Instituto Politécnico Nacional

Centro de Investigación en Computación

Automatic Generation of Fiction Stories Using Human-Like Writing Style

T E S I S

que para obtener el grado de:

MAESTRÍA EN CIENCIAS DE LA COMPUTACIÓN

presenta:

Ing. Brian Daniel Herrera Gonzalez

Tutores:

Dr. Francisco Hiram Calvo Castro
Dr. Alexander Gelbukh

Ciudad de México
Julio 2021

SIP-14
REP 2017

INSTITUTO POLITÉCNICO NACIONAL
SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

ACTA DE REVISIÓN DE TESIS

En la Ciudad de México siendo las 12:00 horas del día 23 del mes de junio del 2021 se reunieron los miembros de la Comisión Revisora de la Tesis, designada por el Colegio de Profesores de Posgrado del: **Centro de Investigación en Computación** para examinar la tesis titulada:

"Automatic Generation of Fiction Stories Using Human-Like Writing Style"

del (la) alumno (a):

Apellido Paterno:	HERRERA	Apellido Materno:	GONZALEZ	Nombre (s):	BRIAN DANIEL

Número de registro: B 1 9 0 4 0 1

Aspirante del Programa Académico de Posgrado: **Maestría en Ciencias de la Computación**

Una vez que se realizó un análisis de similitud de texto, utilizando el software antiplagio, se encontró que el trabajo de tesis tiene 19 % de similitud. **Se adjunta reporte de software utilizado**.

Después que esta Comisión revisó exhaustivamente el contenido, estructura, intención y ubicación de los textos de la tesis identificados como coincidentes con otros documentos, concluyó que en el presente trabajo **SI** ☐ **NO** ☒ **SE CONSTITUYE UN POSIBLE PLAGIO.**

JUSTIFICACIÓN DE LA CONCLUSIÓN: *(Por ejemplo, el % de similitud se localiza en metodologías adecuadamente referidas a fuente original)*

11% de la similitud corresponde a un artículo publicado por el alumno durante la realización de sus estudios: Herrera-González, B. D., Gelbukh, A., & Calvo, H. (2020, October). Automatic Story Generation: State of the Art and Recent Trends. In *Mexican International Conference on Artificial Intelligence* (pp. 81-91). Springer, Cham.
El resto corresponde a fragmentos menores a 1% (ver reporte adjunto)

****Es responsabilidad del alumno como autor de la tesis la verificación antiplagio, y del Director o Directores de tesis el análisis del % de similitud para establecer el riesgo o la existencia de un posible plagio.**

Finalmente, y posterior a la lectura, revisión individual, así como el análisis e intercambio de opiniones, los miembros de la Comisión manifestaron **APROBAR** ☒ **NO APROBAR** ☐ la tesis, en virtud de los motivos siguientes:
La tesis contiene trabajo de investigación original, acorde con el nivel que corresponde a Maestría. El artículo fue publicado en una revista indexada en SCOPUS.

COMISIÓN REVISORA DE TESIS

Dr. Francisco Hiram Calvo Castro
Director de tesis

Dr. Grigori Sidorov

Dra. Olga Kolesnikova

Dr. Alexander Gelbukh
2°. Director de Tesis

Dr. Ildar Batyrshin

Dr. Marco Antonio Moreno Ibarra

INSTITUTO POLITECNICO NACIONAL
CENTRO DE INVESTIGACION
EN COMPUTACION
DIRECCION

Dr. Marco Antonio Moreno Ibarra
PRESIDENTE DEL COLEGIO DE PROFESORES

INSTITUTO POLITÉCNICO NACIONAL

SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

CARTA CESIÓN DE DERECHOS

En la Ciudad de México el día **28** del mes **junio** del año **2021**, el que suscribe **Brian Daniel Herrera Gonzalez** alumno del Programa de **Maestría en Ciencias de la Computación** con número de registro **B190401**, adscrito al **Centro de Investigación en Computación**, manifiesta que es autor intelectual del presente trabajo de Tesis bajo la dirección del **Dr. Francisco Hiram Calvo Castro y Dr. Alexander Gelbukh** y cede los derechos del trabajo intitulado **"Automatic Generation of Fiction Stories Using Human-Like Writing Style"**, al Instituto Politécnico Nacional para su difusión, con fines académicos y de investigación.

Los usuarios de la información no deben reproducir el contenido textual, gráficas o datos del trabajo sin el permiso expreso del autor y/o director del trabajo. Este puede ser obtenido escribiendo a la siguiente dirección **bdherrerag@gmail.com**. Si el permiso se otorga, el usuario deberá dar el agradecimiento correspondiente y citar la fuente del mismo.

Brian

Brian Daniel Herrera Gonzalez

Resumen

A lo largo de la historia el ser humano ha utilizado los relatos como una manera de comunicar ideas y transmitir conocimiento. La gran capacidad de imaginación que posee el hombre ha logrado la creación de diferentes obras, de historias que evocan diferentes emociones y que reflejan la percepción del autor acerca de la realidad. La misma capacidad de crear realidades alternas es la que nos ayuda a percibirlas a través del autor y a poder imaginarlas como receptores de la historia. El poder de imaginar distintos escenarios para una situación en concreto ha ayudado al ser humano a evadir peligros y resolver problemas de tal manera que se genera nuevo conocimiento para interpretar y cambiar la realidad percibida. Todos los aspectos anteriores muestran la importancia que tienen las historias para el ser humano. En un principio las historias eran comunicadas de manera oral, sin embargo uno de los grandes avances en esta área fue obtenido gracias a la escritura, hecho que facilitó sustancialmente la transmisión de estas historias. Siendo una actividad tan antigua en la historia humana, no resulta extraño que haya sido objeto de estudio de la inteligencia artificial, desde una época muy temprana en el desarrollo de esta última, dando como resultado, entre otras cosas, el surgimiento de la generación automática de historias. La generación automática de historias ha sido una tarea que se ha buscado resolver utilizado diferentes enfoques, aunque en los últimos años se han obtenido muy buenos resultados utilizando la tecnología de modelos de lenguaje pre entrenados. En este trabajo se propone una metodología para la generación de historias y se evalúa la importancia del tiempo de pre entrenamiento y el uso de historias de ficción para dicho pre entrenamiento.

Abstract

Throughout history, human beings have used stories as a way of communicating ideas and transmitting knowledge. Man's great capacity for imagination has led to the creation of different tales that evoke different emotions and reflect the author's perception of reality. The same capacity to create alternate realities is what helps us to perceive them through the author and to be able to imagine them as receivers of the story. The power to imagine different scenarios for a particular scenario has helped human beings to evade dangers and solve problems in such a way that new knowledge is generated to interpret and change the perceived reality. All the above aspects show the importance of stories for human beings. In the beginning, stories were communicated orally, but one of the great advances in this area was obtained thanks to writing, which substantially facilitated the transmission of these stories. Being such an ancient activity in human history, it is not surprising that it has been the object of study of artificial intelligence, from a very early stage in the development of the latter, leading to, among other things, the emergence of automatic story generation. The automatic generation of stories has been a task that has been sought to be solved using different approaches, although in recent years very good results have been obtained using the technology of pre-trained language models. In this work, we propose a methodology for the generation of stories and evaluate the importance of pre-training time and the use of fictional stories for such pre-training.

Acknowledgements

I am extremely grateful to all the people who supported me in the development of this work.

I would especially like to thank my family for their unconditional support and my advisors, Dr. Hiram Calvo and Dr. Alexander Gelbukh, for their thoughtful comments and guidance in conducting this project.

The financial support from CONACYT (The National Council of Science and Technology), the infrastructure provided by AMPLN (The Mexican Association of Natural Language Processing) and the education received from Instituto Politécnico Nacional made the realization of this work possible.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Narration has been a theme that has accompanied human beings for many centuries, ancient evidence suggests the use of stories from a very primitive time of the human being, in fact the transmission of stories in an oral way had been, for many centuries, the predominant way to be able to spread them.

The written narrative was one of the great advances that this field obtained, since it allowed, among other things, the intercultural transmission of stories. In more recent times the development of the printing press meant another progress for the propagation of these stories and in the last century the computer and the web have allowed that stories of any origin can reach almost any corner of the earth, however, all these stories have something in common, they have been the result of man's imagination.

The importance of stories relies on the fact that everything we know is the story of someone or something. In the words of the American poet Muriel Rukeyser[7], 'The Universe is made of stories, not of atoms'.

Artificial intelligence has been used to solve different tasks of human beings, either reducing cost, time, effort, among other elements that have led to a new way of using computers, but among all this repertoire there is a question that has intrigued the scientific community: Can a computer be creative? This doubt seems to be far from having an answer that satisfies both the people that belong to this branch of science and those that do not, but one of the ways that can lead to clarify this enigma is the resolution of tasks that seem to appeal to man's creativity. One of these tasks is the automatic generation of stories.

There has been several attempts in order to generate automatic stories, following two general approaches, the symbolic approach and the conexionist approach. However, this task has not yet a clear solution or a basic methodology to be solved.

In recent years pre-trained language models are used for different tasks in the natural language processing such as text generation. The limitations and maximal capabilities of these models are not specified yet.

In this work we will research the importance of keeping pre-training the model with fictional stories, common sense phrases and the importance of the perplexity of those models.

1.1 Motivation

The automatic generation of stories is a task that has many areas and plenty of investigation to do. Both story generation techniques and artificial intelligence techniques can be applied, for the former we have, for example, different methodologies for world building, storytelling, reader engagement, character creation and development.

We can use a hybrid approach between symbolic and conexionist models in order to have the best of both. In addition, the pre-trained language models and architectures created from them are a viable option of research in order to generate better stories.

In the same way we can help those implementations with tasks such as named entity replacement, automatic ontology generation, coherence and cohesion check, among others.

In conclusion, the automatic generation of stories is an ambitious task with a large area of research, and we will work on the impact of fictional text, common sense text and the perplexity on pre-trained language models.

1.2 Hypothesis

A pre-trained language model can generate more diverse, yet interesting outputs using fictional texts, and the lesser the perplexity the stories generated have a more human-like writing style.

1.3 Objectives

The objectives of this work are as follows.

1.3.1 General objective

To generate diverse, yet interesting outputs with a human-like, writing style.

1.3.2 Particular objectives

- To build a fictional dataset.
- To create a methodology for automatic story generation.
- To train pre-trained autoregressive models with a fictional dataset.
- To compare the outputs of the models.

1.4 Novelty

Most of the recent research with pre-trained models and automatic generation of stories does not take into account neither, the importance of fiction in those stories nor the importance of a hybrid approach in the automatic generation. Our proposal has both aspects therein.

1.5 Contributions

The following contributions were obtained through this work.

- A pre-trained model with low perplexity and a fictional approach.
- A methodology for automatic generation of stories.

Chapter 2

Antecedents

Our proposal falls into the area of computational narratology, which is the intersection of the areas of artificial intelligence and narratology. The works described in this chapter are the basis of our work.

2.1 Narratology

"Narratology is a humanities discipline dedicated to the study of the logic, principles, and practices of narrative representation. Dominated by structuralist approaches at its beginning, narratology has developed into a variety of theories, concepts, and analytic procedures. Its concepts and models are widely used as heuristic tools, and narratological theorems play a central role in the exploration and modeling of our ability to produce and process narratives in a multitude of forms, media, contexts, and communicative practices". [8]

There are plenty of methodologies and techniques for creating a story, the oldest known works date back to the writings of Plato and Aristotle, *The Republic* and *Poetics*, respectively. They speak about the relations between characters and actions. Multiple methodologies have been used to define the types of narrative, the taxonomy of stories and the elements that compose a story. For our work we use the situational approach, because it is not linked to a specific literary genre, nor to a thematic classification of the stories.

2.1.1 The Thirty-Six Dramatic Situations

Les 36 situations dramatiques: Maîtriser l'art narratif grâce à l'exploration des principes dramatiques [9] is an essay written by Georges Polti in 1895. It is based on the work of the Italian Carlo Gozzi and the German Johann Wolfgang von Goethe, both of them tried to decompose a story in smaller parts. In this book, Polti proposed thirty-six situations that comprehend all the possible scenarios in the previous narrative works. Every situation has its own subclassification, the elements needed for the specific situation and a brief description. The description includes the emotions evoked and its link to other situations.

The thirty-six dramatic situations are as follows:

1. Supplication
2. Deliverance
3. Crime pursued by vengeance
4. Vengeance taken for kindred upon kindred
5. Pursuit
6. Disaster
7. Falling prey to cruelty/misfortune
8. Revolt
9. Daring enterprise
10. Abduction
11. The enigma
12. Obtaining
13. Enmity of kindred
14. Rivalry of kindred
15. Murderous adultery
16. Madness

17. Fatal imprudence
18. Involuntary crimes of love
19. Slaying of kindred unrecognized
20. Self-sacrifice for an ideal
21. Self-sacrifice for kindred
22. All sacrificed for passion
23. Necessity of sacrificing loved ones
24. Rivalry of superior vs. inferior
25. Adultery
26. Crimes of love
27. Discovery of the dishonour of a loved one
28. Obstacles to love
29. An enemy loved
30. Ambition
31. Conflict with a god
32. Mistaken jealousy
33. Erroneous judgment
34. Remorse
35. Recovery of a lost one
36. Loss of loved ones

Taking the Fifth situation as an example, the title is Pursuit, it is the converse of the First and represents an outcome of the Third and Fourth. The Fifth situation aims to make the reader accomplice in even the worst of scenarios. The specific situations and their examples, taken form the original source [9] are as follows:

- "A - Fugitives From Justice Pursued For Brigandage, Political Offenses, Etc.:- "Louis Perez of Galicia" and "Devotion to the Cross," both by Calderon; the beginning of the mediaeval Miracle "Robert-le-Diable;" "The Brigands" by Schiller; "Raffles" (Hornung,1907). Historical examples: the proscription of the Conventionnels; the Duchesse de Berry. Examples from fiction: "Rocambole" by Gaboriau; "Arsene Lupin" Leblanc). Familiar instances: police news. Example in comedy: "Compere le Renard" Polti, L905"

- "B - Pursued For a Fault of love: Unjustly. "Indigne!" Barbier, L884); more justly, Moliere's "Don Juan" and Comeille's "Festin de Pierre," (not to speak of various works of Tirso de Molina, Tellez, Villiers, Sadwell, Zamora, Goldoni, Grabbe, Zorilla, Dumas père); very justly, "Ajax of Locris," by Sophocles. Familiar instances run all the way from the forced marriage of seducers to arrests for sidewalk flirtations."

- "C - Hero Struggling Against a Power: Aeschy- "Prometheus Bound;" Sophoces "Laocoon;" the role of Porus in Racine's and also in Metastasio's "Alexandre;" Corneille's "Nicomede;" Goethe's "Goetz von Berlichingen" and a part of "Egmont;" Metastasio's "Cato;" Manzoni's "Adelghis" and a part of his "Count of Carmagnola;" the death of Hector in Shakespeare's "Troilus and Cressida;" "Nana-Sahib" (Richepin, 1883); "Edith" (Bois, 1885); the tetralogy of the "Nibelungen;" "An Enemy of the People" (Ibsen); "Le Roi sans Couronne" (de Bouhtfier, 1909)."

- "D - A Pseudo-Madman Struggling Against an Iago-Like Alienist: - "La Vicomtesse Alice" (Second 1882)."

There is also a more recent book [10] written by the British film-director Mike Figgis. Figgis changed some of the situations in order to update the book of Polti, with changes as gender equality and considerations for film scripts.

We use both books for our proposal described in chapter 4.

2.2 Computational Narratology

"Computational narratology is the study of narrative from the point of view of computation and information processing. It focuses on the algorithmic processes involved in creating and interpreting narratives, modeling narrative structure in terms of formal, computable representations. Its scope includes the approaches to storytelling in artificial intelligence systems and computer (and video) games, the automatic interpretation and generation of stories, and the exploration and testing of literary hypotheses through mining of narrative structure from corpora". [11]

As we will see most of the work of the symbolic research for automatic story generation falls into this area, and also the conexionist works with an explicit outline are part of the computational narratology.

2.3 Symbolic Artificial Intelligence

2.3.1 Typed Feature Structures

There is a joint between symbolic artificial intelligence and situations. This bond is the Minsky's framework along with typed feature structures, and we also use this approach in our proposal. This framework is a semantic approach and a way of representing knowledge. Using this framework we can build events and create rules in order to bond the objects in the frames.

We use a Minsky's fragment of text [12] to clarify how to use his framework.

"There was once a Wolf who saw a Lamb drinking at a river and wanted an excuse to eat it. For that purpose, even though he himself was upstream, he accused the Lamb of stirring up the water and keeping him from drinking..."

Minsky exposes that in order to understand the text we need to understand the following situations.

1. The wolf is lying.
2. The contamination never flows upstream.
3. The "upstream" word itself.
4. What is "stirring up" and why would it keep the wolf from drinking?
5. Stirring river-water means that the first frame should have "mud" assigned to it or is related by default to stirred water?

As we can deduce the common sense knowledge needed to understand the story is huge, and it seems a bit far from the actual state of the art in artificial intelligence. To be able to manage this information we use typed feature structures or TFE to represent the situations, as shown in [13].

We need to specify the following.

- Characters
- Places
- Objects
- Actions
- Wh questions
- Its place in the storyline

For some questions, such as what and why, we can have either a thing or another situation as their values. With all of the above we can represent the situation as an attribute-value matrix or AVM.

The symbolic approach was the main way of addressing the task of computational narratology, in recent years the conexionist approach is used for automatic story generation but the focus of those works is still far from the computational narratology, we use a hybrid approach for our proposal.

2.3.2 Previous works

In the first works carried out to perform the activity of automatic generation of stories, the symbolic guideline was used. One of the main strengths of this approach is the ease with which it is possible to delimit the path followed by the actions of the story. However, one of its disadvantages is the lack of “creativity”, in other words, stories tend to be very repetitive and have a very reduced repertoire of events. These works can be divided into two main areas, those case-based and those event-based

2.3.2.1 Case-based methods

Case-based methods are so called because they are strongly restricted to solving a theme, either the stories domain, the author goals or both.

In 1977 Meehan [14] with TALE-SPIN, generates different characters and their respective goals. The system must look for an outcome for the resolution of these goals using an inference engine based on common sense reasoning theories, all the stories are based on King Arthur's stories.

Then in 2010, Riedl et al. use the IPOCL methodology [15]. In this work, they write fables from POCL (Partial Order Causal Link) algorithms using operators, preconditions and effects of the operators.

In 2014, McCoy et al. [16] generate prom's stories using knowledge bases about the social structure, its norms, the cultural aspect and the desires of the characters, as well as concepts of social interaction.

Finally in 2016, Daza et al. [17] use literary structures for writing a story and evaluate the stories generated.

In 1987, Lebowitz [18] with UNIVERSE, uses a hierarchical planner to turn the author's goals into a story. This is achieved with hierarchically structured rules to find related tasks that can achieve that goal.

In 2009 Porteous et al. [19] use reference points obtained from the author for finding different events that lead to a goal established by the human agent.

2.3.2.2 Event-based methods

The event-based methods use a concept called events to construct the storyline. The events are the actions or the transitions form one state to another. In 2012, Onodera et al. [20] use different states generated in a virtual world from a storyline and values of characters, objects and places. Then the user selects the states they want in the story, the system transforms them into events and finally uses a circular generation process to transform these events into sentences.

In 2014, Gervás et al. [21] use states of the world and simulate the interactions between the world and the different characters, choosing the simulation that best corresponds to the established narrative characteristics.

In addition, Adolfo et al. in [22] propose a model with events that emphasizes the development of characters through the use of events. The objective of this model is to develop stories for children.

Finally in 2019 Farrell et al. Farrell2019 propose a system based on Indexter Indexter to obtain the relevance of the events in different stories and to direct the story based on the relevance of those events.

We want to take some aspects from the symbolic approach works, such as the utilization of a narrative methodology and the focus on the coherence, cohesion and level of emotion of the stories.

2.4 Conexionist Artificial Intelligence

The lack of novelty in the symbolic methods has been covered by the conexionist approach, however, the lack of coherence and cohesion was a great obstacle with these methods.

2.4.1 Previous works

Some of the proposals for generating text have been sequence-by-sequence models (seq2seq) with recurrent neural networks (RNN), in particular gated recurrent networks and long short term memories GRU and LSTM respectively. Another proposal is the unsupervised learning by means of the generative adversarial networks GAN, in the literature we have also found pre-trained language models. However, all the alternatives that use only this type of approach have failed in the task of writing a history that maintains coherence.

2.4.1.1 Methods for stories without an explicit outline

These methods do not have a delimited section for generating the storyline. They can be classified as those that write a story without an specific focus and those who have specific requirements.

Jain et al. [23] in 2017 start from small descriptions and use statistical machine translation and sequence-to-sequence networks to obtain a story. Their results contain very little cohesion, however it has served as a basis for the use of hybrid systems.

Furthermore, Clark et al. [24] in 2018 use a generation model based on entities. This model takes into account the general content of these entities, the content of the previous sentence and the already written content of the same sentence and generates the next word. Then encodes these three elements and uses a recurrent neural network in order to generate a sentence. In this research a significant advance is achieved, however, human judges perceive a lack of fluency in the story.

Fan et al. [25] in 2018 also separate the problem into two parts: The generation of a premise and the generation of a story from that premise. The story base was obtained through instructions provided by the researchers to different writers. To obtain the premise they use a seq2seq encoder-decoder convolutional model. In particular they use a non-context model with two attention elements, multiscale and closed. Finally they use a fusion model to retrieve information in the hidden layers. The premise becomes a story using a top-k random sampling scheme on the models. With this approach, the stories have significant coherence but little cohesion.

Peng et al. [26] in 2018 focus on achieving controllability of automatic story generation. They are based on the closing valence and the story line. To obtain the closure validity they use a logistic regression classifier based on a LSTM and for the generation they use a conditional language model in the same way with LSTM. For the argument line they obtain the keywords and they obtain a graph of the document with weights for each keyword. The generation is also made with a LSTM with attention. An effective control of the end of an unfinished story is achieved. However the generation of a story from some words is not yet a very coherent story.

In 2019, Yao et al. [27] proposed generating a story line from a title and then turning it into a story. For the creation of the static storyline and the generation of the text they use a LSTM. The stories have a good performance in coherence.

This time in 2019 Fan et al. [28] propose a semantic role labeling to improve their proposal in [25], obtaining much better results than in their previous work. The semantic role labeling achieves that two sentences containing two different structures and the same meaning can be interpreted as the same entity.

Some methods try to solve specific problems, for example Roemmele et al. [29] in 2018 show a "creative" assistant to continue a story with a sentence that is the result of a model with a recurrent neural network and that uses a variable that adjusts the probability distribution of the model, showing that the most creative proposals were also the least useful for the human author.

Ding et al. [30] in 2018 use 3 components, the first is a seq2seq generator with a LSTM, which given a context, generates an ending, the second component is a binary discriminator that receives both the context and the ending and classifies it as a human-made or machine-generated ending, the third component is an adversarial training process between the two previous components. In this work "good" endings are obtained according to human evaluation.

Additionnally, Luo et al. [31] in 2019 use two components for their system, the first one is a sentiment analyser that adopts three methods, an unsupervised rule-based method, a linear regression model and an adversarial learning model. For the second component they use a seq2seq model controlled by the intensity of the feeling. They manage to obtain a good result according to the intensity of the feeling sought, but like the previous papers, some sentences lack coherence.

We can also observe the work in Guan et al. [32] in 2019 who use two components, the first one is an incremental encoder with a LSTM to obtain the clues that are used to reach a conclusion. The second element is a multi-source attention mechanism that is used to obtain the context of a common sense knowledge base. With this work an increase in the fluency of the story is obtained, but it does not improve the coherence of the story.

2.4.1.2 Methods for stories with an explicit outline

These methods show some kind of planning for the generation of the storyline. In 2018, Martin et al. [33] separate the task into two parts. The first in the generation of the series of events that make up the story and the second in the narrative that describes these events. The events are separated into tuples of subject, verb, object, a wild card, and literary genre; to find the next event they use a multi-layered recurrent code-decoder network. An LSTM network with Beam Search is used to transform events into sentences. They perform favourably in obtaining new events, however the generation of sentences is deficient as it lacks of a coherent structure.

Furthermore, Xu et al. [34] in 2018 use a scheme that they call skeleton and then convert it into sentences. The system consists of two modules, the first one obtains the skeletons from the database using a seq2seq model with encoder and decoder based on LSTM. The second module contains two sub-modules, which are input to skeleton and skeleton to sentence. To obtain the input to skeleton a seq2seq structure is used, where the encoder is hierarchical and the decoder is attention based, the submodule in charge of transforming the skeleton into a sentence uses a seq2seq model with encoder and decoder based on a single layer LSTM with attention mechanism. The two modules are linked by a reinforcement learning algorithm. The results shown have a good coherence and a better fluency.

Tambwekar et al. [35] in 2019 propose to use verbs as history objectives. The reward is obtained by multiplying two parameters: the distance, that is, the number of verbs between the candidate verb and the target verb, and the frequency of the candidate verb before the target verb. To avoid a short story they use a cluster of verbs. The output of the system are events in tuples of subject, verb, object and wildcard. That paper presents a little used proposal, however, the absence of an event to sentence component limits the correct interpretation of the results.

Additionnally, Ammanabrolu et al. [36] in 2019 use 5 different methods to transform events into sentences, using as a basis the events obtained with [31]. The first method is based on Hashimoto RetEdit [37] in 2018, the second method is a template filling, using a simplified grammar and a LSTM, the third method is a Monte Carlo beam search applied to a seq2seq model, the fourth method is a restricted beam search with finite state machines as a guide. The result is an assembly of the five proposals based on the highest confidence obtained by the five methods. In that paper, a better performance for the translation of events into sentences is observed, although there is still a lack of development to obtain a coherent text.

All of the conexionist works above have a lower performance than the recent works that use the transformer architecture, as we will see in chapter 3.

2.4.2 Transformer architecture

In 2017 Vaswani et al. published the paper "Attention is all you need" [38], in this paper they propose the transformer architecture as shown in Figure 2.1 [38], we can see the transformer architecture.

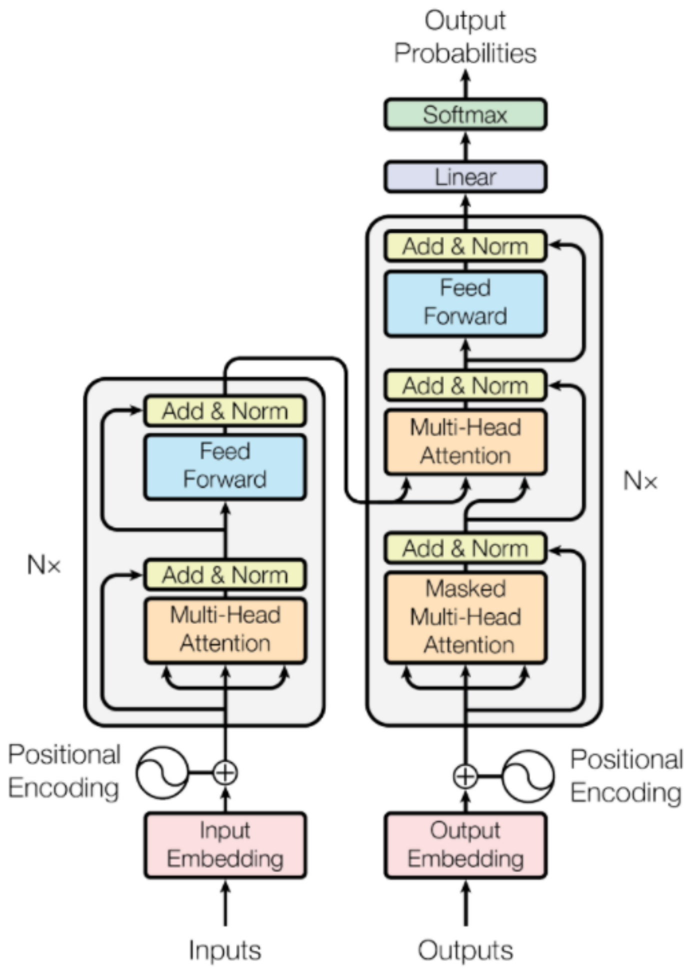

Figure 2.1: Transformer architecture

Transformers networks have characteristics such as the following.

- They can handle long-range dependencies.
- They do not have problems with gradient.
- They require fewer steps to train.
- Since there is no recurrence, parallel computation can be used.
- We can analyse more data.

Transformer architecture is based on an attention mechanism. The attention mechanism seeks to return the value V_i, for a query q, based on a key k_i in some database. As we can see in Equation 2.1.

$$attention(qk, v) = \sum_{i}^{n} similarity(q, k_i) x V_i \tag{2.1}$$

2.4.2.1 Multi-Head Attention

As we can see, one of the most representative elements of the architecture is its attention stage. In Figure 2.2 [38], values of v, q and k are used for a number h of heads. We can also see the equations for multi-head attention in Equation 2.2 [38].

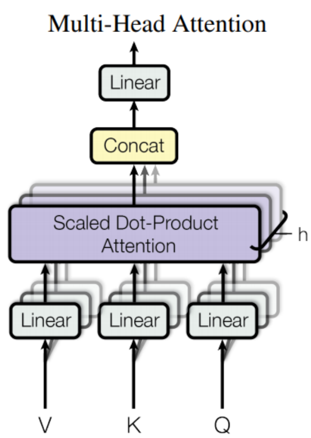

Figure 2.2: Multi-head attention

$$\begin{aligned} \mathrm{MultiHead}(Q, K, V) &= \mathrm{Concat}(\mathrm{head_1}, ..., \mathrm{head_h})W^O \\ \text{where } \mathrm{head_i} &= \mathrm{Attention}(QW_i^Q, KW_i^K, VW_i^V) \end{aligned} \tag{2.2}$$

2.4.2.2 Normalization by layer

The values in each layer are normalised with the intention of having an average of 0 and a variance of 1, thereby reducing the number of iterations for training.

2.4.2.3 Positional Embeddings

In order to keep the information about the position of the words, different types of encoding can be used, but the standard encoding is shown in Figure 2.3 [38].. As we can see it depends on the sine and cosine functions, but as already mentioned, other functions can be used.

$$PE_{(pos,2i)} = sin(pos/10000^{2i/d_{\text{model}}})$$
$$PE_{(pos,2i+1)} = cos(pos/10000^{2i/d_{\text{model}}})$$

Figure 2.3: Positional Embeddings

2.4.3 Architectures based on Transformers

Diverse architectures are based on transformers, and almost every couple of months there is a newer architecture. In this subsection we will describe some of the architectures used as baselines for different approaches.

2.4.3.1 Architectures for Pre-trained Language Models

Pre-trained language models have shown a great impact in plenty of natural language processing tasks, the relevance of these models is the capability to improve the performance of tasks for which they were not specifically trained. You can even train them in other languages and then fine-tuning the models for a specific task. A few disadvantages are the computational power required, the big amount of time for training and the biases learnt from the data. Another disadvantage is that the input block size should be short, because of the quadratic nature of the attention matrix. A great advantage is that it only should be trained once and then the pre-trained language model can be fine-tuned for any task.

The Bidirectional Encoder Representations from Transformers [39] or BERT was developed by Google in 2018, it has 340 million parameters. This architecture takes into account the context, both left and right sides for training. Its training is divided into two parts, in the first part it presents a sentence with a gap and the system must predict what the missing word is. In the second part it presents two sentences and the system must tell if it is true that the two sentences go together.

We also have the Generative Pre-trained Transformer or GPT family, it was in 2019 and with GPT-2 [40] that OpenAI researchers managed to get attention with their generation model, which, in the words of its developers, was "too dangerous to release". The second version has 1.5 billion parameters, some interesting facts about this Transformer are that the language model obtained only depends on the context on the left and that it has a Byte-Pair encoding.

It may be because of the statements of version two that version three made a big impact on the artificial intelligence community. Version GPT-3 [41]was released in mid-2020 and proved to be a media phenomenon due to the apparent intelligence it demonstrated with some tasks for which it had not been particularly trained on, another point that was notable in this proposal was its 175 billion parameters.

RoBERTa [42] or Robust Bidirectional Encoder Representations from Transformers is a proposal to improve BERT [39], it was developed by Facebook and gets its name due to the robust optimisation, a few of the contributions are the dynamic masking, a report about the length of the training blocks that is more convenient and a Larger Byte encoding. According to RoBERTa less perplexity means better performance in the tasks after fine-tuning the model for specific tasks.

2.4.3.2 Architectures for reducing the Attention Matrix dimensionality

One of the current problems with this architecture is the short length of the input sequence. In most previous Transformers the length is between 512 and 1024 characters.In previous architectures, increasing the size of the sequence increases the computation time in a quadratic way. For this reason, several researchers have developed different techniques to solve this problem.

The Reformer [43] model is one of many attempts to increase the sequence size without quadratically increasing the computation time. Its idea is to separate the data into blocks called chunks, and based on that it trains between chunks. We can see the division in Figure 2.4 [43].

Another model that seeks to use longer sentences is Longformer [44].This model is based on convolution windows as shown in Figure 2.5 [44]. As we can see it does not have full attention, instead it presents attention in the sliding window, attention in key places and global attention.

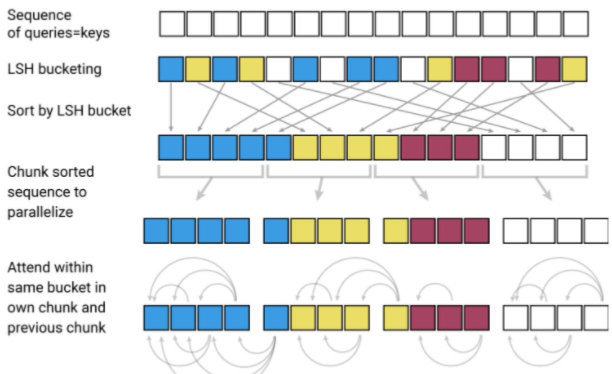

Figure 2.4: Reformer Chunks

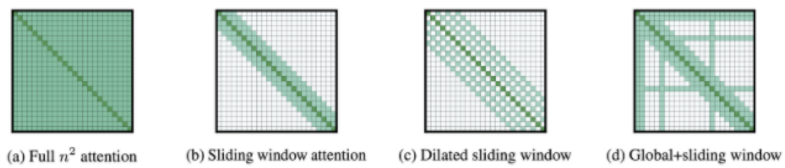

Figure 2.5: Longformer Attention Windows

Another architecture is Linformer [45] which uses the same principle as Reformer with a block attention, as can be seen in Figure 2.6 [45].

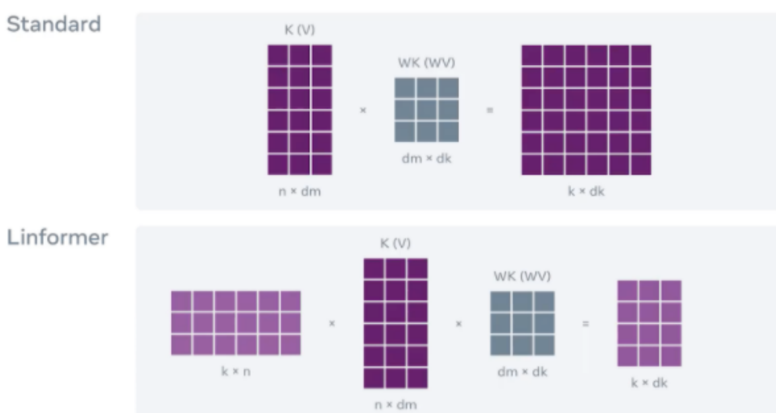

Figure 2.6: Linformer Blocks

However, despite its name, it is non-linear, its computation time is n log(n).

One more model that seeks to read large sequences of data is Big Bird [46] which uses random attention, global attention and windowed attention. However, in order to be useful it has shown that it needs more layers or relies heavily on random attention which adds more training time. We can see Big Bird's approach to attention in Figure 2.7 [46].

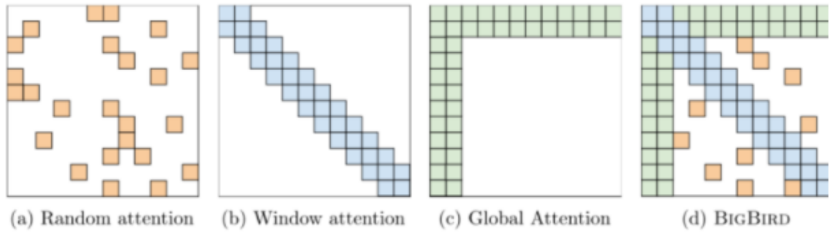

Figure 2.7: Big Bird Attention

The following model, called Performer [47], uses several approaches and manages to obtain a transformation that significantly reduces the computational time. The model uses Kernels to achieve a change of space to a larger space and achieve a matrix multiplication in such a way that, theoretically, it becomes linear. The transformation used by the Performer model can be seen in Figure 2.8 [47].

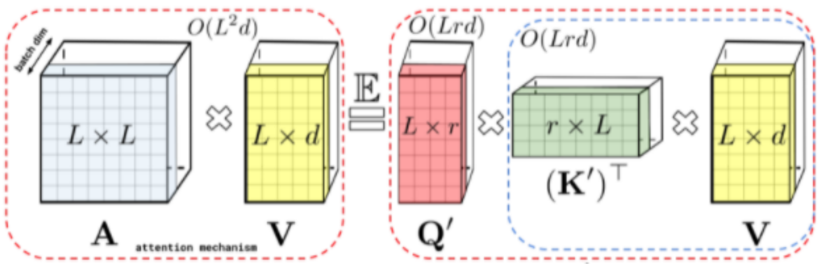

Figure 2.8: Performer Transformation

2.4.3.3 Distillation for Pre-trained Language Models

Among the techniques for reducing the size of the pre-trained model, distillation stands out. One of the most widely used of these is the DistilBERT [48] model, which has an effectiveness in many tasks of 97 percent of the original BERT [39] and at the same time is 40 percent smaller and 60 percent faster. It uses knowledge distillation, which is based on a teacher-student technique and requires the model to already exist in order to distil it.

Chapter 3

State of the Art

3.1 Works with a symbolic case-based approach

Even though the case-based works have been published decades ago, their performance is better in the majority of the aspects pursued by the computational narratology, and one of the advantages of case-based methods is that we have the subjective evaluation of two of the methods.

3.1.1 MINSTREL

In 1993, Turner with MINSTREL[1], generated new stories using existing story concepts as a basis. The system must adapt these concepts through a model of computational creativity that satisfies the objectives of the author and the story. The basis stories and the stories that MINSTREL can generate are short stories about King Arthur and his Knights of the round table.

The stories are generated as a problem solving task by the author. Therefore, it can analyse all the specific constraints for satisfying the author's goals. There are four kinds of goals [49]:

- Thematic goals
- Drama goals

- Consistency goals
- Presentation goals

The thematic goals are considered as a story fragment. The drama goals considered in MINSTREL are based on four techniques implemented by human authors, the techniques are suspense, tragedy, foreshadowing and characterization. The consistency goals are satisfied when the narrative reflects general understandings of the way the world works. The presentation goals are referred as the way and order that the story elements are presented to the reader.

According to [50] "In MINSTREL, all the elements that conform a story are represented as schemas... schemas are divided in two classes: 1) those employed by the system to satisfy rhetoric constraints, referred to as author-schemas; 2) those employed by the system to represent events in a story, referred to as character-schemas. Examples of author-schemas are author-level goals, e.g. the goal of including suspense in a story. MINSTREL has 21 author-level goals divided in four main groups... Author-level goals have associated instructions to achieve their goals. These instructions are independent blocks of Lisp code, and are referred to as Author-Level Plan (ALP). MINSTREL counts with 34 ALPs which explicitly indicates how to create scenes to include revenge, deception, beliefs, how to create the introduction of a story, its denouement, to check its consistency, and so on. Examples of character-schemas are character-level goals (MINSTREL employs 13 of these goals, like satisfying one's hunger, changing location, causing fear in someone, finding a romantic love, etcetera), representations of humans, monsters, animals, physical objects, beliefs, emotions, actions, states (a state is a representation of a fact that is true, e.g. "Lancelot is in the city"), etcetera. Character-schemas can be linked to establish relationships between them. In this way, it is possible to construct elaborated scenes. MINSTREL develops stories about six predefined schemas-themes known as Planning Advice Themes (PATs)... MINSTREL performs two main processes: the planning process, which controls the author-level goals, and the problem-solving process, which focuses in achieving these goals."

Another interesting feature from MINSTREL is the Transform Recall Adapt Methods or TRAMS proposal. The TRAMS are a series of heuristics used to generate new content, the program has a restriction of not presenting events that were already set in the episodic memory. In other words is the creative alternative for this symbolic approach.

"The Vengeful Princess" is a story generated by MINSTREL [1] and is shown in Figure 3.1.

"The Vengeful Princess":

Once upon a time there was a Lady of the Court named Jennifer. Jennifer loved a knight named Grunfeld. Grunfeld loved Jennifer.

Jennifer wanted revenge on a lady of the court named Darlene because she had the berries which she picked in the woods and Jennifer wanted to have the berries. Jennifer wanted to scare Darlene. Jennifer wanted a dragon to move towards Darlene so that Darlene believed it would eat her. Jennifer wanted to appear to be a dragon so that a dragon would move towards Darlene. Jennifer drank a magic potion. Jennifer transformed into a dragon. A dragon moved towards Darlene. A dragon was near Darlene.

Grunfeld wanted to impress the king. Grunfeld wanted to move towards the woods so that he could fight a dragon. Grunfeld moved towards the woods. Grunfeld was near the woods. Grunfeld fought a dragon. The dragon died. The dragon was Jennifer. Jennifer wanted to live. Jennifer tried to drink a magic potion but failed. Grunfeld was filled with grief.

Jennifer was buried in the woods. Grunfeld became a hermit.

Figure 3.1: The Vengeful Princess

3.1.2 MEXICA

In 2001, Pérez y Pérez et al. with MEXICA [2] produced stories about the Mexicas, the old inhabitants of the place today known as Mexico City. MEXICA uses a cognitive method of engagement-reflection, in order to use already known narratives and generate a new story that maintains coherence and an increasing tension throughout the story.

"During the engagement-mode the system produces material driven by content and rhetorical constraints avoiding the use of explicit goal-states or story-structure information. During the reflection-mode the system breaks impasses generated during engagement, satisfies coherence requirements, and evaluates the novelty and interestingness of the story in progress. If the results of the evaluation are not satisfactory, MEXICA can modify the constraints that drive the production of material during engagement. In this way, the stories produced by the program are the result of the interaction between engagement and reflection" [50].

MEXICA needs a memory of stories, and it build the knowledge with the following steps:

- The user defines a set of story-actions.

- The user defines a set of previous stories.
- MEXICA builds in memory content and rhetoric knowledge through information obtained from the previous stories.

The structure used for representing this knowledge is named as story-world context or SWC.

"In the MEXICA system, it is assumed that a coherent sequence of actions can be produced by linking events through the story-world context surrounding them, avoiding in this way the use of explicit goal states or pre-defined story structures... In MEXICA, consequences of actions produce emotional links between characters, modify the dramatic tension in the story, or produces changes of location for the characters." [50].

The engagement cycle [2] is formed by the following steps:

1. An action is performed by a character (the first action in the story is given by the user).
2. The consequences of this action modify characters' SWC.
3. MEXICA employs the SWCs as cue to probe memory.
4. When an SWC matches a structure in memory, the system retrieves the set of possible next actions associated to it.
5. Routines, known as Filters, eliminate all those possible next actions that do not satisfy a group of constraints known as guidelines
6. One of the retrieved actions is selected at random as the next action in the story.
7. Such an action is performed by a character, modifying the SWCs, and the cycle starts again in step 3.

During reflection [2] MEXICA performs three main processes:

1. Breaks impasses.
2. Verifies the coherence of the story in progress.

3. Evaluates the novelty and interest of the story in progress.

This is the methodology of engagement-reflection that uses MEXICA for generating stories. An example of a story [51] generated by MEXICA is shown in Figure 3.2.

> Jaguar Knight was an inhabitant of the great Tenochtitlan. Princess was an inhabitant of the great Tenochtitlan. A bad spirit took Jaguar Knight's soul provoking Jaguar Knight to become intensely jealous of Princess. Jaguar Knight tried to scare Princess by pretending that Jaguar Knight wanted to kill Princess with a lance. But instead, Jaguar Knight stumbled and wounded itself. Princess, knowing that Jaguar Knight's life was at risk, did not try to cure Jaguar Knight. In this way, Princess expected Jaguar Knight's death. Princess's state of mind was very volatile and without thinking about it Princess charged against Jaguar Knight. Suddenly, Jaguar Knight and Princess were involved in a violent fight. Princess threw some dust in Jaguar Knight's face. Then, using a dagger, Princess perforated Jaguar Knight's chest. Imitating the sacred ceremony of the sacrifice, Princess took Jaguar Knight's heart with one hand and raised it towards the sun as a sign of respect to the gods. Princess got so depressed that committed suicide.

Figure 3.2: Story generated by MEXICA

3.1.3 Comparison of MINSTREL and MEXICA

"A story produced by MINSTREL was evaluated by means of an Internet questionnaire. Nine subjects responded to it. They did not know that the story they were evaluating was written by a computer program. The subjects were asked to answer questions related to the age, education and sex of the hypothetical author of the story, as well as questions regarding the quality of the story. The following lines show the mean scores for some of the answers regarding the author: author-age 15.8; author-education: 0.9 (0=grade school, 1= high school, 2=college, 3=Graduate school); sex: 0.4 (0=female, 1=male). The following lines show the results obtained for the story (in a scale from 1 to 5 where 5 indicates best): Overall rating of story: 1.5; Clever plot: 2.5; Attention to details: 2.8; Coherency: 3.6; Use of language: 2.1" [50].

"Pérez y Pérez designed a questionnaire where 50 subjects were asked to compare different aspects of computer-made stories, such as narrative flow and coherence, structure, content and suspense" [50].

"In a different questionnaire, where seven computer-created stories were evaluated by fifty persons from twelve different countries" [50], as we can see in Figure 3.3.

	ER2	Human	MINSTREL	ER2-boring	E2	E2-boring	GESTER
Narrative flow	3.8	3.5	2.9	3.5	2.2	2.8	2.1
Narrative structure	3.7	3.7	3.2	3.2	2.6	2.7	2.1
Content	4.1	3.7	3.6	2.8	2.8	2.4	2.6
Suspense	3.8	3.8	3.3	2.3	2.3	2.0	2.1
Overall quality	3.8	3.6	3.3	2.9	2.6	2.5	2.4

Figure 3.3: Results of evaluation on a 5 point scale from 'very poor' to 'very good'.

ER2 is the engagement-reflection model. ER2-boring is the engagement-reflection model ignoring the tension. E2 is only the engagement model, in other words, without the reflection part, E2-boring is the engagement model without tension. GESTER [52] is used as a baseline because of its rigidity.

3.2 Works using Transformers

Most of the research with the conexionist approach, previous to transformers, lack of coherence, readability and cohesion. These three aspects are greatly improved by the use of transformers architectures, while maintaining a good performance in novelty. There are also works whose objective is the emotion flow. In other words, the transformer utilization has brought relevant improvements to the automatic generation of stories.

3.2.1 KEPMCSG

In the paper "A Knowledge-Enhanced Pretraining Model for Commonsense Story Generation" [3] Guan et al. use a system that generates a pre-trained model from two common sense knowledge bases, ConceptNet and ATOMIC. First, they transform the ontological base into simple sentences and with this information, they pre-train the GPT-2 model [40]. After obtaining the knowledge-enhanced pre-trained model they fine tune it with ROC Stories using multi-task learning.

With this work they obtain decent results in terms of fluency and improve the state of the art corresponding coherence. We can see an example of stories generated in Table 3.1. They also compare their work with a few of the works seen in chapter 2. They offer both automatic and human evaluations. Their results are shown in Table 3.2 and Table 3.3, respectively.

Context: The first time I saw the results of an **accident**, I was in a bus.
Golden Continuation by Human: We were stopped and a friend pointed out the accident on the road. A man was under a car after having been run over. People tried to move the car off him but only made it worse. After seeing that, I am much more careful when crossing the street. **pretrained GPT-2:** *She was trying to get home but the other passengers would not let her go.* He thought she was going to die because of her weight. *She was trying to get home but the other passengers would not let her go.* The bus driver *would not let her go.* **Fine-tuned GPT-2:** *I was on my way to a party. I was on my way to a party* and I 'd gotten out of my seat, and started driving. I got a flat tire, so I stopped driving. I drove to the party and had a great time.

Table 3.1: Examples of stories generated in [3].

Models	PPL	BLEU-1	BLEU-2	Coverage	Repetition-4(%)	Distinct-4(%)
ConvS2S	N/A	0.312	0.132	13.64	22.87	72.78
Fusion	N/A	0.322	0.137	12.02	24.23	72.82
Plan&Write	N/A	0.308	0.126	13.38	17.06	67.20
SKRL	N/A	0.267	0.088	10.82	18.34	69.42
DSRL	N/A	0.293	0.117	10.38	**15.36**	73.08
GPT-2 (Scratch)	11.82	0.311	0.134	10.76	22.87	73.33
GPT-2 (Pretrain)	33.50	0.257	0.085	8.04	39.22	64.99
GPT-2 (Fine-tune)	7.96	0.322	0.141	12.40	29.41	73.85
KEPMCSG	7.85	**0.326**	**0.143**	**18.48**	21.93	78.96
w/o Pretrain	11.04	0.316	0.134	16.33	21.52	77.17
w/o Knowledge	**7.70**	0.314	0.136	13.95	25.08	73.24
w/o Multi-task	8.04	0.324	0.140	17.19	24.40	**79.43**
Golden Story	*N/A*	*N/A*	*N/A*	*19.28*	*7.64*	*89.51*

Table 3.2: Automatic evaluation in [3].

Models	Grammaticality				Logicality			
	Win (%)	Lose (%)	Tie (%)	κ	Win (%)	Lose (%)	Tie (%)	κ
vs. Fusion	50.0**	27.0	23.0	0.421	57.0**	28.0	15.0	0.455
vs. DSRL	58.0**	24.0	18.0	0.441	58.0**	29.0	12.0	0.475
vs. GPT-2 (Scratch)	54.0**	24.5	21.5	0.385	54.0**	26.0	20.0	0.304
vs. GPT-2 (Pretrain)	52.0**	31.5	16.5	0.483	56.5**	32.5	11.0	0.493
vs. GPT-2 (Fine-tune)	42.0**	28.0	30.0	0.344	51.0**	27.5	21.5	0.371
vs. KEPMCSG w/o Pretrain	51.0**	31.0	18.0	0.378	56.0**	28.0	16.0	0.375
vs. KEPMCSG w/o Knowledge	46.0**	23.0	21.0	0.289	48.0**	29.0	23.0	0.314
vs. KEPMCSG w/o Multi-task	37.5	31.0	31.5	0.313	48.5**	25.5	26.0	0.297

Table 3.3: Human evaluation in [3].

Context	[FEMALE] was on a long road trip .
Guan et al. (2020)	she had just driven into town . she had been driving for hours . she was so tired she decided to take a nap . when she woke up , she still had n't slept well .
MEGATRON-CNTRL-124M keyword: driving	she was on her way to see a strange country. she had never seen anything like that before. finally she saw a small town that was so small. it was a great little town with a nice park !
MEGATRON-CNTRL-8B keyword: driving	she was driving in the middle of [PLACE]. all of a sudden the tire pressure light came on. she got out to check her tire. it was flat so she used the roadside assistance.
MEGATRON-CNTRL-8B-ANT keyword: attract	she really wanted to see a few attractions. the first one she saw was a giant water park. it was amazing. it ended up being a fun experience.

Table 3.4: Examples of stories generated in [4].

Name	PPL ↓	Repeat ↓	Distinct ↑	Consistency ↑ (Human Eval)
GPT-2-124M	6.98	27.2	74.1	69.5
Yao et al. (2018)	NA	**13.3**	63.7	49.0
Guan et al. (2020)	7.04	22.1	77.1	67.0
M-CNTRL-124M	9.37	20.0	80.1	74.5
M-CNTRL-355M	8.02	19.9	81.6	75.5
M-CNTRL-774M	6.58	21.3	81.6	80.5
M-CNTRL-2B	6.31	21.2	82.6	89.0
M-CNTRL-8B	**6.21**	21.2	**82.8**	**93.0**

Table 3.5: Automatic evaluation and consistency in [4].

3.2.2 MEGATRON-CNTRL-8b

Some of the best stories generated were generated by the MEGATRON-CNTRL-8b [4] system. We can see an example of stories generated by the system in Table 3.4 and the architecture of this system is shown in Figure 3.4. 124M and 8B refer to the size of their conditional generator. ANT refers to a model trained with antonyms [4].

They use MEGATRON-LM [53] for all the pre-trained language models to initialize their contextual knowledge ranker and generative models. We can observe their performance in Table 3.5. They also have a human evaluation for determining which story has more coherence and fluency, as shown in Table 3.6. For this project Xu et al. used an architecture for a total of 160 Tesla V100 32GB GPUs at NVIDIA labs.

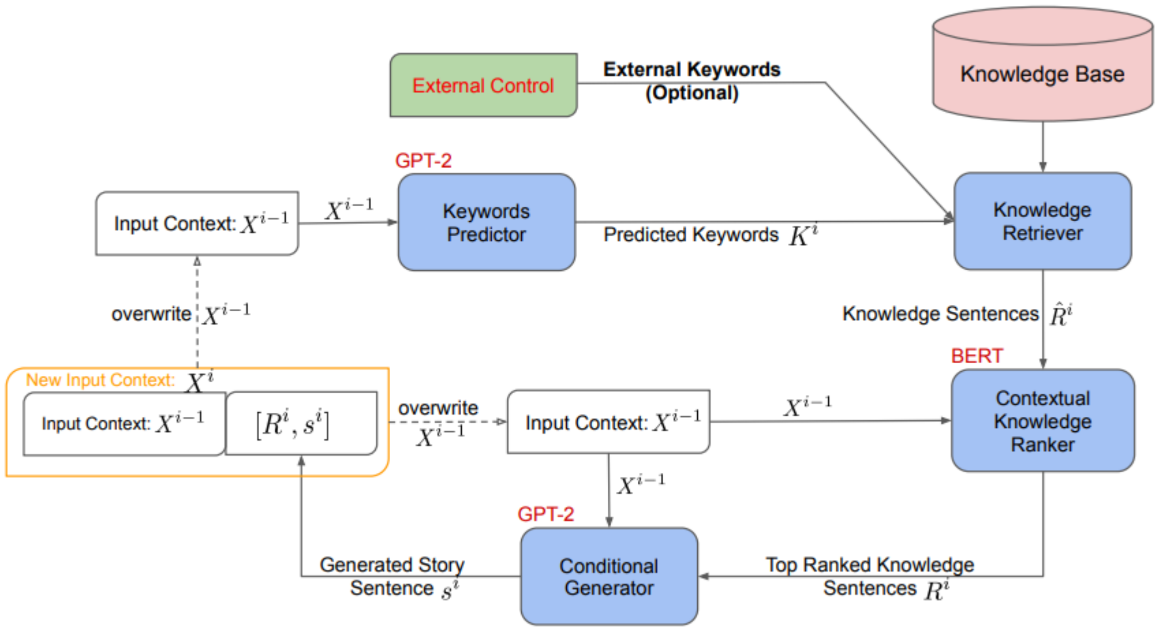

Figure 3.4: Architecture of MEGATRON-CNTRL-8b [4].

Source A	Coherence ↑	Fluency ↑	Source B
M-CNTRL-124M	**78.5%** - 13.0%	**66.5%** - 22.5%	Yao et al. (2018)
M-CNTRL-124M	**46.0%** - 39.0%	**44.5%** - 43.5%	Guan et al. (2020)
M-CNTRL-355M	**56.0%** - 30.5%	**46.5%** - 30.5%	Guan et al. (2020)
M-CNTRL-355M	**52.0%** - 31.5%	**46.5%** - 39.0%	M-CNTRL-124M
M-CNTRL-774M	**44.5%** - 41.5%	**56.0%** - 33.5%	M-CNTRL-355M
M-CNTRL-2B	**50.5%** - 30.5%	**53.0%** - 39.0%	M-CNTRL-774M
M-CNTRL-8B	**46.0%** - 39.5%	46.5% - 46.5%	M-CNTRL-2B

Table 3.6: Human evaluation in [4].

We can compare the results from Table 3.5 and Table 3.6 in Table 3.7 and we can see a relation between a lower perplexity and a better performance in the metrics evaluated by humans as consistency, coherence and fluency.

We can obtain a total preference of the cohesion and fluency as seen in Table 3.8 and Table 3.9

In Table 3.10 we can see the comparison of the MEGATRON-CTRL language models and in Figure 3.5 we can see a graph of this comparison.

Name	PPL ↓	Repeat ↓	Distinct ↑	Consistency ↑ (Human Eval)
GPT-2-124M	6.98	27.2	74.1	69.5
Yao et al. (2018)	NA	**13.3**	63.7	49.0
Guan et al. (2020)	7.04	22.1	77.1	67.0
M-CNTRL-124M	9.37	20.0	80.1	74.5
M-CNTRL-355M	8.02	19.9	81.6	75.5
M-CNTRL-774M	6.58	21.3	81.6	80.5
M-CNTRL-2B	6.31	21.2	82.6	89.0
M-CNTRL-8B	**6.21**	21.2	**82.8**	**93.0**

Source A	Coherence ↑	Fluency ↑	Source B
M-CNTRL-124M	**78.5%** - 13.0%	**66.5%** - 22.5%	Yao et al. (2018)
M-CNTRL-124M	**46.0%** - 39.0%	**44.5%** - 43.5%	Guan et al. (2020)
M-CNTRL-355M	**56.0%** - 30.5%	**46.5%** - 30.5%	Guan et al. (2020)
M-CNTRL-355M	**52.0%** - 31.5%	**46.5%** - 39.0%	M-CNTRL-124M
M-CNTRL-774M	**44.5%** - 41.5%	**56.0%** - 33.5%	M-CNTRL-355M
M-CNTRL-2B	**50.5%** - 30.5%	**53.0%** - 39.0%	M-CNTRL-774M
M-CNTRL-8B	**46.0%** - 39.5%	46.5% - 46.5%	M-CNTRL-2B

Table 3.7: Comparison of Table 3.5 and Table 3.6

MCLM - left	vs. Preference- left	vs. Preference - right	MCLM - right	Total preference - right
355M	52.0%	31.5%	124M	31.50%
774M	44.5%	41.5%	355M	52.00%
2B	50.5%	30.5%	774M	55.76%
8B	46.0%	39.5%	2B	92.32%
			8B	107.51%

Table 3.8: Total preference in coherence

MCLM - left	vs. Preference- left	vs. Preference - right	MCLM - right	Total preference - right
355M	46.5%	39.0%	124M	39.00%
774M	56.0%	33.5%	355M	46.50%
2B	53.0%	39.0%	774M	77.73%
8B	46.5%	46.5%	2B	105.63%
			8B	105.63%

Table 3.9: Total preference in fluency

MCLM	Perplexity	Consistency	TP in coherence	TP in fluency
124M	9.37	74.50%	31.50%	39.00%
355M	8.02	75.50%	52.00%	46.50%
774M	6.58	80.50%	55.76%	77.73%
2B	6.31	89.00%	92.32%	105.63%
8B	6.21	93.00%	107.51%	105.63%

Table 3.10: Comparison of MEGATRON-CTRL LM

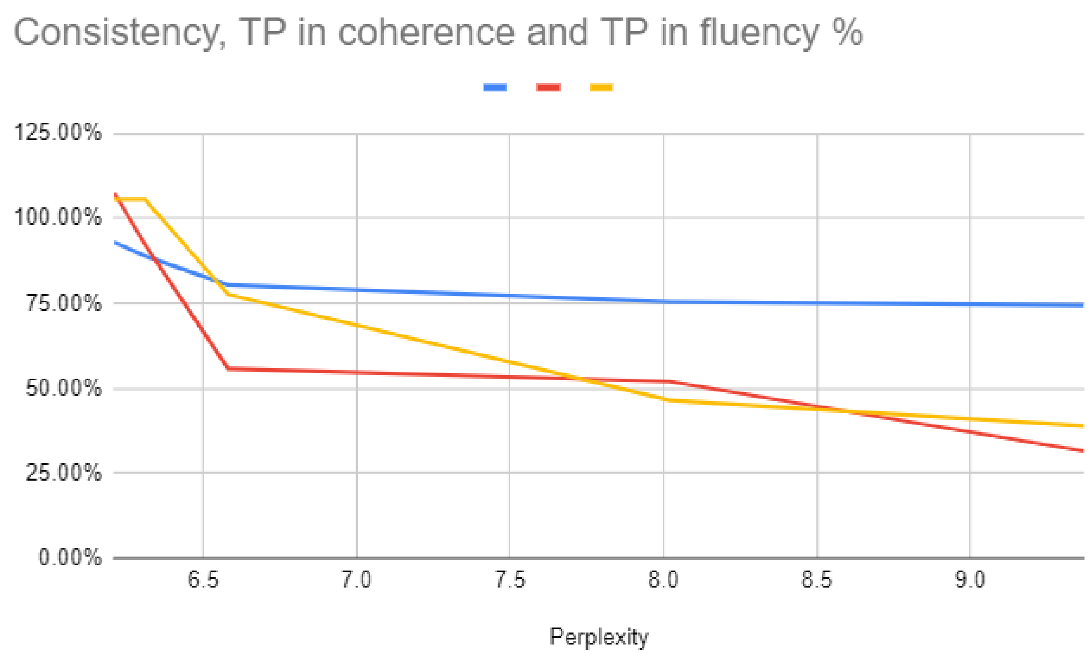

Figure 3.5: Graph of the comparison of MEGATRON-CTRL LM

3.2.3 Narrative Interpolation

In [5] Wang et al. propose a model based in the generation of text with the GPT-2 pre-trained language model[40] and a coherence ranker trained on a RoBERTa pre-trained language model with the ROCStory dataset. We can see one step of the INTERPOL process in Figure 3.6.

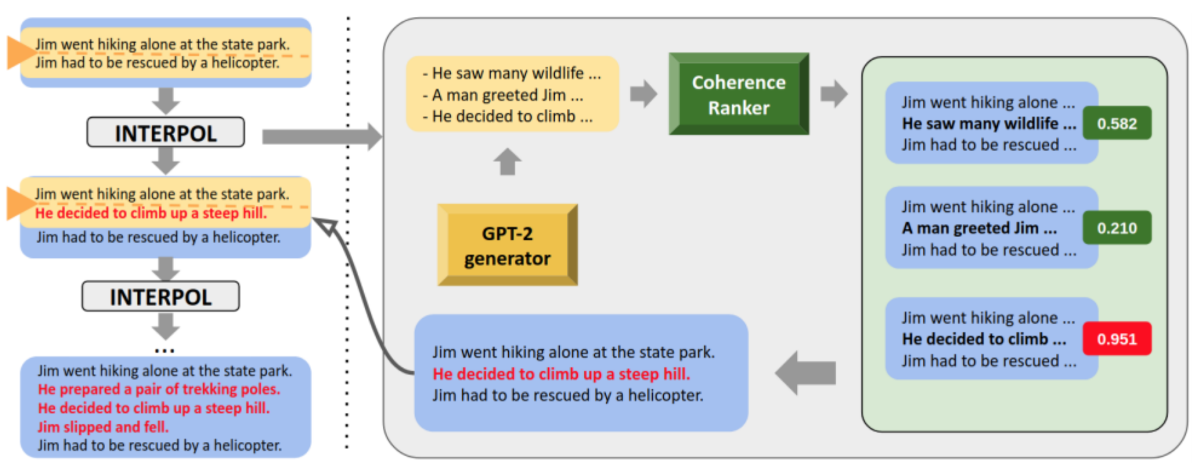

Figure 3.6: Iterative narrative generation by INTERPOL[5].

Beginning: *Sarah cherished her favorite toy that her mother gave her.*
Storyline-guided model (PAW, Yao et al. (2019))
Storyline 1: `played toy-toy disappeared-never found it-bought new toy` **Generated**: *She loved to play toy with it every day. One day the toy disappeared. Sarah searched all over the house and never found it. She bought herself a new toy.*
Storyline 2: `played-disappeared-lost-new` **Generated**: *Sarah played in the courtyard. Her mother disappeared. Sarah found herself lost in the woods. She bought a new house.*
Ending-guided model (INTERPOL, ours)
Ending: *Her parents had to buy Sarah a new toy.* **Generated**: *Sarah cherished her favorite toy that her mother gave her. Sarah's mother forgot about the toy at school. Sarah was upset that her mother had left the toy at school. Sarah was angry and cried.*

Table 3.11: Example of a story generated by INTERPOL [5].

Model	*Single-sentence*		*Full-story*	
	L2R	NR	L2R	NR
Perplexity	8.90	**6.76**	9.93	**7.53**

Table 3.12: LEFTTORIGHT vs . NORANKING in [5]

"Left: high-level flow of the interpolation procedure — for each insertion point, produce an interpolation sentence. Right: detailed view of INTERPOL. First, with a selected left-right context pair, the text generator proposes a list of interpolation candidates. Then the coherence ranker picks out the globally best candidate in the context of the story-in-construction. NB: the order of interpolation is "bisectional": for a 5-sentence story, taking s1, s5 we generate s3, then taking s1, s3 we generate s2; finally generating s4 given s3, s5"[5]. We can see an example of a story generated by this work in Table 3.11.

They have three evaluations, in the first they evaluate the perplexity of using only the first sentence or the first and fifth sentences to the model without coherence ranking, we can see the results in Table 3.12. Their second table shows a human evaluation that highlights the importance of the coherence ranker, we can see the results in Table 3.13. Finally they compare their stories with those from PAW[27], as shown in Table 3.14.

	NR (-ranking) better	INTERPOL (+ranking) better	Both Good	Both Bad
Coherence	0.033	**0.611**	0.089	0.267
Preference	0.078	**0.589**	0.044	0.289
	NR	INTERPOL		
Faithfulness	0.278	**0.834**		

Table 3.13: NORANKING vs. INTERPOL in [5]

	PAW better	INTERPOL better	Both Good	Both Bad
Coherence	0.178	**0.444**	0.233	0.144
Preference	0.156	**0.507**	0.167	0.170
	PAW	INTERPOL		
Faithfulness	0.333	**0.744**		

Table 3.14: INTERPOL vs. PAW in [5]

3.2.4 Consistency and Coherence

In "Consistency and Coherency Enhanced Story Generation" [6] the authors emphasise the lack of coherence, consistency and coreference in the texts generated by pre-trained language models. We can see the framework proposed in Figure 3.7.

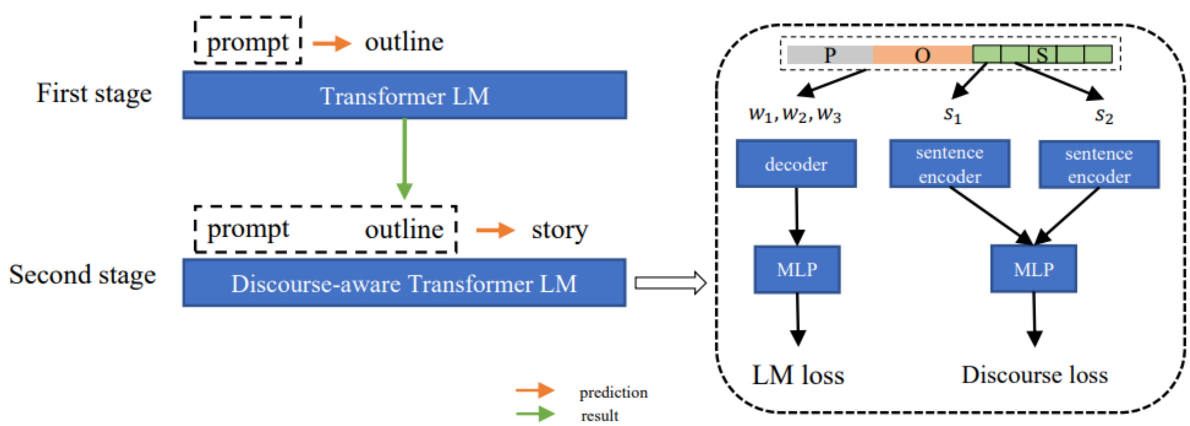

Figure 3.7: Framework for story generation in [6]

We can see an example of a story generated by this work in Table 3.15. As we can see the framework has two stages, in the first they take a prompt as an input and then generate the outline, they use a GPT-2 [40] and BERT [39] transformers.

For the first stage they use WRITINGPROMPTS dataset and explore two ways of obtaining the outline: key words and abstract. For the second stage they take into account shallow discourse relations between adjacent sentences. Therefore, they fine tune BERT in order to get a golden discourse marker prediction model, and use a dataset comprehending Treebank 2.0 pairs and Books8 dataset. They also encourage the model to attend correct entities adding a supervision on attention weight of entity mention tokens. They use Stanford's CoreNLP tool to extract coreference annotation of stories.

We can see an automatic evaluation in Table 3.16, a human evaluation in Table 3.17 and a comparison of different outlines in Table 3.18. "Unknown" refers to unknown labels in generated stories, and they use Stanford's CoreNLP tool to extract coreference chains of generated stories.

Prompt: The zombie apocalypse is upon us but all humans are immune to the virus . All other mammals are susceptible .

GPT2P: As I sat there I tried to remember how many days it had been since the last one died . [My wife and daughter were dead , I knew they would n't be alive any time soon .]$_{1a}$ The food ran out quickly , they had to be taken away to stay in the house , but that was no reason to be angry if there were a few people left living there . After a month of food being gone a little girl sat next to me . She was the only other survivor who was n't bitten , as it was hard for us to eat them , we were all sickly . That would be another month to go and if the other animals were still alive in the food , they would n't be able to eat her . " Hey little one . " I looked up and saw her . [I had a big smile on my face at this point .]$_{2a}$ [My wife was the only one who noticed , but her smile was all that was visible of me , and not of her .]$_{1b}$ [I looked at her , and smiled sadly .]$_{2b}$ She looked sad , but did n't care . She never cared for me .

C&C (generated abstract): He was the last to see the end of this world . The only reason John had survived was by his own hand , and it took much more than he could bear with a few months to heal his wounds . It took a few months for him to be able to see the end and get his arm back from the dead , and he did n't care for a few more years without having to go back .

C&C (generated story): When he had been first laid to rest the first few days he did n't care , and he had thought it was just a small thing , but it started getting worse . He was the last to see the end of this world . The only reason John had survived was by his own hand , and it took much more than he could bear with a few months to heal his wounds . John had lived the rest of his life on Earth without anything else but the clothes on his back , and the rifle he kept at home when he was alone . The first few days were easy . A little bit of food , a bit of protection . He spent a few days walking around the city and drinking a beer , thinking he was safe , but the next few months were hell . It took a few months for him to be able to see the end and get his arm back from the dead , and he did n't care for a few more years without having to go back . It was better to go back , to be safe , so he would be safe for a while , and so he would n't get infected .

Table 3.15: Comparison of different methods in [6]

Method	Perplexity↓	Dis-1(%)↑	Dis-2(%)↑	Unknown(%)↓	Coref Chains↑
ConvS2S	34.61	0.400	5.191	76.01	5.52
FConvS2S	33.97	0.482	6.271	75.60	5.43
GPT2	29.50	0.474	6.796	74.95	5.67
GPT2P	25.64	0.493	7.333	73.61	5.61
C&C(0% ground truth outline)	30.84	0.531	7.379	75.19	5.98
C&C(50% ground truth outline)	19.21	1.311	13.253	75.15	5.97
C&C (100% ground truth outline)	10.32	1.509	15.266	74.97	5.80

Table 3.16: Automatic evaluation in [6]

Method	Relevance			Grammaticality			Logicality		
	Win(%)	Tie(%)	Lose(%)	Win(%)	Tie(%)	Lose(%)	Win(%)	Tie(%)	Lose(%)
C&C vs. FConvS2S	**23**	66	11	**28**	53	19	**40**	33	27
C&C vs. GPT2P	**21**	60	19	**17**	69	14	**31**	47	22

Table 3.17: Human evaluation in [6]

Method	Perplexity↓	Dis-1(%)↑	Dis-2(%)↑	Unknown(%)↓	Coref Chains↑
First stage					
keyword	74.46	0.964	7.132	/	/
abstract	35.53	0.776	10.060	/	/
Second stage					
story with keyword	17.82	0.461	6.188	74.26	5.67
story with abstract	10.65	0.512	7.358	74.54	5.81

Table 3.18: Comparison of different outlines in [6]

Chapter 4

Our proposal

4.1 Key elements

The biggest challenge of this task is the lack of metrics that evaluate automatically the performance of the stories generated by the variety of models in the state of the art. However, there are key elements that have been evaluated by humans.

Another major issue is the change of the nomenclature of those key elements, in some works the coherence is the consistency of other works, in some cases the coherence is treated as the cohesion of other cases.

In this work, we take the same nomenclature as the one used in the recent state of the art papers.

Some of the key elements and their definitions are as follows:

- Readability - Is the capability of an automatic generator to write stories using natural language.
- Coherence - Coherence is the relevance of a pair of linked events.
- Cohesion - Cohesion is the adequate writing of an event and the links that they have. It is also known as fluency.
- Consistency - Consistency is the the respect the events in the past with the rules of the specific world in the story.

- Novelty - Novelty is the creation of new situations and stories.
- Interestingness - Interestingness is the level of engagement on the reader. Most of the works do not take this element into account.

The works in the recent state of the art have shown that the generation using pre-trained language models do not have a problem with the readability of their stories.

The coherence in the state of the art, is an inter-sentence coherence, some works also take into account the coreference for their works. The coreference is outside the scope of the present work. There is not a metric for the evaluation of inter-sentence coherence.

Most of the works either new architectures of pre-trained language models or new models for automatic generation of stories have shown that the perplexity is closely related to the fluency or cohesion of the texts generated.

For having consistency the model needs a huge capability of natural language understanding as well as common sense knowledge in order to verify that a new event or situation respect the rules of the world and the previous events and situations. It should be noted that it does not have a metric associated with it.

Novelty is associated with Distinction and Repetition metrics. Distinction shows the percentage of different n-grams with respect to the other n-grams. Repetition evaluates the percentage of generated stories that repeat at least once an n-gram.

Is really difficult to measure the interestingness of a story, most of the research that takes this element into account make their models emotionally aware or emotionally-guided but there is not a metric to measure the level of engagement the reader could have with the written stories.

4.2 Framework

Our proposal is divided into various stages and modules. We will present the modules and stages independently and then the framework.

4.2.1 AVM and BCO generator

For every situation in Polti's book and with the help of Figgis' book, there is a set of templates to create automatically three linked situations and their AVM. The three phrases are based on the beginning, climax and outcome of a situation. It is worth noting that the same process can be applied to every Polti's subsituation. For the scope of this proposal a full story is a Polti's situation represented in natural language.

With the above in mind, we will have 108 different templates and a way to write a sentence with each of them. An example of an AVM is shown in Figure 4.1, inspired by [13]. For the possible values, p is a Polti's situation, n in an integer, and the WHY attribute can be linked to another AVM.

TFS situation	
P_SIT	p
TIME	n
WHO	character
WHAT	verb
WHERE	place
WHOM	character
TOOL	thing
WHY	situation
NEG	*boolean*
HAPPENS	*boolean*

Figure 4.1: Example of an AVM

The process starts with a blank AVM. Almost all the values will be automatically filled with a Knowledge Base except for p, n, and the field for a situation in the WHY attribute. Then a situation will be randomly selected and will find a p value. The AVM will pass to a situational script, in order to generate three phrases. We can see this process in Figure 4.2

The BCO phrases correspond to beginning, climax and output situations that represent an outline for a Polti's situation.

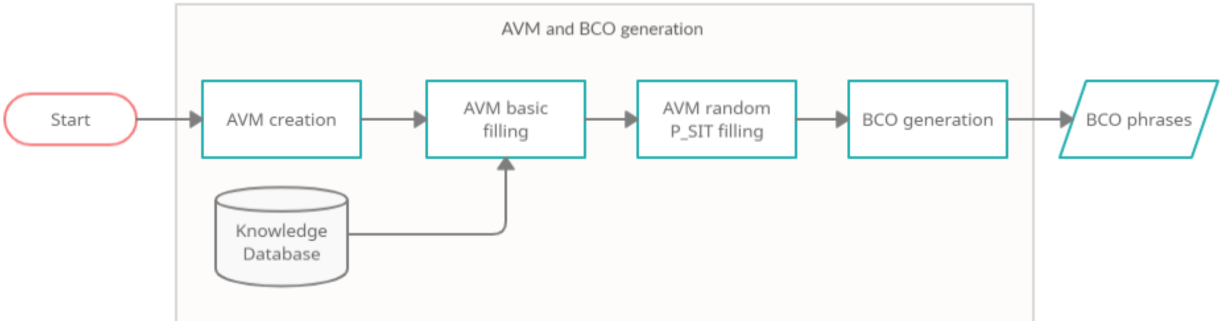

Figure 4.2: AVM and BCO generation

4.2.2 Fictional and semi-coherent generator

We use an autoregressive model in order to generate novel sentences with only a left context. We will pre-train the model with a Fiction dataset and then fine-tune it with a Causal dataset, we propose Books 8 as the causal dataset.

The training stage can be seen in Figure 4.3.

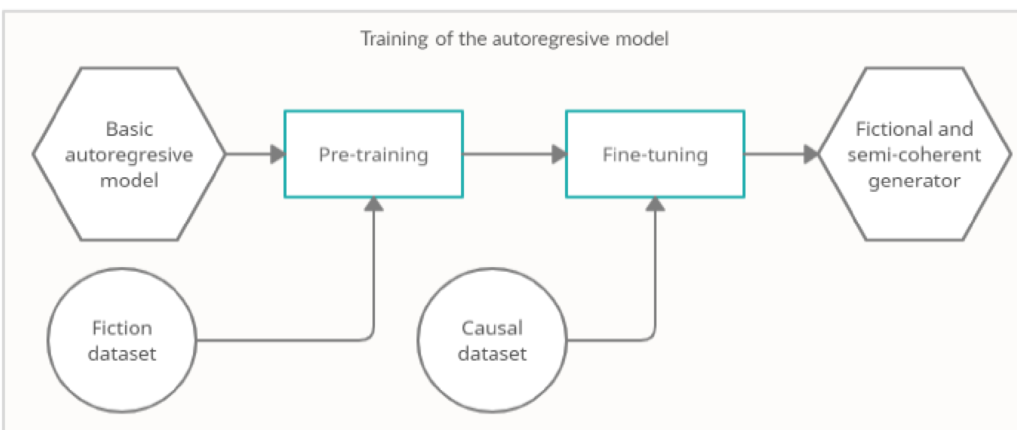

Figure 4.3: Training of the Autoregresive model

We refer as semi-coherent to the fact that the outputs will have coherence and cohesion with the input.

4.2.3 Coherence filter

In order to have a narrative interpolation [5] we need a coherence filter, in contrast to them we named it filter because we will keep only one of the candidate sentences.

As in the narrative interpolation paper, we propose a masked language model. We will use a combination of ROC Stories and WRITINGPROMPTS as the Correct stories dataset.

The stories dataset will have some changes as Repetition, Irrelevant phrases and Out of order phrases in order to generate stories with a bad coherence level [6]. We use a masked pre-trained language model because of the fact that it takes into account both, the left and right contexts. The process is shown in Figure 4.4.

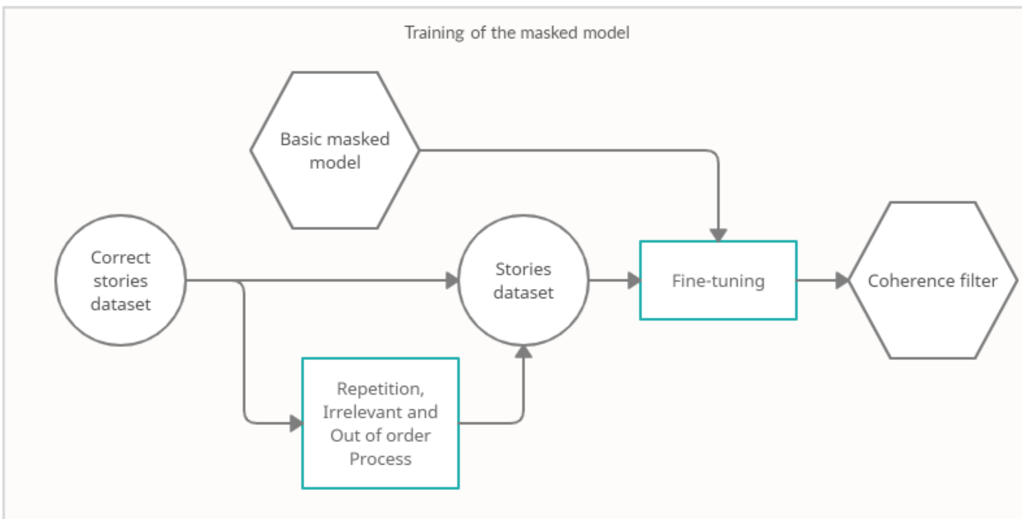

Figure 4.4: Training of the Masked model

4.2.4 Fictional and coherent interpolation

We use a process similar to [5] the main differences are that we use a fictional and semi-coherent generator, instead of a basic autoregressive pre-trained language model; and also the changes proposed for the coherence filter. An iteration of the interpolation proposed in this work is shown in Figure 4.5.

As we can see two phrases, Alpha and Omega, go into the interpolation system, Alpha is the input for the text generator and it generates ten candidate sentences. The coherence filter ranks the candidate sentence Beta, that fits the best between the Alpha and Omega phrases. There is a process that creates an AVM for the Beta phrase and that updates the content of the Alpha and Beta AVM for the TIME and WHY attributes. Therefore, we have updated AVM's and a new sentence in the story.

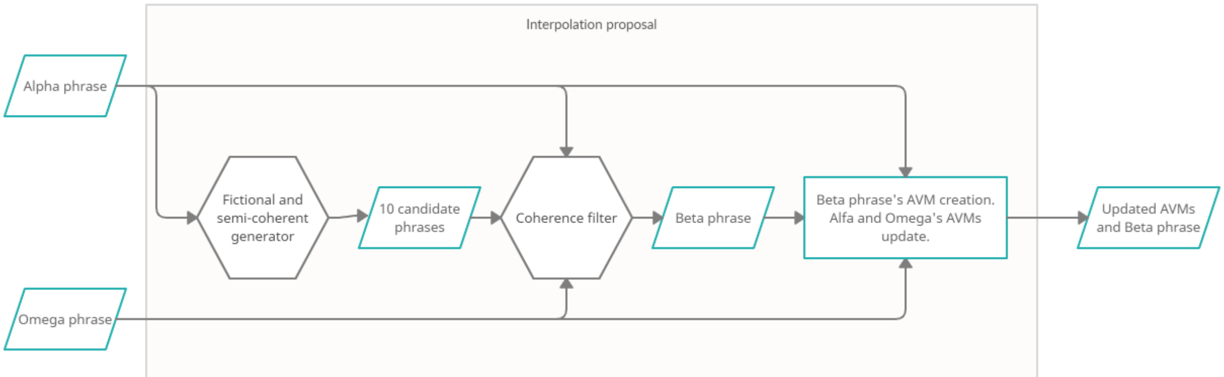

Figure 4.5: Interpolation proposed

4.2.5 Full story generation

For creating a full story that represents a Polti's situation we propose a schema where we interpolate three new phrases between the B and C phrases, and three new phrases between the C and O phrases. We can see the process in Figure 4.6.

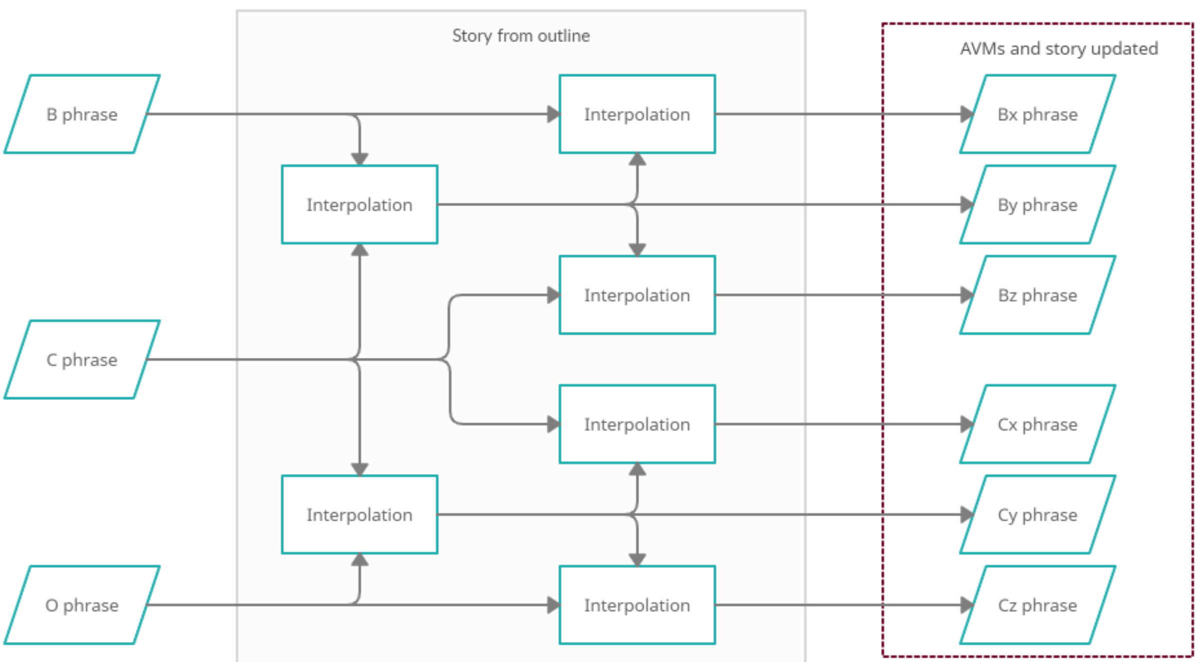

Figure 4.6: Full story generation

4.2.6 Automatic generation of fiction stories

We can see the general proposal in Figure 4.7.

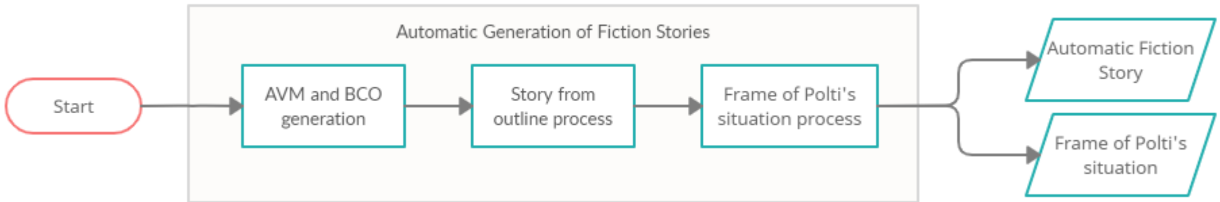

Figure 4.7: Automatic generation of fiction stories

4.3 Implementation scope

Because of time constraints, in this work we only implement the pre-training of the autoregressive language model. We trained the language model with a fictional dataset and a common sense dataset [3] in order to observe the impact of a fictional dataset in the pre-trained language model. We evaluate the perplexity of the models for comparison and used diverse stages of the training to generate stories and compare them. Further details are described in chapters 5 and 6.

Chapter 5

Implementation

As mentioned in chapter 4, our goal is to train the autoregressive pre-trained language model with Fiction and Common sense datasets.

5.1 Datasets

5.1.1 Common sense dataset

The dataset used for the Commonsense model is the same used in [3]. It is based in ATOMIC and ConceptNet knowledge bases. They used the relations of tuples in the knowledge bases in order to generate simple phrases. For example a relation of the kind *IsA* they write it as *Object1 is a Object2*.

5.1.2 Fiction dataset

The iction dataset was created with the one thousand most downloaded fiction books from the Project Gutenberg library. We preprocess the text from the books in order to eliminate the extra content. We followed two approaches in the first we used the preprocessed text and in the second approach we used the algorithm TextRank for obtaining the summaries of the books. However, preliminary experiments showed that training with disordered text affects negatively the fluency of the text generated by the model, for that reason we used the first dataset as the final Fiction dataset.

5.2 Technical details

The autoregressive pre-trained language model used was GPT-2 [40], we used the base sized model. In order to train with the Fiction and Common sense datasets, we build a tokenizer. As in GPT-2 we crate the tokenizer with a byte pair encoding. We set the following special tokens "<s>", "<pad>", "</s>" and "<unk>".

For training the pre-trained language model with our datasets, we should add all our texts in a single string and the tokenize it. For both models we used the same parameters and the HugginFace library. The parameters used are the following:

- Block size = 512
- Masked language model = False
- Optimizer = Adam with default parameters.

We used the frameworks TensorFlow and PyTorch, both have a similar performance, the principal difference is the VRAM of the GPU consumed and the transformer language models available for them.

We used an NVIDIA Tesla V100 Volta GPU with 32GB of VRAM, this is the principal reason for the choice of block size equals to 512.

We trained the fictional pre-trained model for 45 epochs, the training time was superior to 270 hours.

We trained the Common sense pre-trained model for 15 epochs, the training time was inferior to 27 hours.

It should be noted that there were a few issues with the experimentation, the real consumed time was fairly superior mainly due to the fact that it was a borrowed and shared GPU. It was difficult to find the right version of the frameworks in order to use the GPU with the incapability of modifying its drivers. As it was shared there were idle times. The fact that this topic is cutting-edge knowledge represents a difficulty in finding responses to specific issues in the implementation. In addition to all of above, we only have had access to the GPU for 2 months.

Nevertheless, it should be noted that without this valuable resource would be impossible this implementation on time, since 1 epoch for training the fictional pre-trained language model with the former resources took more than 120 hours. That implies a training time superior to 8 months for both models.

Chapter 6

Experiments and results

6.1 Fiction model

We trained the fiction pre-trained language model for 45 epochs We can see the loss depending the epoch in Figure 6.1. The model has a perplexity value of 21.54, close to 19.93 reported in [40]. We present a set of stories generated by this model in different epochs as well as the perplexity of such models.

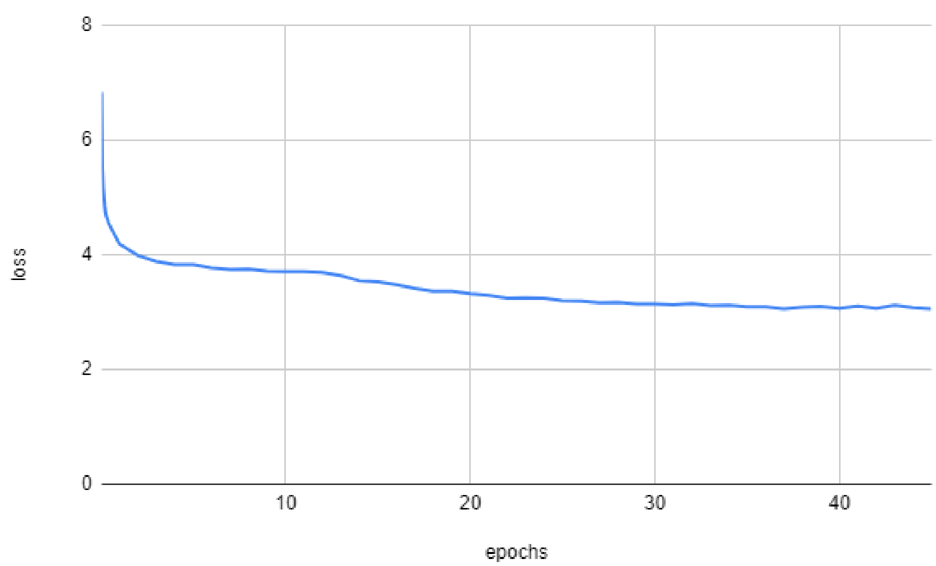

Figure 6.1: Loss in the fiction training

In Table 6.1, we show the performance of the Fiction pre-trained language model with

Input	Fiction-1
once upon a time	" which all the events,000 miles from the water and from the water. The water and the water had now been given for one time, and it was still as a couple of miles as it opened to the water.
i was looking forward to roasting my own turkey .	his eyes were in a low tone. "I know that the ship have done it. I could not talk. But I wish we were getting along and that's all." "The fleet of our fleet are in the
she really loved the sun .	. He had been very glad if he had come out of it so that his mind would have been so far from the right hand of the city. * * * * * "He's an
that person knew the solution to a problem .	The very-time and with great power. For the first time we were all silent--a couple of and white-shaped white, two-world of the walls before us, and the huge, yellow-green-
he was on a road trip .	. "It was an old man about the road," she said. "But I do not feel like that." That was all her words. But the words were moving away, and the road came out.

Table 6.1: Examples of stories generated by Fiction-1

5 different outputs. The perplexity value of this model is 72.3. We can see even with only 1 epoch that the fiction is present, specifically in the first and second phrases. Most of the outputs have syntactic and higher level mistakes.

In addition, in Table 6.2 we see the outputs of the model Fiction-5 with a perplexity value of 51.41. As in the previous model, the outputs generated lack of cohesion, nevertheless, the impact of fiction is more visible.

We show the text generation of model Fiction-10 in Table 6.3. The model's perplexity is 40.85. At this point the enhance of the fiction is evident and we can see an improvement in the cohesion of the generated text.

The results from the model Fiction-15 are shown n Table 6.4. The model has a perplexity value of 34.47. These are the first outputs where the fluency is acceptable comparing them with the regular outputs of the GPT-2 [40] model.

The Fiction-30 model obtained a perplexity value of 23.1 and their results can be seen in Table 6.5. With this model the outputs have a decent fluency in terms of a pre-trained language model [40]. The presence of fiction in the generated text is undeniable.

In Table 6.6 we observe the final results from our pre-trained language model. It

Input	Fiction-5
once upon a time	miles away," he said, "when we have a chance, I've failed to stop that. When our return, I am going to take it with me, and I am going to meet you at the time. It
i was looking forward to roasting my own turkey .	had not been a big one of their masters's masters. He felt slightly excited about a hundred times at his own length, then turned to the other. He made sure it wasn't so much, then was very much hurt
she really loved the sun .	also: And it seemed, when the world swept through the air by the heavens, it became like a sky-white sun, and it was the sun that the heat had so slowly occurred. The wind, that
that person knew the solution to a problem .	had been the best day in its last year. A woman's mother was a child whom she would consider an honest woman if you see so much if you would not see. And not one woman.
he was on a road trip .	of its former, it was no longer a man or a woman but no, of course it could be a woman or a woman. They must know that the time to go out and talk on the spot, and I must

Table 6.2: Examples of stories generated by Fiction-5

Input	Fiction-10
once upon a time	, a thing like it, that would keep its own course away from me, and that would kill me, when the first thing had appeared, and I had tried so, and at least I could never die nor die nor
i was looking forward to roasting my own turkey .	in what they were of any sort. Then--to be dead, they were more--that there might be any living--the girl and the man--a man of human nature, an ideal or a little human being--
she really loved the sun .	to me, my head was not over five miles away from me in my mind. I felt a little of my thoughts and I went down. I went back to my own desk and went to the sofa. I turned
that person knew the solution to a problem .	"Oh, I'll bet I don't! Now that's what's wrong about it," he mumbled. "It's still a hard chance on us--in the middle of the evening, after a bit of work
he was on a road trip .	1sS: _The sun of Mars_ _But surely we wouldn't have stopped us from the Earth's atmosphere yet? That's the best." "And what's the matter with you?" "We

Table 6.3: Examples of stories generated by Fiction-10

Input	Fiction-15
once upon a time	, and they never will. But of course the war has been put in the same place. And while they've been running, the war has been turned out by the war. So, it is my own. Your enemy
i was looking forward to roasting my own turkey .	. A moment, it seems to anybody. The man moved closer to Craig and a moment later he looked back from the control panel. "All right, George. I don't think he can help me. But in case he can
she really loved the sun .	. And there was the great river overhead with the dead and there was nothing there for them, but their last hope of escaping in the great forest, that the forest must be a great water cliff where the forest had no place
that person knew the solution to a problem .	, what are you doing about it? He was in no state of mind to be noticed, and yet he seemed to think he would rather not be noticed, than the rest of us. And now, as soon as us
he was on a road trip .	A million miles to Jupiter Jupiter to Saturn for ever,000 miles, and an air-vessel which could easily be found by Jupiter's astronomers. But a vast series of vast quantities of rock was of a sort that existed

Table 6.4: Examples of stories generated by Fiction-15

Input	Fiction-30
once upon a time	and even for a mile on foot, was the very top of a huge pit. From its summit rose the dark forest, the forest between two of them as if it had been a hill, but which must have had the place of shadow in
i was looking forward to roasting my own turkey .	At the equator, a number of comets are visible, appearing beyond all doubt as the distance between their light, and, at the same time, the cause of their existence. How the comets go, comets don't
she really loved the sun .	, or to increase speed, at about a hundred miles a minute. In the distance there was little, if not much. One of the nearest dots reached the surface of the planet; the other, in passing, plunged
that person knew the solution to a problem .	and have a mighty pleasure in watching a tremendous globe rise vertically in the sky, and a tremendous sphere become a sphere of immense size or the planet a star of the zenith. There was a time when this tiny sphere had
he was on a road trip .	¼ in a volume that would enable him to see through the open air. The sky was black with the black blue of dawn, and all that it seemed was a gigantic crater, a crater similar

Table 6.5: Examples of stories generated by Fiction-30

Input	Fiction-45
once upon a time	, the star _vua_ was no more than twenty miles away in diameter! "Look at it!" yelled the skipper, darting the map at the end, then pointing to the surface again. "It is an
i was looking forward to roasting my own turkey .	.... He could tell by her eyes the familiar pattern of stars from the distance and sound and sound of the night and sound of the night--only when one was aware of the sounds of the stars, and another also from
she really loved the sun .	" was the best the firmament could have seen. "We can't have reached that. There is a constant force of some kind. If there are any means of determining a human civilization, we can't just as soon
that person knew the solution to a problem .	.... The world was no more than a half-life. In one year, it came up into another, as much as the years. The days were coming near. There were a number of small gray suns
he was on a road trip .	at the time when the speed of light from each sphere was greater than from the sun, plus a speed which would only take the distance of the world into a great scale, while the stars

Table 6.6: Examples of stories generated by Fiction-45

reaches a perplexity value of 21.54. From our perspective the outputs from the model Fiction-30 have more cohesion, this is the first time in our experiments that a model with lower perplexity has worst fluency than another with a greater perplexity. The outputs also denote a tendency to generate stories with a galaxy thematic.

As we can see by the results, a good level of cohesion is reachable with only a pre-trained language model, notwithstanding the level of coherence is really low. There lies the importance of having two processes focused in coherence.

6.2 Commonsense model

We trained the Common sense pre-trained language model for 15 epochs. We show the loss depending the epoch in Figure 6.2

This model's perplexity is 8.08 and is close to 8.04, the perplexity reported in [3]. We could improve it with a few more epochs, but it remains this way for the sake of a better comparison between the common sense and the fictional datasets, therefore, we could see

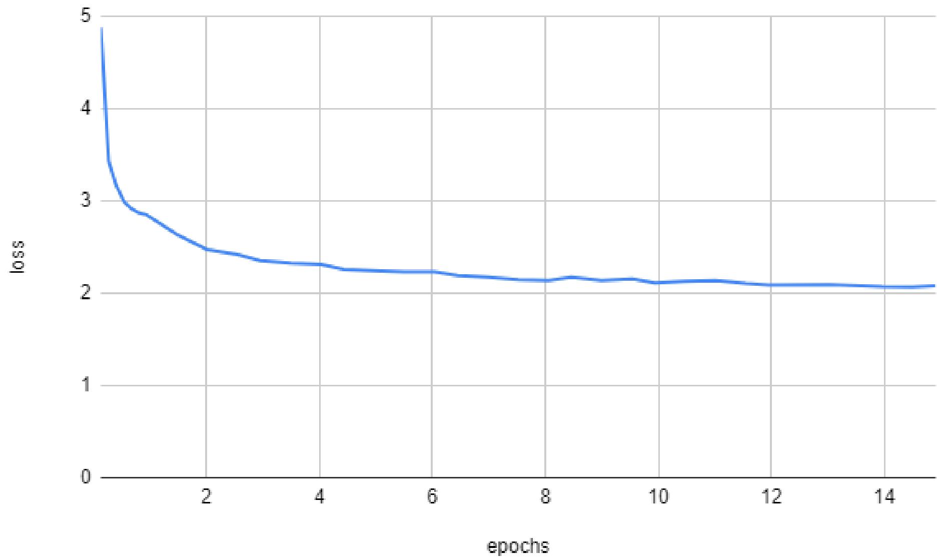

Figure 6.2: Loss in the common sense training

a fair difference of the impact of fiction in autoregressive pre-trained language models.

As in the previous section, we present a set of stories generated by the model in different epochs as well as the perplexity of such models.

In Table 6.7 we can see the results of the Commonsense-1 model. From this stage we can see that the model chooses to generate the text either using ConceptNet knowledge or using ATOMIC knowledge. The level of cohesion is really poor and the perplexity for this model is 17.81.

We can see the results of this model with a training of 5 epochs in Table 6.8. This model's perplexity is 9.39. We see the same tendency of choosing a specific dataset for generating the text, we can see an increase of fluency but is still poor. We can see the outputs of the model trained 10 epochs in Table 6.9. The model has a perplexity value of 8.5. There is not an observable improvement in fluency.

The results of our final Commonsense model are presented in Table 6.10. As mentioned before, this model's perplexity is 8.08. It shows an increase of cohesion, however, is barely acceptable compared with GPT-2 [40] model outputs. It is surprising that in [3], they used a model with the same dataset and a similar perplexity level and even with this level of cohesion they manage to obtain coherent and cohesive stories only by fine-tuning the model. Another element that should be highlighted is the presence of phrases like first

Input	Commonsense-1
once upon a time	yx person paint has first subevent of learn how to send someone has property lonely go for jog has subevent you want to have fun make patchwork quilt has prerequisite for learn someone is at an building
i was looking forward to roasting my own turkey .	MALE] needs to know what [FEMALE] is. [MALE] pulls the none out of the store. [MALE] wants to be kind. [MALE] sees a great job. [FEMALE] will [
she really loved the sun .	of a train has subevent a a good health a car room-six and the the university has property fun a toy dog is used for store the meal a car room has a a right the right of the wall
that person knew the solution to a problem .	ine piece of new people. [MALE] wants to stop the same the doctor.. [MALE] calls none on the floor. [MALE] needs to take the floor. [MALE] tells [FEMALE] to
he was on a road trip .	heFEMALE] a school. [MALE] is rude. [MALE] buys a new pair of bed. [MALE] is mean. [MALE] pays personx's none in order. [MALE] [MALE

Table 6.7: Examples of stories generated by Commonsense-1

Input	Commonsense-5
once upon a time	of people's home. [MALE] goes fishing with persony's father. [MALE] wants to get a good book. [MALE] needs a new haircut. [MALE] wants play outside. [MALE
i was looking forward to roasting my own turkey .	and a person does not desire be near your head the time they has property beautiful paint a house has first subevent of put water in the water a baby is not capable of sleep in the water we
she really loved the sun .	mh is nenses in a store some people think people is capable of cause being understood lakes has a more than one person person desires be around people some people think people is capable of like water
that person knew the solution to a problem .	[MALE] uses personx's none to prevent. [MALE] will upset. [MALE] sees a moose. [MALE] is thoughtful. [MALE] has had it. [MALE] is brave.
he was on a road trip .	person x dose not have property part of all human sock has property easy to look up in shape golf has property fun bicycle is at at hotel person is at at hotel person is capable of

Table 6.8: Examples of stories generated by Commonsense-5

Input	Commonsense-10
once upon a time	waging war has subevent you break head of your bed the sun receives action known for hold the lawn wasted a brush has subevent you start writing buy presents for others has prerequisite for make them pay attention
i was looking forward to roasting my own turkey .	this is a lot of books this is an example of an old planet this is a television pipe this is an old planet in japan this is a very tall lion in colorado this is an art
she really loved the sun .	the ground is at the arctic water plant has subevent open your mouth lucid dream has prerequisite for make people laugh go fishing has subevent you trip a little get some physical activity has first subevent of decide what
that person knew the solution to a problem .	is in florida in a person is capable of own good health a finger is used for get a book people is capable of find a room to relax people used to have a job when people does not desire be
he was on a road trip .	the walls is capable of be given a ticket a person's mother desires learn different things the human being has property cold a person is capable of be using a ship a person does not desire be caught without bleeding

Table 6.9: Examples of stories generated by Commonsense-10

Input	Commonsense-15
once upon a time	of the man. [MALE] tells [FEMALE] to shut up. [FEMALE] will gets into trouble. [MALE] sees a rainbow. [MALE] wants to avoid them. [MALE] uses
i was looking forward to roasting my own turkey .	is an american part of his family a person desires have children iq is defined as last name of a person who always plays go to a sporting event has prerequisite for get out your shopping spree waking up
she really loved the sun .	. [MALE] loses ground. [MALE] wants to try to swim. [MALE] tries and get persony. [MALE] wants to talk to y. [MALE] takes none back to the store
that person knew the solution to a problem .	. is motivated by relax in their bed. [MALE] [MALE] is sad. [MALE] loses the first game. [MALE] needs to wear your clothes. [MALE] decides to try out for
he was on a road trip .	. [MALE] wants to run around. [MALE] asks my friend. [MALE] [MALE] gets a phone call. [MALE] eats [FEMALE] alive. [MALE] will satisfied with himself

Table 6.10: Examples of stories generated by Commonsense-15

and last phrases. This model is intended to represent the common sense knowledge, but those phrases help us to remember than ethics is a serious matter in the usage of these architectures.

6.3 Comparison of models

We compare both generated models with GPT-2 [40] Base outputs. We can see the results in Table 6.11. In Table 6.12 we can see the comparison of the perplexity of our model and the perplexity of GPT-2 in its base form. The comparison of the three models is shown in Table 6.13.

The level of cohesion or fluency is similar with GPT-2 Base and Fiction-45 models, the Commonsense-15 model has a lower level of cohesion. As stated before the impact of fiction in the model to the right is undeniable.

Input	GPT-2 output
once upon a time	Fantastic Bastion was never really my favorite game. I think the first few hours of Bastion were the best and most polished of my time. So much work, work... just so you know I
i was looking forward to roasting my own turkey .	This week's post was a bit different. The idea was very different–maybe you have another topic you have talked to someone about that you could help them understand. Here are three of our favorite stories (in no particular order
she really loved the sun .	In its latest update to its software for Android, Google announced, Google will soon begin to offer a beta of the Android Wear operating system. A preview of the app is here. Google said that it will begin opening
that person knew the solution to a problem .	A Florida man accused of threatening to kill three men just outside the city of Clearwater Saturday has been arrested for a second time on drug charges, a new law said Thursday. Prosecutors say the suspect was known for threatening
he was on a road trip .	Gavin and Sean's new app RedLight-Grow, will be released on August 11 in the U.S. and October 5 in Canada. RedLight-Grow will let you pick up your smartphone at a

Table 6.11: Examples of stories generated by GPT-2 Base

Model	Perplexity
GPT-2 Base	19.93
Fiction-1	51.41
Fiction-5	40.85
Fiction-15	34.47
Fiction-30	23.1
Fiction-45	21.54

Table 6.12: Comparison of perplexity of Fiction and GPT-2 Base models

Input	GPT-2 Base	Commonsense-15	Fiction-45
once upon a time	Fantastic Bastion was never really my favorite game. I think the first few hours of Bastion were the best and most polished of my time. So much work, work... just so you know I	of the man. [MALE] tells [FEMALE] to shut up. [FEMALE] will gets into trouble. [MALE] sees a rainbow. [MALE] wants to avoid them. [MALE] uses	, the star _vua_ was no more than twenty miles away in diameter! "Look at it!" yelled the skipper, darting the map at the end, then pointing to the surface again. "It is an
i was looking forward to roasting my own turkey .	This week's post was a bit different. The idea was very different–maybe you have another topic you have talked to someone about that you could help them understand. Here are three of our favorite stories (in no particular order	is an american part of his family a person desires have children iq is defined as last name of a person who always plays go to a sporting event has prerequisite for get out your shopping spree waking up	.... He could tell by her eyes the familiar pattern of stars from the distance and sound and sound of the night and sound of the night--only when one was aware of the sounds of the stars, and another also from
she really loved the sun .	In its latest update to its software for Android, Google announced, Google will soon begin to offer a beta of the Android Wear operating system. A preview of the app is here. Google said that it will begin opening	. [MALE] loses ground. [MALE] wants to try to swim. [MALE] tries and get persony. [MALE] wants to talk to y. [MALE] takes none back to the store	" was the best the firmament could have seen. "We can't have reached that. There is a constant force of some kind. If there are any means of determining a human civilization, we can't just as soon
that person knew the solution to a problem .	A Florida man accused of threatening to kill three men just outside the city of Clearwater Saturday has been arrested for a second time on drug charges, a new law said Thursday. Prosecutors say the suspect was known for threatening	. is motivated by relax in their bed. [MALE] [MALE] is sad. [MALE] loses the first game. [MALE] needs to wear your clothes. [MALE] decides to try out for	.... The world was no more than a half-life. In one year, it came up into another, as much as the years. The days were coming near. There were a number of small gray suns
he was on a road trip .	Gavin and Sean's new app RedLight-Grow, will be released on August 11 in the U.S. and October 5 in Canada. RedLight-Grow will let you pick up your smartphone at a	. [MALE] wants to run around. [MALE] asks my friend. [MALE] [MALE] gets a phone call. [MALE] eats [FEMALE] alive. [MALE] will satisfied with himself	at the time when the speed of light from each sphere was greater than from the sun, plus a speed which would only take the distance of the world into a great scale, while the stars

Table 6.13: Comparison of pre-trained models

Chapter 7

Conclusions and future work

7.1 Conclusions

In relation to the general objective, we generate stories with aspects like readability and cohesion, coherence, even the inter-sentence coherence could not be achieved with the current implementation. However, our Fiction pre-trained model reached a good level of cohesion in comparison with a common sense model, a similar model to our Commonsense model has shown to have a good performance after a fine-tuning process. We conclude that our model has a good fluency compared to the state of the art.

In respect of the particular objectives, we build a fictional dataset, we propose a methodology for automatic generation of fiction stories. We also trained a GPT-2 model with the fictional dataset and got interesting results.

With the fictional model we reach a stage where almost every sentence is restricted to a theme, in this case, astronomy and science fiction. The decrease of perplexity has shown an increase in cohesion, however the distinction between the sentences generated also decreases. Even considering the computation and time costs the results achieved with pre-trained language models are above the results obtained for other works with a conexionist approach. We also conclude that a hybrid approach is a great perspective in order to solve this task.

7.2 Future work

As future work we suggest the implementation of our methodology proposed. With activities as the following:

- To fine-tune the Fiction pre-trained model
- To implement the AVM and BCO generator
- To build the stories dataset.
- To fine-tune a masked pre-trained language model
- To implement the interpolator
- To link the processes into the framework proposed

Considering the results we also propose to use only stories within a specific world in the same dataset. Then to train autoregressive pre-trained models in order to have a various models, by doing so, we would have a model trained specifically for every world.

As a long-sighted work in mind, there is the research in order to set universal automatic metrics for the evaluation of coherence, cohesion, interestingness and consistency.

Bibliography

[1] S. R. Turner, *A computer model of creativity and storytelling.* PhD thesis, University of California, Computer Science Department, 1993.

[2] R. Pérez y Pérez and M. Sharples, "Mexica: A computer model of a cognitive account of creative writing.," *Journal of Experimental & Theoretical Artificial Intelligence 13*, 2001.

[3] J. Guan, F. Huang, Z. Zhao, X. Zhu, and M. Huang, "A knowledge-enhanced pretraining model for commonsense story generation," *arXiv preprint arXiv:2001.05139*, 2020.

[4] P. Xu, M. Patwary, M. Shoeybi, R. Puri, P. Fung, A. Anandkumar, and B. Catanzaro, "Megatron-cntrl: Controllable story generation with external knowledge using large-scale language models," *arXiv preprint arXiv:2010.00840*, 2020.

[5] S. Wang, G. Durrett, and K. Erk, "Narrative interpolation for generating and understanding stories," *arXiv preprint arXiv:2008.07466*, 2020.

[6] W. Wang, P. Li, and H. T. Zheng, "Consistency and coherency enhanced story generation," *arXiv preprint arXiv:2010.08822*, 2020.

[7] M. Rukeyser, *The Collected Poems of Muriel Rukeyser.* Pittsburgh: University of Pittsburgh Press, 2005.

[8] J. Meister, "Narratology." http://www.lhn.uni-hamburg.de/node/48.html, 2014. Accessed on 2021-06-17.

[9] G. Polti, *Les 36 situations dramatiques: Maîtriser l'art narratif grâce à l'exploration des principes dramatiques.* Paris: Editions du Mercure de France, 1895.

[10] M. Figgis, *The Thirty-Six Dramatic Situations.* England: Faber & Faber, 2017.

[11] I. Mani, "Computational narratology." http://www.lhn.uni-hamburg.de/node/43.html, 2014. Accessed on 2021-06-17.

[12] M. Minsky, *Psychology of Computer Vision*. New York: McGraw Hill, 1975.

[13] H. Calvo, *DETERMINACIÓN AUTOMÁTICA DE ROLES SEMÁNTICOS USANDO PREFERENCIAS DE SELECCIÓN SOBRE CORPUS MUY GRANDES*. PhD thesis, Computing Research Center (CIC-IPN), 2006.

[14] J. R. Meehan, "Tale-spin, an interactive program that writes stories," *5th international joint conference on Artificial intelligence*, 1977.

[15] M. O. Riedl and R. M. Young, "Narrative planning: Balancing plot and character," *Journal of Artificial Intelligence Research 39*, 2010.

[16] M. J., T. M., S. B., R. A. A., M. M., and W.-F. N., "Social story worlds with comme il faut.," *IEEE Trans. Comput. Intellig. and AI in Games*, 2014.

[17] A. Daza, H. Calvo, and J. Figueroa-Nazuno, "Automatic text generation by learning from literary structures.," *Proceedings of the Fifth Workshop on Computational Linguistics for Literature*, 2016.

[18] M. Lebowitz, "Planning stories," *9th Annual Conference of the Cognitive Science Society*, 1987.

[19] J. Porteous and M. Cavazza, "Controlling narrative generation with planning trajectories: The role of constraints," *Joint International Conference on Interactive Digital Storytelling*, 2009.

[20] K. Onodera, T. Akimoto, and T. Ogata, "A state-event transformation mechanism for generating micro structures of story in an integrated narrative generation system.," *Proceedings of the Annual Meeting of the Cognitive Science Society 34*, 2012.

[21] L. C. and G. P., "Creativity in story generation from the ground up: Nondeterministic simulation driven by narrative," *5th International Conference on Computational Creativity*, 2014.

[22] A. B. *et al.*, "Generating children's stories from character and event models," *International Workshop on Multi-disciplinary Trends in Artificial Intelligence*, 2017.

[23] P. Jain, P. Agrawal, and M. L. A. S. K. Mishra, A.; Sukhwani, "Story generation from sequence of independent short description," *SIGKDD Workshop on Machine Learning for Creativity (ML4Creativity)*, 2017.

[24] E. Clark, Y. Ji, and N. A. Smith, "Neural text generation in stories using entity representations as context," *NAACLHLT*, 2018.

[25] A. Fan, M. Lewis, and Y. Dauphin, "Hierarchical neural story generation," *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, 2018.

[26] N. Peng, M. Ghazvininejad, J. May, and K. Knight, "Towards controllable story generation," *Proceedings of the First Workshop on Storytelling*, 2018.

[27] L. Yao, N. Peng, R. Weischedel, K. Knight, D. Zhao, and R. Yan, "Plan-and-write: Towards better automatic storytelling," *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligences*, 2019.

[28] A. Fan, M. Lewis, and Y. Dauphin, "Strategies for structuring story generation," *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019.

[29] M. Roemmele and G. A., "Automated assistance for creative writing with an rnn language model," *Proceedings of the 23rd International Conference on Intelligent User Interfaces Companion*, 2018.

[30] Z. Li, X. Ding, and T. Liu, "Generating reasonable and diversified story ending using sequence to sequence model with adversarial training," *COLING*, 2018.

[31] F. Luo, D. D., Y. P., T. Liu, B. Chang, Z. Sui, and X. Sun, "Learning to control the fine-grained sentiment for story ending generation," *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019.

[32] J. Guan, Y. Wang, and M. Huang, "Story ending generation with incremental encoding and commonsense knowledge," *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019.

[33] L. J. Martin, P. Ammanabrolu, X. Wang, W. Hancock, S. Singh, B. Harrison, and M. O. Riedl, "Event representations for automated story generation with deep neural nets," *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[34] J. Xu, X. Ren, Y. Zhang, Q. Zeng, X. Cai, and X. Sun, "A skeleton-based model for promoting coherence among sentences in narrative story generation," *EMNLP*, 2018.

[35] P. Tambwekar, M. Dhuliawala, L. J. Martin, A. Mehta, B. Harrison, and M. O. Riedl, "Controllable neural story plot generation via reward shaping," *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, 2019.

[36] P. Ammanabrolu, E. Tien, W. Cheung, Z. Luo, W. Ma, L. J. Martin, and M. O. Riedl, "Story realization: Expanding plot events into sentences," *arXiv preprint arXiv:1909.03480*, 2019.

[37] T. B. Hashimoto, K. Guu, Y. Oren, and P. Liang, "A retrieve-and-edit framework for predicting structured outputs," *32nd Conference on Neural Information Processing Systems*, 2018.

[38] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, pp. 5998–6008, 2017.

[39] D. J., M. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2019.

[40] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and Sutskever, "Language models are unsupervised multitask learners," *OpenAI Blog*, 2019.

[41] T. B. Brown *et al.*, "Language models are few-shot learners," *arXiv preprint arXiv:2005.14165*, 2020.

[42] Y. Liu *et al.*, "Roberta: A robustly optimized bert pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.

[43] N. Kitaev, K. Łukasz, and A. Levskaya, "Reformer: The efficient transformer," *arXiv preprint arXiv:2001.04451*, 2020.

[44] I. Beltagy, M. E. Peters, and A. Cohan, "Longformer: The long-document transformer," *arXiv preprint arXiv:2004.05150*, 2020.

[45] S. Wang, B. Li, M. Khabsa, H. Fang, and H. Ma, "Linformer: Self-attention with linear complexity," *arXiv preprint arXiv:2006.04768*, 2020.

[46] M. Zaheer *et al.*, "Big bird: Transformers for longer sequences," *arXiv preprint arXiv:2007.14062*, 2020.

[47] K. Choromanski *et al.*, "Rethinking attention with performers," *arXiv preprint arXiv:2009.14794*, 2020.

[48] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter," *arXiv preprint arXiv:1910.01108*, 2019.

[49] S. R. Turner, *The Creative Process: A Computer Model of Storytelling.* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1994.

[50] R. Pérez y Pérez and M. Sharples, "Three computer-based models of storytelling: Brutus, minstrel and mexica," *Knowledge-based systems*, vol. 17, no. 1, pp. 15–29, 2004.

[51] R. Pérez y Pérez, "From mexica to mexica-impro: The evolution of a computer model for plot generation," in *Computational Creativity Research: Towards Creative Machines* (T. R. Besold, M. Schorlemmer, and A. Smaill, eds.), ch. 13, pp. 267–284, Germany: Atlantis Press, 2015.

[52] L. Pemberton, "A modular approach to story generation," *Fourth Conference of the European Chapter of the Association for Computational Linguistics*, 1989.

[53] M. Shoeybi, M. Patwary, R. Puri, P. LeGresley, J. Casper, and B. Catanzaro, "Megatron-lm: Training multi-billion parameter language models using model parallelism," *arXiv preprint arXiv:1909.08053*, 2019.