. The functional words (prepositions, pronouns, auxiliary verbs, or conjunctions) in  $f$  were not substituted and remained in the sentence that was to be generated. Figure 1 illustrates this first step, where the functional words are represented by filled boxes in the PGS and the POS tags that replaced the lexical words are represented by the empty boxes.



**Figure 1.** Illustration of the first step of our algorithm based on the *canned text* method.

##### 4.2. Choice of Vocabulary to Generate the New Phrase

Once the PGS was formed in the previous step, each of its POS tags were replaced by a word chosen from a vocabulary produced by an algorithm that implements a semantic analysis supported by the Word2vec method [7]. For the replacement, we used the method 3CosAdd, introduced by Drozd, Gladkova, and Matsuoka in [31], which mathematically captures the analogy relations among words. This method considers the relationship between a set of words, for example *Italy*, *Rome*, and *Argentina*, and an unknown word,  $y$ . Assume, now, that *Argentina*, *Italy*, and *Rome* belong to the vocabulary of a corpus that was used to train Word2vec, called **CorpT**, and consequently,  $\vec{Argentina}$ ,  $\vec{Italy}$ , and  $\vec{Rome}$  are the respective embedding vectors found during training that are associated with these words. To find  $y$ , we then find a vector  $\vec{y}$  associated with a word in **CorpT**, such that  $\vec{y}$  is

closest to  $\vec{x} = \overrightarrow{Rome} - \overrightarrow{Italy} + \overrightarrow{Argentina}$ , according to the cosine similarity between  $\vec{x}$  and  $\vec{y}$  given by

$$\cos(\vec{x}, \vec{y}) = \frac{\vec{x} \cdot \vec{y}}{\|\vec{x}\| \|\vec{y}\|}. \tag{1}$$

The answer to this specific example is considered correct if  $\vec{y}$  is the embedding of *Buenos Aires* in the vocabulary of **CorpT**.

In the present work, we always refer to word embeddings that are produced by Word2vec when trained with MEGALITE<sup>FR</sup>, for our literary ATG algorithm. Let us consider the words  $Q, o$ , and  $A$ , where:

- $Q$  represents the context specified by the user;
- $o$  is the original word in  $f$ , whose POS tag is to be replaced to generate the new phrase;
- $A$  is the word previous to  $o$  in  $f$ .

These words are represented by the embeddings  $\vec{Q}, \vec{o}$ , and  $\vec{A}$  that are used to calculate:

$$\vec{x} = \vec{A} - \vec{o} + \vec{Q}. \tag{2}$$

Vector  $\vec{x}$  is thus produced by reducing the features of  $\vec{o}$  ( $\vec{x}$  is further away from  $\vec{o}$ ) and enhancing the features of  $\vec{Q}$  and  $\vec{A}$ . Next, we provided  $\vec{x}$  as the input to Word2vec and kept the first  $M = 4000$  outputs in a list  $\mathcal{L}$ ; in other words, list  $\mathcal{L}$  is formed by the  $M$  embeddings output from Word2vec that are closest to  $\vec{x}$ . The list  $\mathcal{L}$  then has  $M$  entries, where each entry,  $\vec{L}_k$ , corresponds to the embedding of a word,  $w_k$ , associated with  $\vec{x}$ . The value of  $M$  was chosen so that the execution time was not excessive, while still maintaining the good quality of the results of the experiments that we conducted. We next calculated the cosine similarity,  $\theta_k$ , between each  $\vec{L}_k$  and  $\vec{x}$ , according to Equation (1), so that

$$\theta_k = \cos(\vec{L}_k, \vec{x}), \tag{3}$$

and  $\mathcal{L}$  was then ranked in descending order of  $\theta_k$ .

If we substitute the first POS tag in the PGS,  $A$  is an empty word, so  $\vec{x} = -\vec{o} + \vec{Q}$ . For example, in the sentence  $f = I \overrightarrow{play} \overrightarrow{the} \overrightarrow{guitar}$  and for  $Q = love$ , when replacing the inflected verb  $o = play$ , we compute  $\vec{x} = -\overrightarrow{play} + \overrightarrow{love}$  to obtain the ordered list  $\mathcal{L}$ . Some of the words corresponding to embeddings obtained in  $\mathcal{L}$  for this example are *like, role, enchant, abandon*. The words in this list are then joined with the words adjacent to  $o$  in  $f$ , and an algorithm that analyses bigrams is used to choose which word will substitute  $o$ , as we describe next.

### Bigram Analysis

When choosing the word to replace the POS tag associated with word  $o$ , an important feature to consider is consistency. We used a bigram analysis to reinforce consistency and coherence, by estimating the conditional probability of having word  $w_n$  in the  $n$ th position in a sentence, given that another word  $w_{n-1}$  is adjacent to  $w_n$  on the left, expressed as

$$P(w_n|w_{n-1}) = \frac{P(w_n \wedge w_{n-1})}{P(w_{n-1})}. \tag{4}$$

The conditional probability calculated by Equation (4) corresponds to the frequency of occurrence of each bigram in MEGALITE<sup>FR</sup>. We considered the bigrams of MEGALITE<sup>FR</sup> formed only by lexical and functional words (punctuation, numbers, and symbols were ignored) to create the list,  $LB$ , which we used to calculate the frequencies.

We then formed two bigrams,  $b1_j$  and  $b2_j$ , for each  $\vec{L}_j \in \mathcal{L}$ . Bigram  $b1_j$  is formed by the word adjacent to the left of  $o$  in  $f$  concatenated with the word  $w_j$ , and  $b2_j$  is formed by concatenating  $w_j$  with the word adjacent to the right of  $o$  in  $f$ . To illustrate the formation of these bigrams, we show, in Figure 2, an example where  $f$  is the sentence in English “*I play the guitar*” and the word  $o = play$  will be substituted. Next, we calculated the mean value,  $bm_j$ , of the frequencies with which  $b1_j$  and  $b2_j$  occur in  $LB$ . If  $o$  is the last word in  $f$ ,

$bm_j$  corresponds to the frequency of  $b1_j$ . The value  $bm_j$  of each  $w_j$  is then combined with  $\theta_j$ , the cosine similarity obtained with Equations (1) and (3), and the list  $\mathcal{L}$  is reordered in descending order according to the new values:

$$\theta_j = \frac{\theta_j + bm_j}{2}, \quad 1 \leq j \leq M. \tag{5}$$

We then took the first *embedding* in  $\mathcal{L}$  that corresponds to the highest score,  $\theta_{max}$ , and the best word,  $w_{max}$ , to replace  $o$ . We finally performed a morphological analysis with FreeLing, in order to transform the selected word,  $w_{max}$ , to the correct form according to the inflection specified by its corresponding POS tag (respecting conjugations, gender, or number conversions).

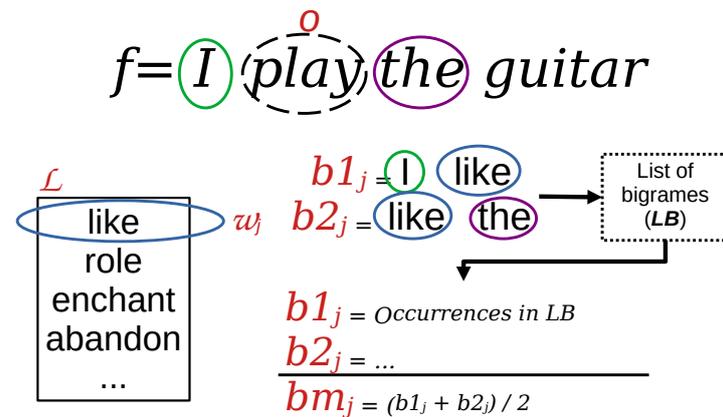


Figure 2. Illustration of the procedure based on bigram analysis for generating a new sentence.

With this procedure, we selected the word that will substitute  $o$  that is semantically closest to  $\vec{x}$ , with an analysis that was performed based on Word2vec, while maintaining coherence and consistency in the generated text by using the linguistic bigram analysis performed on the MEGALITE<sup>FR</sup> corpus. This procedure was repeated for each lexical word that was to be substituted in  $f$ , i.e., for each POS tag in the PGS. We, thus, obtained a newly created sentence that does not exist in MEGALITE<sup>FR</sup>. The model is illustrated in Figure 3.

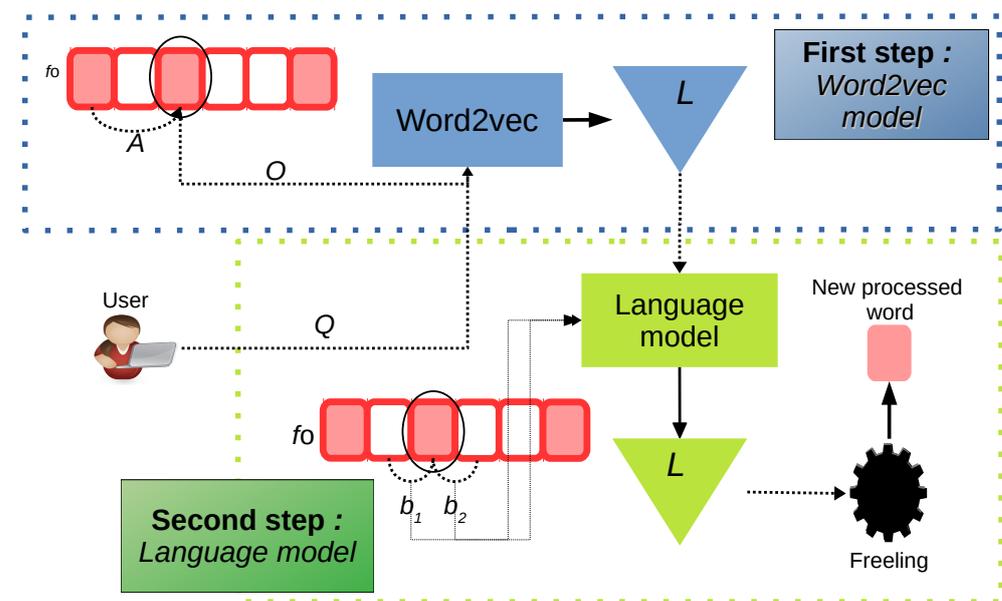


Figure 3. Illustration of the model for generating literary sentences.

## 5. Experiments

In this section, we present a manual protocol to evaluate the sentences produced by our model. Given the inherent subjectivity and ambiguity involved in literary perception, literary ATG algorithms developed for producing objects, in the domain of computational creativity, are frequently evaluated manually. We now show some examples and the results of the evaluation of sentences in French produced by our model, when using the MEGALITE<sup>FR</sup> corpus for training the Word2vec method, the generation of the PGSs, and the bigram analysis. The configuration of the hyper-parameters for the Word2vec procedure used in these experiments is specified as follows:

- The number of learning periods executed with MEGALITE<sup>FR</sup>, *iterations* = 10;
- The least number of times a word must appear in MEGALITE<sup>FR</sup>, to be included in the model vocabulary, *minimum count* = 3;
- The dimension of the embedding vectors, *vector size* = 100;
- The number of words adjacent to a specific word in a sentence that will be considered during the training phase, *window size* = 5.

Here are some examples of sentences in French, generated with three different PGS' and three queries, *Q*, which are shown in the format: sentence in French (*translation of sentence to English*).

For *Q* = love:

1. Il n'y a pas de passion sans impulsif. (*There is no passion without impulse.*)
2. Il n'y a pas d'affection sans estime. (*There is no affection without esteem.*)
3. Il n'y a ni aimante ni mélancolie en sérénité. (*There is neither love nor melancholy in serenity.*)
4. Il n'y a ni fraternelle ni inquiétude en anxiété. (*There is no brotherhood or concern in anxiety.*)

For *Q* = sadness:

1. En solitude, la première tendresse est la plus forte. (*In solitude, the first tenderness is the strongest.*)
2. Il n'y a pas de confusion sans amour. (*There is no confusion without love.*)
3. Il n'y a pas de liaison sans ordre. (*There is no connection without order.*)
4. En douleur, la première mélancolie est la plus grande. (*In pain, the first melancholy is the greatest.*)

For *Q* = friendship:

1. En union, la première sollicitude est la plus belle. (*In union, the first concern is the most beautiful.*)
2. Il n'y a ni fraternelle ni faiblesse en impuissance. (*There is no brotherhood or weakness in powerlessness.*)
3. Il n'y a pas de sympathie sans émoi. (*There is no sympathy without emotion.*)
4. Il n'y a ni amie ni honte en peur. (*There is no friend or shame in fear.*)

In these examples, we can observe reasonably coherent sentences with words belonging to the same semantic field. There are some small syntax errors that occurred within the FreeLing tokenization module, which we expect may be solved by a fine-grained, a posteriori analysis, based on regular expressions.

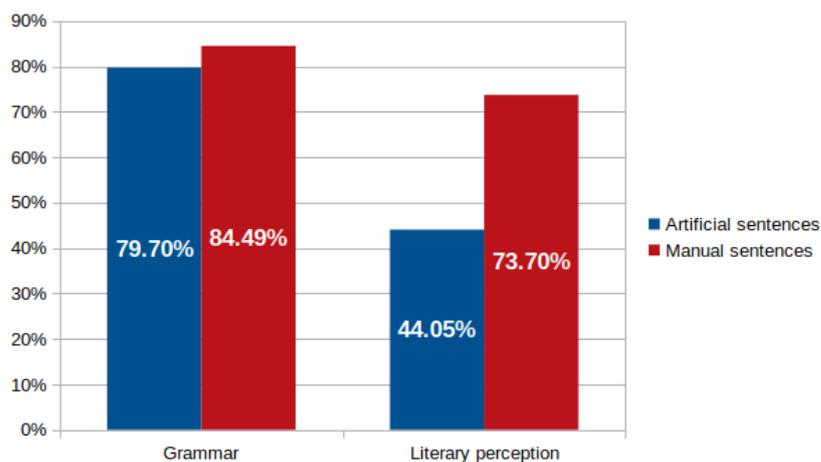
### *Evaluation Protocol and Results*

We developed an evaluation protocol with four criteria that included the Turing test. The evaluators were asked to evaluate the sentences considering: correct grammar, relation to context, literary perception, and the Turing test. The set of sentences that were presented to the evaluators consisted of sentences generated by our literary ATG algorithm mixed with sentences generated by human writers, and this allowed us to compare our artificially generated sentences with non-artificial, human sentences.

We generated  $p = 70$  artificial sentences for the evaluation set with our literary ATG model and with the following contexts (queries): tristesse (sadness), amitié (friendship), and amour (love). We also asked 18 native French speakers, all of which have a university Master's degree, to write sentences. These sentences written by humans were also based on the same previously generated PGSs that we used to generate the artificial sentences. Therefore, each person received six PGSs and was asked to write sentences by replacing the POS tags in the PGSs with French words, respecting the grammatical properties indicated by the POS tags. We, thus, obtained  $p' = 230$  literary sentences written by human beings. We then randomly mixed the  $p$  artificial sentences with the  $p'$  manually created sentences and obtained a total of  $P = 300$  sentences that constituted the evaluation set.

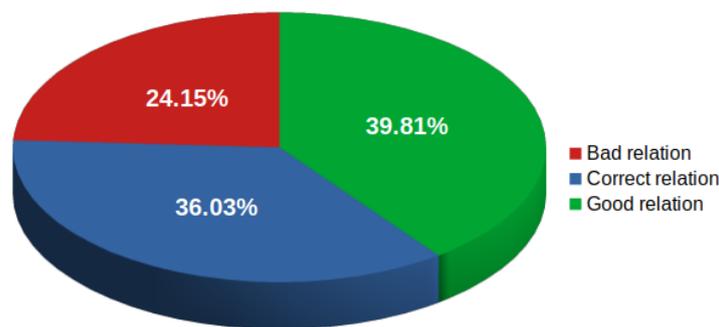
We then sent the sentences in the evaluation set to the evaluators (or annotators), many of which had also participated in the writing efforts, so we made sure that the annotators did not evaluate their own sentences. For the assessment of correct grammar and literary perception, the raters were asked to indicate if the sentences were bad or good (a binary response), according to their personal perception. The relationship to the context was assessed with three categories: poor relationship, good relationship, and very good relationship. Finally, for the Turing test, the raters should indicate if they considered that the sentence was generated by a machine or not.

The results that we received from the evaluators were encouraging. Approximately 80% of the artificial sentences were found to be grammatically correct, which is similar to the case of the sentences written by humans where 84% were classified as correct. The standard deviation calculated for the classification of the artificial sentences was 0.22 and indicated an acceptable level of agreement between evaluators. For the evaluation of literary perception, 44% of the artificial sentences were perceived as literary against 73% of the human sentences, and although the score obtained by the artificial sentences may seem low, it was still encouraging, when we consider the ambiguity inherent to literary perception among people (see Figure 4). However, this result indicated that there is still room for improvement regarding the literariness of the sentences generated by our ATG procedures.



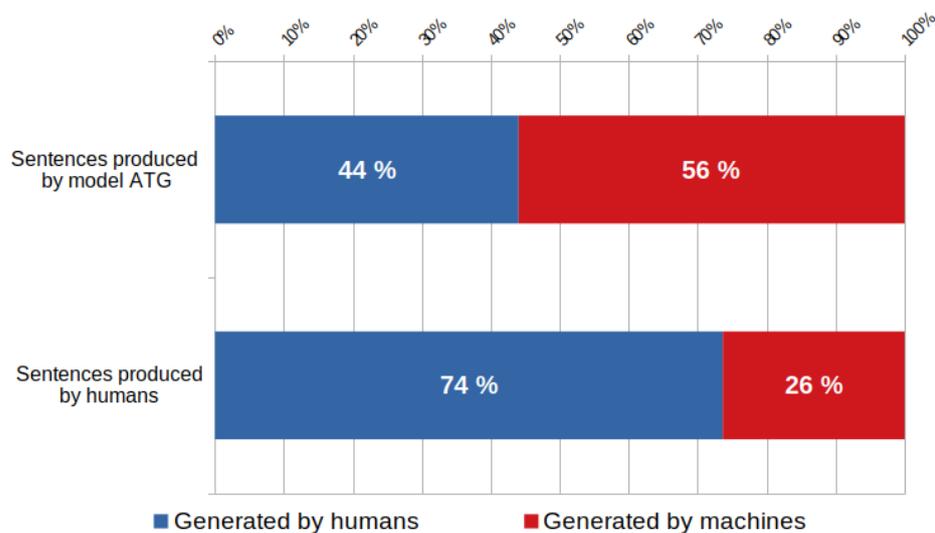
**Figure 4.** Results of the evaluation of grammatical correctness and literariness of the French sentences generated by humans and by our ATG model.

In Figure 5, we display the results obtained for the criterion that regards the semantic relationship of the artificial sentences with the contexts given by the user (the queries). It can be seen that the raters considered that 24% of the artificial sentences had a bad relation with the given context. On the other hand, the raters considered that 36% of the sentences had a good relation to the context, and almost 40% of the sentences were considered to have a very good relation to the expected context.



**Figure 5.** Results for the assessment of the semantic relation between the artificially generated sentences and the given query.

Finally, with respect to the Turing test, 44% of the artificial sentences were perceived as sentences generated by humans. Although, at first, it may seem otherwise, this result was also encouraging, when we consider that 26% of the human sentences were perceived as being artificial (see Figure 6). Given that only 74% of human sentences were perceived as sentences written by humans, the difference between the positive score of the ATG sentences and the human sentences was 30%, and this gap suggested that possible improvements to our algorithm may produce a similar perception of natural human sentences by evaluators for our ATG model.



**Figure 6.** Results of the Turing test applied to the sentences produced by the ATG model. Red is used for the sentences perceived as artificial and blue for the sentences perceived as written by humans).

Some basic features of our work, including our main aims and important characteristics of our algorithms and training corpora, did not allow a formal comparison with other related work. Our main purpose in recent work and also here was the generation of literary sentences, and therefore, the training corpora we constructed and used are strictly literary. In contrast, most of the other works found in the recent scientific, specifically the ATG literature, analyzed texts from other types of sources, such as social networks and journalistic and scientific literature [9,13,32]. Works based on the study of non-literary texts are evaluated automatically using metrics such as BLEU or ROUGE. The sentences generated with our proposal cannot be evaluated in the same way, since criteria such as literary perception necessarily need to be evaluated by humans, and this is why we employed a manual evaluation of a limited number of sentences. Among the works that studied the generation of literary texts, we found some such as [10,23,27], where the goal was the production of poems or short stories. We, however, avoided the restriction of specific literary genres, as our goal was to produce sentences that can be perceived as

literary, by evoking a subjective experience and emotion in the reader through a carefully selected vocabulary. We expect to extend the generation of sentences to paragraphs and eventually to longer texts in the future. We thus note that the aims and criteria evaluated in our literary ATG model were not the same as those of other ATG proposals that we found and cited previously.

## 6. Conclusions

In the present work, we described a model composed of algorithms that is capable of producing literary sentences in French, based on the canned text and Word2vec methods, and a bigram analysis of the MEGALITE<sup>FR</sup> corpus that we constructed and presented here. The model produces sentences with linguistic features frequently observed in literary texts written by humans. Moreover, the grammar analysis module of our ATG system, which was based on the canned text method, enabled the simulation of a specific literary style and also of the writing style and psychological trait intended by an author, as discussed in [25]. This module can thus be used to simulate specific literary genres and styles, without affecting the semantic analysis of the model.

Approximately 80% of the sentences generated by our model were perceived as being grammatically correct. In the case of the relationship to an expected context, approximately 76% of the sentences were classified as having a good contextual relationship. As for the Turing test, only 44% of the artificial sentences were perceived as having been written by a human. Although, at first, this seems not to be a good result, it is necessary to take into account the difficulty of this test, considering that the evaluation was also subjective. According to the literature, misleading humans in 50% of the evaluated texts is considered an acceptable result [33]. If we note that, in our tests, only 74% of manually produced sentences were perceived as having been written by a human being, the percentage of artificial sentences generated by our model that were considered to have been written by a human was 30% lower, and this can be considered as a margin of improvement for our algorithms in our future developments. It is also possible to envision plans to produce other literary structures such as rhymes and paragraphs with these types of algorithms.

### *Future Works*

We believe it is possible to enhance our algorithms so that they may be capable of generating text with a longer length, such as longer sentences and paragraphs. For this purpose, we regard using techniques based on the concept of textual energy [34] or neural network models such as transformers, although without relying entirely on the latter. This could help us to confirm the working hypothesis that it is possible to automatically generate long texts with classical methods, without relying entirely on neural network models, where one does not understand the full logic of the generation process. These methods can be applied to generate text in French, Spanish, and Portuguese with the use of the MEGALITE corpora, MEGALITE<sup>FR</sup>, MEGALITE<sup>SP</sup>, and MEGALITE<sup>PT</sup>. We also plan to increase the size of the MEGALITE corpora and tune the algorithms to specialize them for producing literary texts in specific genres or regarding specific historical periods or classes of emotions.

**Author Contributions:** conceptualization, L.-G.M.-J. and J.-M.T.-M.; data curation, J.-M.T.-M.; formal analysis, R.S.W.; funding acquisition, J.-M.T.-M. and R.S.W.; investigation, L.-G.M.-J. and J.-M.T.-M.; methodology, L.-G.M.-J., J.-M.T.-M. and R.S.W.; project administration, J.-M.T.-M.; resources, J.-M.T.-M.; software, L.-G.M.-J. and J.-M.T.-M.; supervision, J.-M.T.-M. and R.S.W.; validation, L.-G.M.-J., J.-M.T.-M. and R.S.W.; writing—original draft, L.-G.M.-J., J.-M.T.-M. and R.S.W.; writing—review and editing, L.-G.M.-J., J.-M.T.-M. and R.S.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was funded by the Consejo Nacional de Ciencia y Tecnología (Conacyt, Mexico), Grant Number 661101, and partially funded by the Université d'Avignon, Laboratoire Informatique d'Avignon (LIA), Agorantic program (France). It also received partial financial support from the Brazilian science funding agencies: Conselho Nacional de Desenvolvimento Científico e Tecnológico

(CNPq), Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ), and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior-Brasil (CAPES).

**Data Availability Statement:** The MEGALITE<sup>FR</sup> corpus can be freely downloaded from the Ortolang website <https://www.ortolang.fr/market/corpora/megalite#!>, accessed on 3 January 2023. The versions in Spanish and Portuguese are also available.

**Acknowledgments:** R.S.W. is also grateful for the kind hospitality of the Laboratoire Informatique d'Avignon (LIA), Université d'Avignon, where she participated in this research.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Sharples, M. *How We Write: Writing as Creative Design*; Routledge: London, UK, 1996.
2. Sridhara, G.; Hill, E.; Muppaneni, D.; Pollock, L.; Vijay-Shanker, K. Towards automatically generating summary comments for Java methods. In Proceedings of the IEEE/ACM International Conference on Automated Software Engineering, Antwerp, Belgium, 20–24 September 2010; Association for Computing Machinery: New York, NY, USA, 2010; pp. 43–52. [CrossRef]
3. Mikolov, T.; Zweig, G. Context dependent recurrent neural network language model. In Proceedings of the 2012 IEEE Spoken Language Technology Workshop (SLT), Miami, FL, USA, 2–5 December 2012; IEEE: New York, NY, USA, 2012; pp. 234–239. [CrossRef]
4. Boden, M.A. *The Creative Mind: Myths and Mechanisms*; Routledge: London, UK, 2004.
5. de Carvalho, L.A.V.; Mendes, D.Q.; Wedemann, R.S. Creativity and Delusions: The dopaminergic modulation of cortical maps. In Proceedings of the 2003 International Conference on Computational Science (ICCS 2003), St. Petersburg, Russia, 2–4 June 2003; Sloot, P.M.A., Abramson, D., Bogdanov, A.V., Dongarra, J.J., Zomaya, A.Y., Gorbachev, Y.E., Eds.; Springer: Berlin/Heidelberg, Germany, 2003; Volume 2657; pp. 511–520. [CrossRef]
6. Moreno-Jiménez, L.G.; Torres-Moreno, J.M.; Wedemann, R.S.; SanJuan, E. Generación automática de frases literarias. *Linguamática* **2020**, *12*, 15–30. [CrossRef]
7. Mikolov, T.; Yih, W.t.; Zweig, G. Linguistic regularities in continuous space word representations. In Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL—HLT 2013), Atlanta, GA, USA, 9–14 June 2013; Association for Computational Linguistics: Stroudsburg, PA, USA, 2013; pp. 746–751.
8. Ormazabal, A.; Artetxe, M.; Agirrezabal, M.; Soroa, A.; Agirre, E. PoELM: A meter-and rhyme-controllable language model for unsupervised poetry generation. *arXiv* **2022**, arXiv:2205.12206.
9. Ta, H.T.; Rahman, A.B.S.; Majumder, N.; Hussain, A.; Najjar, L.; Howard, N.; Poria, S.; Gelbukh, A. WikiDes: A Wikipedia-based dataset for generating short descriptions from paragraphs. *Inf. Fusion* **2023**, *90*, 265–282. [CrossRef]
10. Bena, B.; Kalita, J. Introducing aspects of creativity in automatic poetry generation. In Proceedings of the 16th International Conference on Natural Language Processing, Alicante, Spain, 28–30 June 2019; NLP Association of India: International Institute of Information Technology: Hyderabad, India, 2019; pp. 26–35. [CrossRef]
11. Ji, Z.; Lee, N.; Frieske, R.; Yu, T.; Su, D.; Xu, Y.; Ishii, E.; Bang, Y.; Madotto, A.; Fung, P. Survey of hallucination in natural language generation. *ACM Comput. Surv.* **2022**, 1–36. [CrossRef]
12. Szymanski, G.; Ciota, Z. Hidden Markov models suitable for text generation. In Proceedings of the WSEAS International Conference on Signal, Speech and Image Processing (WSEAS ICSSIP 2002), Budapest, Hungary, 10–12 December 2002; Mastorakis, N., Kluev, V., Koruga, D., Eds.; WSEAS Press: Athens, Greece, 2002; pp. 3081–3084.
13. Molins, P.; Lapalme, G. JSrealB: A bilingual text realizer for Web programming. In Proceedings of the 15th European Workshop on Natural Language Generation (ENLG), Brighton, UK, September 2015; Association for Computational Linguistics: Stroudsburg, PA, USA, 2015; pp. 109–111. [CrossRef]
14. Van de Cruys, T. Automatic poetry generation from prosaic text. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 5–10 July 2020; Association for Computational Linguistics: Stroudsburg, PA, USA, 2020; pp. 2471–2480. [CrossRef]
15. Clark, E.; Ji, Y.; Smith, N.A. Neural text generation in stories using entity representations as context. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL—HLT 2018), New Orleans, LA, USA, 1–6 June 2018; Association for Computational Linguistics: Stroudsburg, PA, USA, 2018; Volume 1; pp. 2250–2260. [CrossRef]
16. Fan, A.; Lewis, M.; Dauphin, Y. Hierarchical neural story generation. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, Melbourne, Australia, 15–20 July 2018; Association for Computational Linguistics: Stroudsburg, PA, USA, 2018; Volume 1; pp. 889–898. [CrossRef]
17. Lin, C.Y. ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*; Association for Computational Linguistics: Barcelona, Spain, 2004; pp. 74–81.

18. Papineni, K.; Roukos, S.; Ward, T.; Zhu, W.J. BLEU: A method for automatic evaluation of machine translation. In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL'02), Philadelphia, PA, USA, 7–12 July 2002; pp. 311–318. [\[CrossRef\]](#)
19. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, MN, USA, 2–7 June 2019; Association for Computational Linguistics: Stroudsburg, PA, USA, 2019; Volume 1 (Long and Short Papers), pp. 4171–4186. [\[CrossRef\]](#)
20. Oliveira, H.G. A survey on intelligent poetry generation: Languages, features, techniques, reutilisation and evaluation. In Proceedings of the 10th International Conference on Natural Language Generation (CNLG), Santiago de Compostela, Spain, 4–7 September 2017; Association for Computational Linguistics: Stroudsburg, PA, USA, 2017; pp. 11–20. [\[CrossRef\]](#)
21. Oliveira, H.G.; Cardoso, A. Poetry generation with PoeTryMe. In *Computational Creativity Research: Towards Creative Machines*; Atlantis Press: Paris, France, 2015; Volume 7; pp. 243–266. [\[CrossRef\]](#)
22. Agirrezabal, M.; Arrieta, B.; Astigarraga, A.; Hulden, M. POS-tag based poetry generation with WordNet. In Proceedings of the 14th European Workshop on Natural Language Generation (ENLG), Sofia, Bulgaria, 8–9 August 2013; Association for Computational Linguistics: Stroudsburg, PA, USA, 2013; pp. 162–166.
23. Zhang, X.; Lapata, M. Chinese poetry generation with recurrent neural networks. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 24–29 October 2014; Association for Computational Linguistics: Stroudsburg, PA, USA, 2014; pp. 670–680. [\[CrossRef\]](#)
24. Moreno-Jiménez, L.G.; Torres-Moreno, J.M. MegaLite-2: An extended bilingual comparative literary corpus. In *Proceedings of the Intelligent Computing*; Arai, K., Ed.; Springer: Cham, Switzerland, 2022; pp. 1014–1029. [\[CrossRef\]](#)
25. Moreno-Jiménez, L.G.; Torres-Moreno, J.M.; Wedemann, R.S. Literary natural language generation with psychological traits. In *Natural Language Processing and Information Systems—NLDB 2020*; Métails, E., Meziane, F., Horacek, H., Cimiano, P., Eds.; Springer: Cham, Switzerland, 2020; Volume 12089; pp. 193–204. [\[CrossRef\]](#)
26. Moreno-Jiménez, L.G.; Torres-Moreno, J.M.; Wedemann, R.S. Generación automática de frases literarias: Un experimento preliminar. *Proces. Del Leng. Nat.* **2020**, *65*, 29–36.
27. Moreno-Jiménez, L.G.; Torres-Moreno, J.M.; Wedemann, R.S. A preliminary study for literary rhyme generation based on neuronal representation, semantics and shallow parsing. In Proceedings of the XIII Brazilian Symposium in Information and Human Language Technology and Collocated Events (STIL 2021), Online, 29 November–3 December 2021; Ruiz, E.E.S., Torrent, T.T., Eds.; Sociedade Brasileira de Computação: Rio de Janeiro, Brasil, 2021; pp. 190–198. [\[CrossRef\]](#)
28. Morgado, I.; Moreno-Jiménez, L.G.; Torres-Moreno, J.M.; Wedemann, R.S. MegaLite<sup>PT</sup>: A corpus of literature in Portuguese for NLP. In Proceedings of the 11th Brazilian Conference on Intelligent Systems, Part II (BRACIS 2022), Campinas, Brasil, 28 November–1 December 2022; LNAI; Xavier-Junior, J.C., Rios, R.A., Eds.; Springer International Publishing: Cham, Switzerland, 2022; Volume 13654; pp. 251–265. [\[CrossRef\]](#)
29. van Deemter, K.; Theune, M.; Kraemer, E. Real versus template-based natural language generation: A false opposition? *Comput. Linguist.* **2005**, *31*, 15–24. [\[CrossRef\]](#)
30. Padró, L.; Stanilovsky, E. FreeLing 3.0: Towards wider multilinguality. In Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12), Istanbul, Turkey, 23–25 May 2012; European Language Resources Association (ELRA): Luxembourg, 2012; pp. 2473–2479.
31. Drozd, A.; Gladkova, A.; Matsuoka, S. Word embeddings, analogies, and machine learning: Beyond king - man + woman = queen. In Proceedings of the 26th International Conference on Computational Linguistics: Technical Papers (COLING 2016), Osaka, Japan, 11–16 December 2016; The COLING 2016 Organizing Committee: Osaka, Japan, 2016; pp. 3519–3530.
32. Kiddon, C.; Zettlemoyer, L.; Choi, Y. Globally coherent text generation with neural checklist models. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP'16), Austin, TX, USA, 1–5 November 2016; Association for Computational Linguistics: Stroudsburg, PA, USA, 2016; pp. 329–339. [\[CrossRef\]](#)
33. Turing, A.M., Computing machinery and intelligence. In *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer*; Epstein, R., Roberts, G., Beber, G., Eds.; Springer: Dordrecht, The Netherlands, 2009; pp. 23–65. [\[CrossRef\]](#)
34. Torres-Moreno, J.M.; Molina, A.; Sierra, G. La energía textual como medida de distancia en agrupamiento de definiciones. In Proceedings of the 10th International Conference on Statistical Analysis of Textual Data (JADT 2010), Rome, Italy, 6–11 June 2010; pp. 215–226.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.