Natural Language Generation and Creative Writing A Systematic Review

Abdulla Alsharhan

Faculty of Engineering & IT, The British University in Dubai, United Arab Emirates

Email: alsharhan@outlook.com

Abstract

Among studies on natural language generation (NLG), computational creativity, and human-computer interaction; there is a vision of witnessing these tools collaborating with humans in generating and authoring creative content. This study aims to systematically review published studies discussing creative writing and story generation during the period of 2016-2021. This work seeks to identify the primary research methods used in NLG and creative writing studies, to locate how these studies are distributed geographically, and finally, to classify the subfields or common keywords primarily used in NLG involving creative writing. The findings suggest that experiment studies and problem-solving were the most common research methods in NLG and creative writing. Major identified themes in the reviewed articles include story generation, language models, and co-creativity, along with some gaps in foreign language translation and humour generation studies. The majority of the studies suggest that NLG tasks had a positive impact on creative writing. Common tasks related to NLG and creative writing are typically using keywords such as story generation, co-creativity, co-writing, user interface and writing tools. In future work, we aim to explore more GPT-3 capabilities in creative writing, in addition to creative writing applications in foreign language translation tasks.

Keywords:

Natural Language Processing; Natural Language Generation; Creative writing, Story Generation; language model; Co-Creativity; Co-writing; Poem Generation; writing tools; Computational creativity; Systematic Literature Review

1. Introduction

This study was initiated with the question: how far are we from asking a machine to write this study by itself? Or at least, to minimize the effort needed to complete a decent article? The development of automatic text generation techniques in the past decade represents an interesting twist to this question. Philip M Parker wrote over than 200,000 books that are accessible on amazon.com. Parker did not write all these titles manually. Instead, he created a computer algorithm that collected all the required data from publicly-available online resources, processed them in a natural language, and compiled them into books. This made him “the most published author in the history of the planet”, as he describes himself in his own words [1].

Doi: https://doi.org/10.54216/IJAACI.010105
Received: January 14, 2022 Accepted: May 26, 2022

Natural Language Processing (NLP) is rapidly-developing discipline; and many NLP application being developed on a daily basis, including virtual newspapers from sensor data, soccer reports, environmental concerns, weather reports, summaries of patient information in clinical contexts, and persuasive text [2]. One particular area of NLP is Natural Language Generation (NLG), which is attracting increasing attention because of its ability to automatically create complex personalisation and transform textual data sets to understandable narratives.

The development of NLG started to become noticed in the mid-to-late 1990s, where most NLG applications have emerged, including the first software houses specialising in the development of NLG technology [3]. The importance of NLG stems from the increasing demand for human interaction with technology. This relationship between humans and technology has formed the need for machines to generate language, instead of just working on understanding natural human language [4]. Some NLG applications include machine translation, text simplification, summarisation, automatic generation of paraphrases, and questions [2].

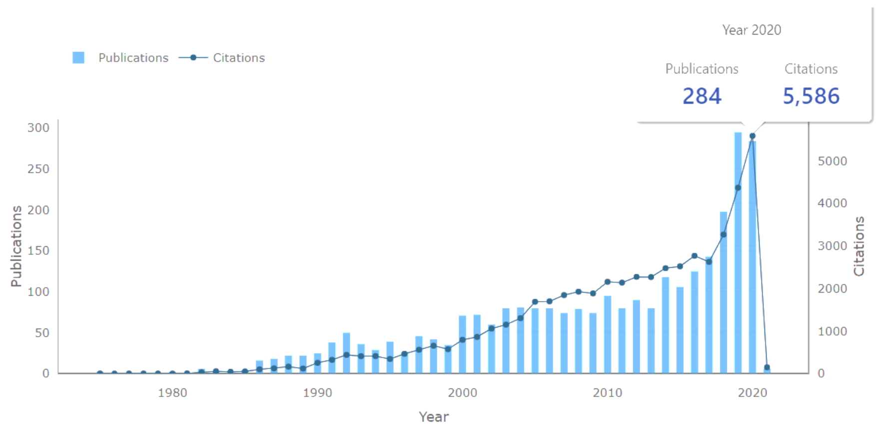

Figure 1: NLG Publications & Citations Over Time (Microsoft Academic 2020).

NLG is expected to experience a significant growth rate at a CAGR over the forecast period from 2019 to 2026. According to Grand View Research (2019), the global NLG market was valued at USD 336.2 in 2018, and is expected to reach USD 1.2 billion By 2025, with a CACR of 19.8% from 2019 to 2025.

Among NLG, computational creativity, and human-computer interaction studies, there is a vision in seeing these tools collaborate with humans in authoring creative content [6]. The quality of the natural language generated is advancing rapidly, and is believed to go beyond grammar-checking and spell-checking, such as suggesting content to spark new ideas. The possibility of machines generating creative writing is becoming more apparent with time [7].

Existing research on the ability of machines to generate creative writing has focused on several perspectives with respect to NLG tasks. Various NLG and automatic content generation tasks were carried out. It is believed that each of these studies provides a valuable part to the synthesis of NLG generation processes, yet further analysis is required, depending on other research perspectives. Existing reviews neglected the examination of NLG studies with regards to creative writing. Accordingly, this work aims to systematically review and synthesise the NLG studies

Doi: https://doi.org/10.54216/IJAACI.010105
Received: January 14, 2022 Accepted: May 26, 2022

related to creative writing, in order to offer a comprehensive analysis of the collected studies. More specifically, this study poses the five following research questions:

- **RQ1:** What are the main NLG tasks investigated, considering their relationship with creative writing?
- **RQ2:** What are the main research methods used in the collected studies?
- **RQ3:** What is the impact of creative computing on creative writing?
- **RQ4:** How are the NLG studies that consider creative writing distributed across countries of implementation and year of publication?
- **RQ5:** what are the subfields of NLG studies involving creative writing, and what are the types of methods used in the collected studies?

2. Literature review

2.1 Natural Language Generation

Natural Language Generation (NLG) has attracted a considerable amount of attention from many researchers. NLG is usually classified as a sub-field of artificial intelligence and computational linguistics, and can be defined as the field concerned with constructing computer systems that can produce understandable texts in human languages from a non-linguistic representation of information [8].

Common NLG tasks can vary from content determination to referring expression generation. Figure 1 shows the five major NLG tasks [9].

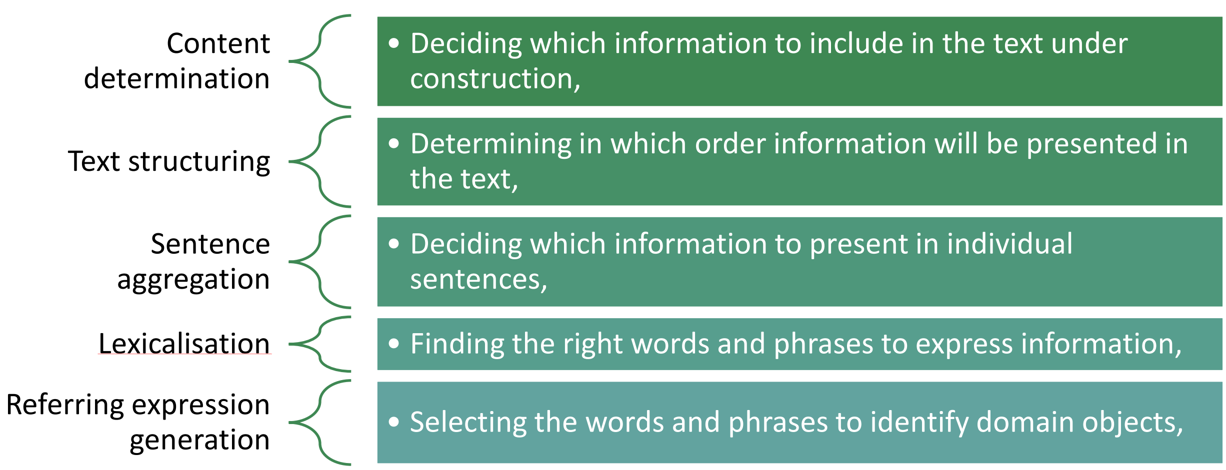

Figure 1:The main NLG tasks.

Doi: https://doi.org/10.54216/IJAACI.010105
Received: January 14, 2022 Accepted: May 26, 2022

2.2 NLG Applications

NLG applications cover a wide range of functions, starting from machine translations, up to the generation of paraphrased content. Recent advancements with large-scale pre-trained language modes such as OpenAI GPT-3 and BERT (Devlin et al. 2018 in Yang & Tiddi 2020) have demonstrated the ability of machines to generate a paragraph of understandable text based on a given topic [10]. Figure 2 shows some prominent uses for NLG applications [11].

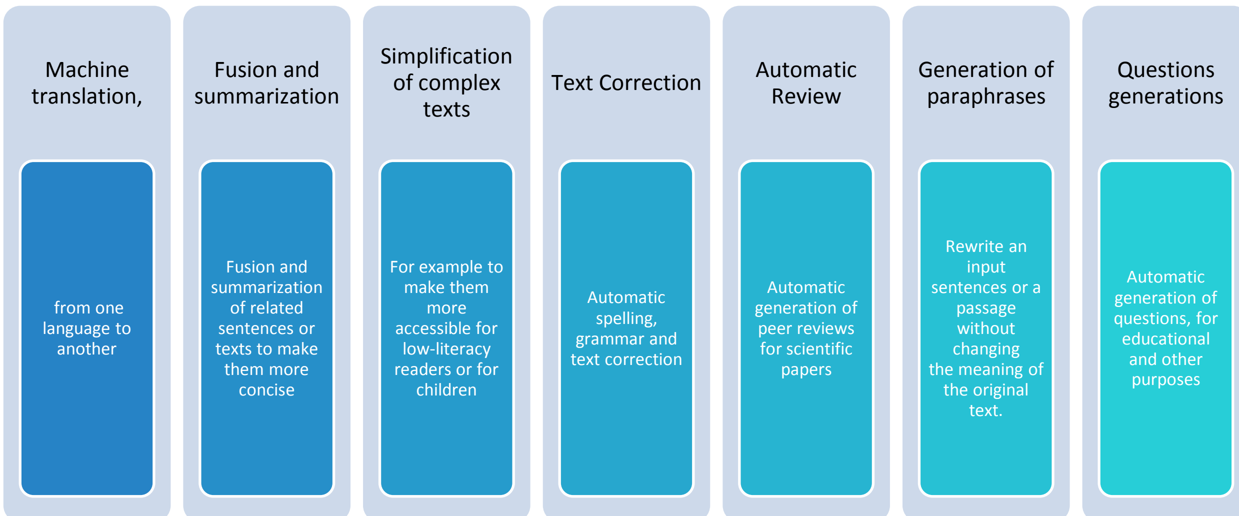

Figure 2: NLG Applications.

According to a recent state-of-the-art survey [2], future NLG applications can cover social media feeds summaries, situated language generation, corporate and financial reports, press applications, medical applications, sport applications, educational applications, as well as entertainment and media applications.

2.3 Natural Language Generation and Creative Writing

Oxford English Dictionary defines creative writing as the opposite of academic or editorial writing. Creative writing is often associated with writing fiction or poetry, which demonstrate some type of fantasy or novelty. The intersection of NLG with creative writing is responsible for many debates among researchers. One school of thought claims the ability of machines to generate creative content can save hundreds of hours of work, especially with applications of creative writing in the media and videogame sectors. Games such as "Lord of the Rings", "Elder Scrolls V: Skyrim", "Fallout games", and "Final Fantasy" often have players spending hundreds of hours exploring the world. It can take a tremendous amount of creative work to create a believable world and its associated themes [12]. Popular sports games such as football and basketball might have a journalist covering them, but what about other sports? Companies like "Narrative Science" automatically generate sports reports for less popular sports and games. "Automated Insights" even generates reports based on user-provided `fantasy football' data [2].

Doi: https://doi.org/10.54216/IJAACI.010105
Received: January 14, 2022 Accepted: May 26, 2022

Another group of scholars claims that this can be a threat to human authenticity and millions of creative jobs worldwide. A yearly competition like NaNoGenMo challenges writers and programmers to generate a novel programmatically [13]. At the same time, in Germany, the Stuttgarter Zeitung's AI-augmented reporting on air pollution recently won a journalism award. Some describe it as "the Kasparov moment," associating the victory to the moment a supercomputer defeated the chess grandmaster Gary Kasparov [14].

In contrast, some scholars believe that creative language generation is still a difficult task for natural language processing systems to handle. Gatt and Krahmer (2018) argue that the recent successes of using deep learning techniques for NLG do not offer decent results, as they define good writers from a different perspective. Good writers do not present their ideas in a coherent and organized text only, but they also successfully attract the attention of readers using their skilled narration techniques. An example is occasionally surprising the reader with creative language, small jokes and well-placed metaphors, within this context. Gatt and Krahmer (2018) describe the current automatically-generated creative writing to be somewhat boring and repetitive.

Some researchers have reached a middle ground, where they see machines and humans working together to enhance creative output. Davis (2013) (in Manjavacas et al. 2017) defines co-creativity as a collaborative process between several agencies. In this context, Davis sees co-creativity as a blending improvisational force. Developing a co-creative computational NLG system allows researchers to include the human factor as an essential input to the NLG system workflow. For example, applications like "Writefull" can check the language for you and browse through sentences from millions of published papers to compare with. It offers a “sentence palette” to browse possible template sentences for your article [16]. Other applications like "AI Writer" can generate an entire draft for you based on a headline. It can also provide a powerful paraphrasing tool to ensure the authenticity of your work [17].

Despite the split between the researchers' group on NLG advancement in creative writing, little is known on the relationship between NLG tasks and creative writing. This study aims to identify the primary research methods used in NLG and creative writing studies, to identify these studies' outcomes, to determine how these studies are distributed geographically, to classify the subfields mainly used in NLG tasks related to creative writing, and finally, to identify any gaps for future research opportunities.

1. Research Methodology

This study aims to carry out an in-depth systematic literature review (Hustadt 2016). According to Dumay and Cai (2014), writing a systematic literature review includes the following six steps: 1) identifying the research questions; 2) writing the research protocol; 3) determining related articles and searching the literature; 4) developing a coding framework; 5) investigating quality appraisal; and 6) conducting data analyses and discussing the results obtained.

This detailed systematic review is based on results from research papers derived from various online journals and databases that focus on NLP and NLG performance, as well as creative writing. The screened papers' time frame was set to be within the last five years (between 2016 and 2021), excluding books chapters and non-computer science contexts.

Doi: https://doi.org/10.54216/IJAACI.010105
Received: January 14, 2022 Accepted: May 26, 2022

3.1 Data sources and search strategies

Different articles were found to match the search criteria. In particular, 78 materials were found from credible online journals. All the results were found using WorldCat.org, Google Scholar, ProQuest, and Science Direct.

The search algorithms included searching for the presence of the required elements in the research objectives. The exclusion aided to sort the sources from undesirable and non-organizational contexts, which side-lined to leave only the required data. The keywords used to locate information from the journals were KW: "natural language generation" AND "creative writing" AND "natural language processing".

3.2 Inclusion and exclusion criteria

Table 1 shows more insights on the inclusion and exclusion criteria, and how articles were selected based on these criteria.

Table 1:Inclusion and exclusion criteria.

Inclusion	Exclusion
Articles published in 2016 and above	Articles published before 2016
Should be related to NLG methodologies, processes, or lifecycle	Related to NLG but not linked to Natural Language Processing or Computer Science fields
should be related to creative writing Techniques	Related to creative writing but not linked to NLG or Natural Language Processing
Should be in the field of natural language processing	Not related to the field of NLG or Computer science
Papers are in English	Papers are not in English
Journal Articles	Books, theses or book chapters
Open access articles	Articles which are not accessible

3.3 Selection of papers

This review is based on the proposed reporting model [19] referred to as the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA). The selected papers were extracted from Google scholar and Science Direct using Perish [20]. Additional results from WorldCat, Science Direct and ProQuest were added manually to an extended spreadsheet. All duplicates were removed, followed by screening the titles and abstracts. Any paper that did not meet the inclusion criteria in Table 1 was removed. The remaining filtered papers were fully screened to ensure their content is relevant to this study. The final filtered list included 16 papers. The following figure shows the steps taken to reach the final list of papers.

Doi: https://doi.org/10.54216/IJAACI.010105
Received: January 14, 2022 Accepted: May 26, 2022

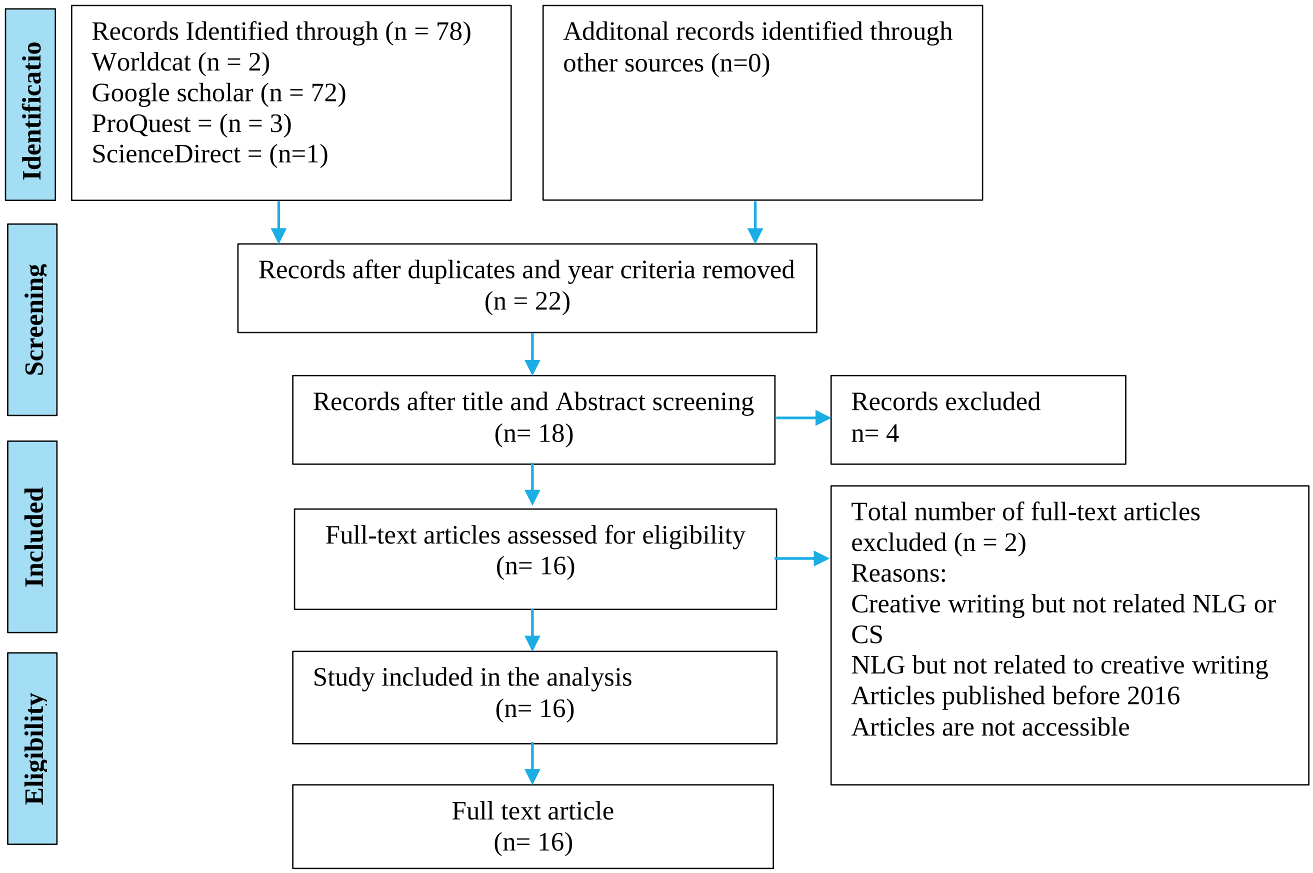

Figure 3 Research PRISMA model

3.4 Quality assessment

It is common to carry out a quality assessment procedure to indicate how right the conclusions are. In this study, we selected five questions for quality assessment to evaluate each study's quality. As for each answer value, three values indicate the answer to each question. The "Yes" answers to questions are marked as (1), the "No" answers to questions are marked as (0), and the partially-answered questions are marked as (0.5).

Table 2: Quality assessment checklist.

#	**Question**
١	Are the research aims clearly specified?
2	Is the data method collection explicitly detailed?
3	Does the paper providing clear purpose, methods and findings?
4	Is the paper a published journal article or a conference proceeding?
5	Does the study add to your knowledge and understanding?

Doi: https://doi.org/10.54216/IJAACI.010105
Received: January 14, 2022 Accepted: May 26, 2022

Table 3: Quality assessment result.

Source	Q1	Q2	Q3	Q4	Q5	Total	%
R01	1	0	0.5	1	1	3.5	70%
R02	1	1	1	1	1	5	100%
R03	1	1	1	1	1	5	100%
R04	1	1	1	1	1	5	100%
R05	1	1	1	1	1	5	100%
R06	1	1	1	1	1	5	100%
R07	1	1	1	1	1	5	100%
R08	1	0	1	1	1	4	80%
R09	1	1	1	1	1	5	100%
R10	1	1	1	1	1	5	100%
R11	1	1	1	1	1	5	100%
R12	1	1	1	1	1	5	100%
R13	1	1	1	1	1	5	100%
R14	1	1	1	1	1	5	100%
R15	1	1	1	1	1	5	100%
R16	1	1	1	1	1	5	100%

3.5 Data Analysis and Coding Framework

This section describes the data analysis system used in this work, which takes a formal and systematic approach, emphasising the use of data visualisation and analysis tools. Sixteen papers fulfilled all of the above selection and screening criteria. As shown in Table 4, some properties of the selected works are listed below. In addition, the table shows (a) the main NLG task, (b) research method used, (c) main findings, (d) outcome or results, and (e) country. The relationships discussed in this work result from a combination of two factors: the relationship between NLG and creative writing.

Table 4: Synthesis matrix

#	Ref.	The main NLG task	Methods	Main finding	Outcome	Country

Doi: https://doi.org/10.54216/IJAACI.010105
Received: January 14, 2022 Accepted: May 26, 2022

R01	Trăuşan-Matu (2019)	To analyze some of the differences between NLP techniques and to discuss the limits of NLG.	Comparative	Deep learning for NLG did not offer good results, and AI lacks empathy. Real creativity by AI is still out of reach. Polyphonic model and chronotopical perspectives can solve pace-time in NLG.	Negative	Romania
R02	Yang & Tiddi (2020)	To study how to combine knowledge graphs with language models and build a creative story generation system named DICE.	Dev. &Application	Injecting entities and relations from knowledge graphs into the generated stories effectively.	Positive	Netherlands
R03	Chakrabarty, Muresan & Peng (2020)	Generating a simile starting from a literal utterance that contains the TOPIC, EVENT and PROPERTY.	Problem-solving	The first work in attempting to generate similes. Their approach generates 88% novel similes that do not share properties with the training data.	Positive	United Sates
R04	Lampridis, Kefalas & Tzallas (2020)	Implementing lyric generation machine learning models in the Greek language for the genre of Éntekhno music.	Problem-solving	A model that can generate high quality lyrics and difficult to distinguishable from actual lyrics.	Positive	Greece
R05	Calderwood et al. (2020)	To explore how fiction writers use generative language models during their writing processes.	Exploratory	Identified six criteria for evaluating creative writing assistants, and proposed design guidelines for future co-writing tools.	Positive	United Sates
R06	Ramos, Monteiro & Paraboni	Personality-dependent content selection in NLG.	Experiment	Among the first to establish a relation between personality and content selection.	Positive	Brazil

Doi: https://doi.org/10.54216/IJAACI.010105
Received: January 14, 2022 Accepted: May 26, 2022

R07	Alexi, Papathanasopoulou & Papakitsos (2016)	Morphological generator.	Descriptive	Crucial research critera for selecting implementation mode are maximizing grammatical production while minimizing computational complexity.	Positive	Greece
R08	Booten 2019	Introducing an interactive writing system designed to notice when writers resort to expected language and encourage them to avoid these linguistic elements.	Problem-solving	Progym pushed writers away from writing expected language. Though, the different versions led to different styles of revision.	Positive	United Sates
R09	Siddharthan et al. (2019)	Transforms raw telemetric data into engaging and informative blog texts that are readily understood by all.	Experiment	Data-to-text technology is sufficiently advanced to achieve more than just factual summarization of data for professional use, opening up new areas for using AI to engage people through data.	Positive	United Kingdom
R10	Elkins & Chun (2020)	To determine whether GPT-3 (Open AI Project) can pass a writer's Turing test.	Experiment	GPT-3 can create realistic yet surprising plots, and recreate key stylistic and thematic traits of an author in just a few lines, but it can sometimes fail at the simplest of linguistic tasks.	Neutral	United Sates
R11	Mcgovern & Scott (2016)	A tool for Automatic Poetry Generation.	Dev. &Application	An algorithm to mine syllable and rhyme structure from a provided set of poems to generate poetry with a high classification accuracy.	Positive	United Sates

Doi: https://doi.org/10.54216/IJAACI.010105
Received: January 14, 2022 Accepted: May 26, 2022

R12	WILKE & BEROV (2018)	Implementation of a Function Unit based plot analysis in the above-mentioned storytelling system.	Case study	A step towards more general intelligence by allowing a system that was formerly only capable of expression-type generative acts to also perform aesthetic- and framing-type generative acts.	Positive	Germany
R13	Roemmele & Gordon (2018)	Examine an emerging NLP application that supports creative writing by automatically suggesting continuing sentences in a story.	Experiment	Authors relied heavily on the suggestions, as they often chose to keep most of the suggestions, with fewer edits.	Positive	United Sates
R14	Kutlak, van Deemter & Mellish (2016)	Production of Referring Expressions for an Unknown Audience	Problem-solving	The new model excels in terms of the proportion of correctly identified entities and in terms of the perceived quality of the generated descriptions.	Positive	United Kingdom
R15	Ippolito et al. (2019)	Unsupervised Hierarchical Story Infilling	Problem-solving	Taking a hierarchical approach to story infilling is an effective strategy for balancing fluent and coherent generated text with the diversity and interestingness necessary to build a useful tool for writers.	Positive	United Sates
R16	Manjavacas et al. (2017)	Describes a co-creative text generation system applied within a science fiction setting to be used by an established novelist.	Experiment	An applied text generation system and graphical user interface, which together facilitate a co-creative environment in which to write science fiction literature.	Neutral	Netherlands

4. RESULTS

This symmetric review covered sixteen screened papers that discuss NLG and creative writing, and were found in Google Scholar, Science Direct, ProQuest and WorldCat between 2016 and 2021. The following are the answers to the research questions raised at the beginning of this review.

Doi: https://doi.org/10.54216/IJAACI.010105
Received: January 14, 2022 Accepted: May 26, 2022

RQ1: what are the main NLG tasks studied, considering their relationship with creative writing?

Several research studies that focus on creative writing and co-creativity were carried out. The following table shows the relationships identified in each selected study. Most studies focus on co-writing and providing writing assistance rather than generating the whole text from scratch (n = 5). Language and computational models were also the main focus of other studies (n = 5). Several studies focused on story generation (n = 3), poetry generation (n = 2) and machine learning (n = 2). The remaining studies covered some important NLG tasks, including simile generation, personality traits, world predication and audience design.

Table 5: Analysis of NLP research with regards to creative writing.

	Source	**Story Generation**	**Poetry generation**	**Co-writing**	**Language models**	**Machine Learning**	**simile generation**	**Personality traits**	**Data-to-Text**	**Word Predication**	**Audience Design**
R01	Trăuşan-Matu (2019)	X			X						
R02	Yang & Tiddi (2020)	X			X						
R03	Chakrabarty (2020)						X				
R04	O. Lampridis et al. (2020)		X			X					
R05	Calderwood et al. (2020)			X							
R06	Ramos et al. (2020)				X			X			
R07	Alexi et al. (2016)										
R08	Booten 2019			X							
R09	Siddharthan et al. (2019)	X							X		
R10	Elkins & Chun (2020)					X					
R11	McGovern (2016)		X								
R12	Wilke & Berov (2019)				X						
R13	Roemmele & Gordon (2018)			X							
R14	Kutlak et al. (2016)			X	X						X
R15	Ippolito et al. (2019)									X	
R16	Manjavacas et al. (2017)			X							

Doi: https://doi.org/10.54216/IJAACI.010105
Received: January 14, 2022 Accepted: May 26, 2022

RQ2: what are the main research methods and outcomes addressed by the collected studies?

Distribution of research methods

Figure 2 indicates that two-thirds of the analysed studies were mainly experiment studies (n = 5) and problem-solving studies (n = 5). The 3rd common research method involved development and application studies (n = 2). Most studies related to NLG involving the experimental development of novel solutions used experiments and problem-solving methods.

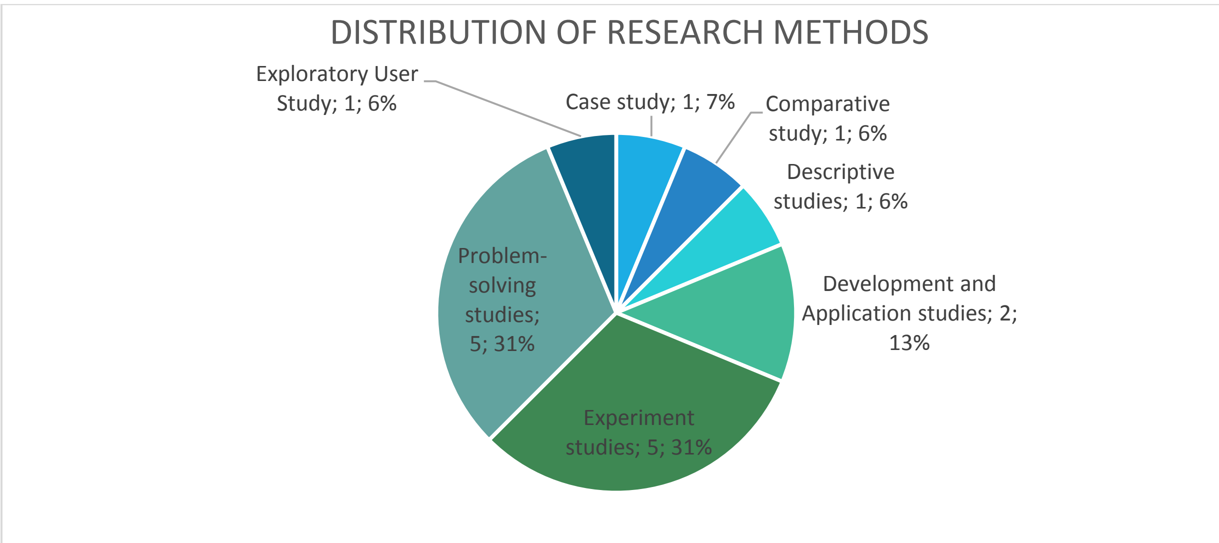

Figure 4: Distribution of research methods.

RQ3: What is the impact of creative computing on creative writing?

The chart in Figure 3 shows that 75% of the analysed studies suggested that NLG tasks had a positive impact on creative writing (n = 12), while only one study suggested NLG tasks had a negative impact on creativity (6%), followed by three studies with a neutral outcome (19%). The chart in Figure 4 reveals that most studies associated with positive outcomes are experiment and problem-solving studies. In contrast, the only method associated with negative outcomes is the comparative study method by Trăuşan-Matu (2019). The study by Calderwood et al. (2020) was neutral.

Doi: https://doi.org/10.54216/IJAACI.010105
Received: January 14, 2022 Accepted: May 26, 2022

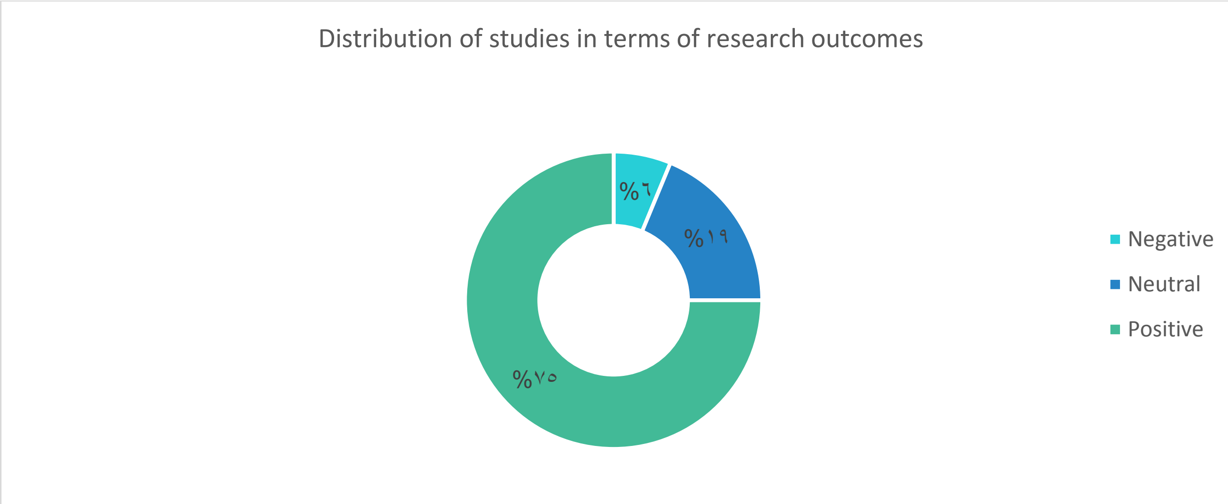

Figure 5: Distribution of studies in terms of research outcomes.

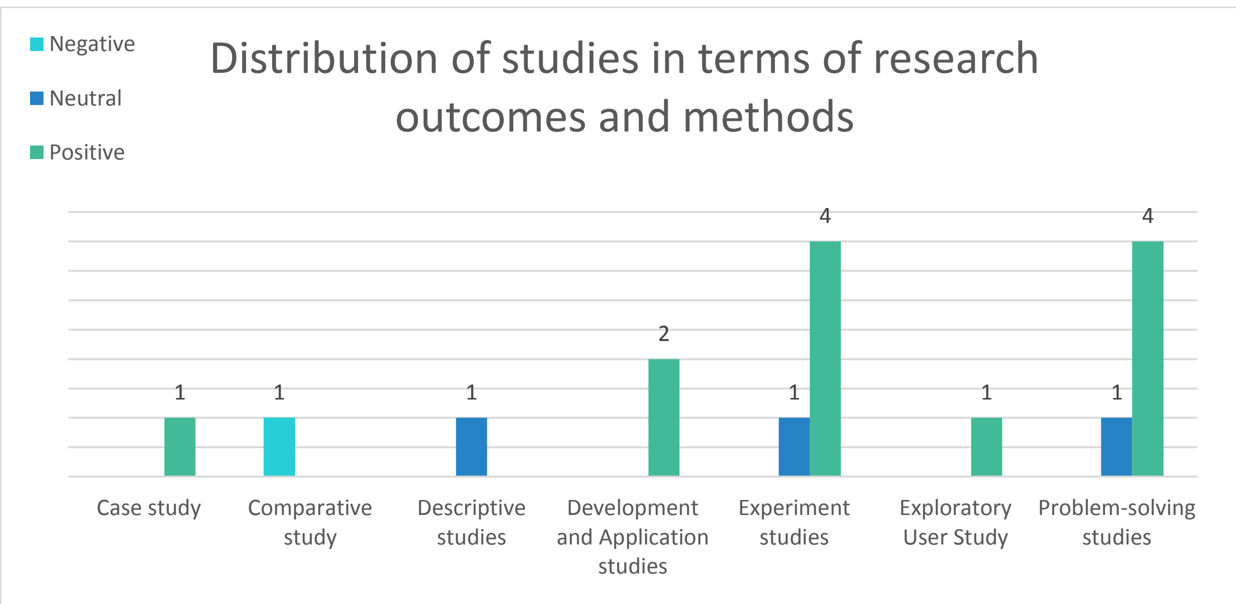

Figure 6: Distribution of studies in terms of research outcomes and methods

RQ4: How are the NLG studies considering creative writing distributed by countries and by year of publication?

Distribution of studies with regards to their country of implementation

Figure 3 demonstrates the distribution of the sixteen analysed articles by country in which these studies were carried out. Most of these studies were conducted in the United States (n =7), followed by the United Kingdom, the Netherlands and Greece (n=2), with two studies for each. Finally, Germany, Brazil and Romania (n=1) involved one study each.

Doi: https://doi.org/10.54216/IJAACI.010105
Received: January 14, 2022 Accepted: May 26, 2022

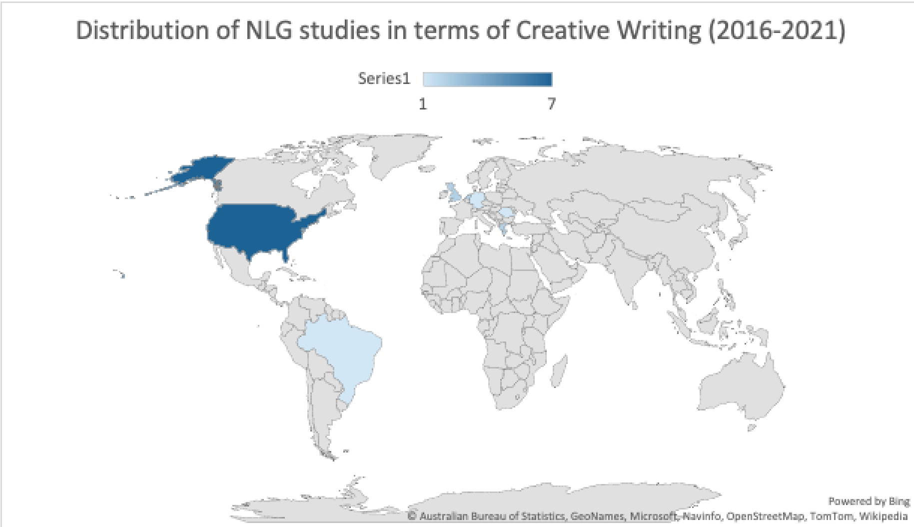

Figure 7: Distribution of studies with regard to their country of implementation.

Distribution of studies based on their year of publication

The number of published articles has increased from three studies in 2016 to six in 2020. When it comes to publication year, Figure 4 describes the distribution of the articles analysed by the year in which they were published. The studies cover the period 2016-2020, and the number of publications per study ranges from 1-6. It is worth noting the number of published articles increased between 2019-2020.

It is worth noting that although there was only one study published in 2018, the number of citations of that particular article [15] was remarkably high compared to the remaining published studies.

Doi: https://doi.org/10.54216/IJAACI.010105
Received: January 14, 2022 Accepted: May 26, 2022

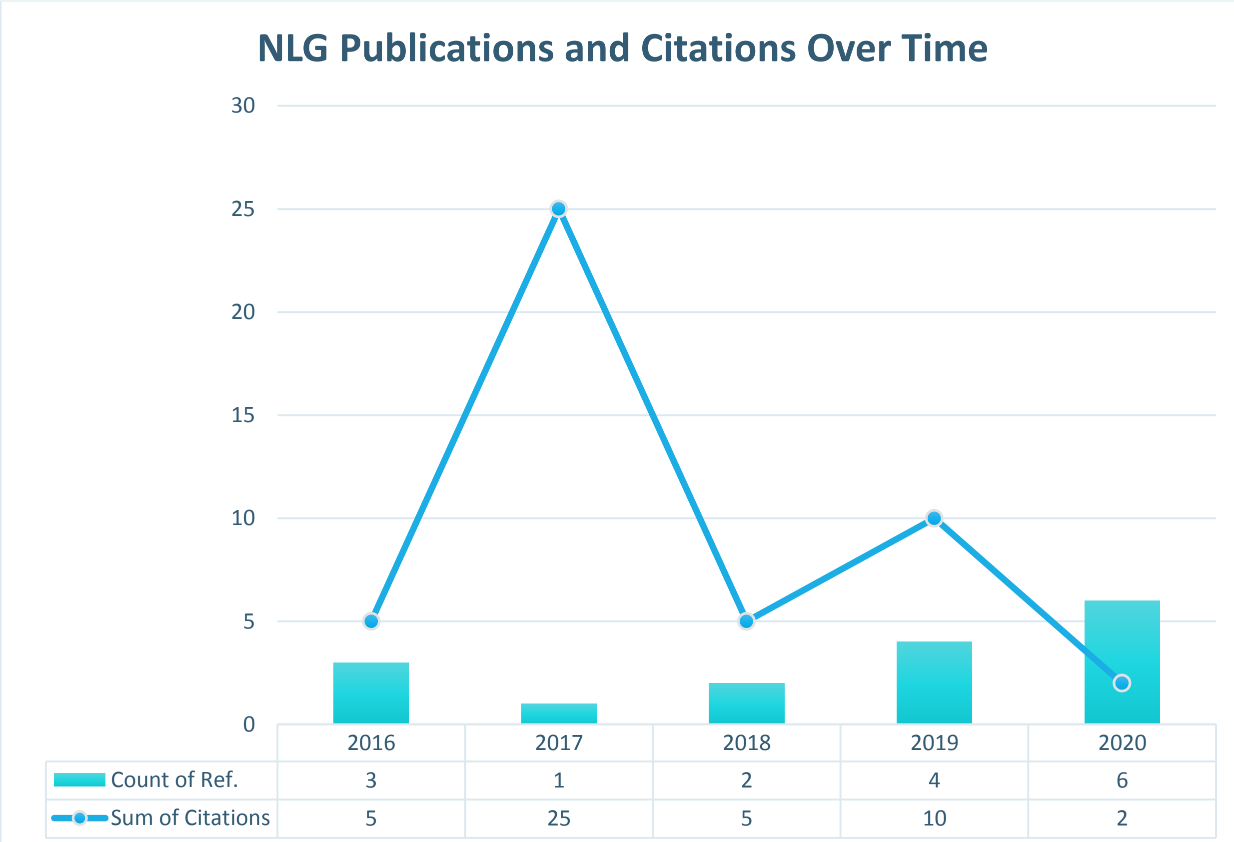

Figure 8: Distribution of studies by publication year.

RQ5: What subfields or keywords of NLG are mainly studied, and what are the types of methods used in the collected studies?

Figure 7 provides an overview of the most common terms and keywords associated with NLG and creative writing studies, as well as the type of research method used in relation to each sub-field. It can be noticed that the most common studies related to NLG and creative writing focus on co-creativity, co-writing, user interface and writing tools. These studies commonly use problem-solving, experiments and exploratory analyses.

Comparative studies were common among those related to story generation, computational models, creative writing and phenomenology. The development and application studies shared some methods with comparative studies methods, such as story generation and computational models, in addition to knowledge graphs, language models and poetry generation.

Doi: https://doi.org/10.54216/IJAACI.010105
Received: January 14, 2022 Accepted: May 26, 2022

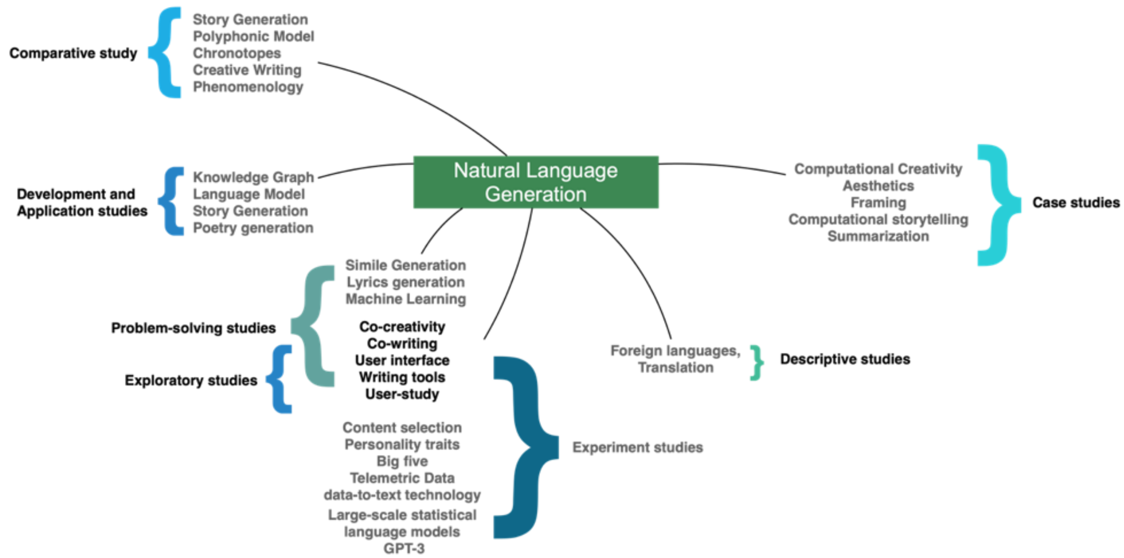

Figure 9: Mind Map of NLG, Creative Writing and methods.

2. Discussion

The results discussed many story generation techniques and state-of-the-art computational models related to language. The comparative studies by Trăuşan-Matu (2019) focused on analysing computer-based story generation techniques, presenting some of the differences between NLP techniques, and discussing the limitations of NLG. Trăuşan-Matu (2019) claims that recent successes of using deep learning techniques for NLG remain limited. According to him, Artificial Intelligence is still out of reach, as it is still missing empathy and real creativity.

Trăuşan-Matu (2019) also found that the polyphonic model and chronotopic perspectives provide a solid foundation for considering the space-time complexity in the generation of text. In contrast, the OpenAI team has released the upgraded GPT-3, with 175 billion parameters, which is 100 times larger than the previous model, GPT-2. These kinds of language models demonstrate significant text generation capabilities that can achieve impressive results, without extra training (Keskar et al. 2019, in Yang & Tiddi 2020).

The results are in line with those by Elkins and Chun (2020), where they conduct experimental studies testing GPT-3, and witnessed a remarkable score in passing the writer's Turning Test. Elkins and Chun (2020) found that GPT-3 excels in many aspects of creative writing, which an ordinary undergraduate student would find challenging. It can create realistic yet surprising plots, and recreate key stylistic and thematic traits of an author in just a few lines. Despite their findings, Elkins and Chun (2020) noted some downside of GPT-3. Some limitations of GPT-3 are maintaining a coherent argument or narrative thread over long periods, maintaining consistency of gender or personality; employing simple grammar rules, and showing basic knowledge and common-sense reasoning. These results are in line with Yang and Tiddi (2020). According to Brown et al. (2020) (in Yang & Tiddi 2020), the performance of these language models becomes limited when dealing with the long tail of rare entities, such as

Doi: https://doi.org/10.54216/IJAACI.010105
Received: January 14, 2022 Accepted: May 26, 2022

numbers and dates. In addition to that, Guan et al. (2019) (in Yang & Tiddi 2020) argues that these models perform poorly when building contextual clues, or use implicit knowledge to generate a rational ending to the stories.

Yang and Tiddi (2020) presented novel development and application studies, answering the question of how to combine knowledge graphs with language models and build a creative story generation model. The proposed model (DICE) uses external knowledge graphs to provide context clues and implicit knowledge to generate coherent and creative stories. The generated results were compared with GPT-2, because GPT-3 was not available at the time when Yang and Tiddi (2020) conducted their experiment. DICE keyword coverage scored higher than the GPT-2 baseline (96% for DICE-CW and 97% for DICE-DBCW) (73% for GPT-2, 88% for GPT-keyword-generation). As a result, According to Yang and Tiddi (2020) the major advantage of the DICE approach relies on its ability to effectively inject knowledge from knowledge graphs into the automatically generated stories by the language model. DICE can inject entities and relations from knowledge graphs into generated stories.

Siddharthan et al. (2019) handled story generation from a different perspective. This study appears to be the first to introduce a novel artificial intelligence program called 'blogging-bird'. The program transforms raw telemetric data into engaging and informative blog texts that are readily understood by all. The program composes blog posts through three different types of analysis: the first one is realistic general telemetric data about the weather, geographical locations and habitat types. The second one involves processing ecological interpretation of movement data in the context of habitat range use. The third one is the exploitation of domain knowledge programmed as a collection of parameters that help the system generate creative and imaginative text about looking for food and other social behaviours, based on environmental and geographic parameters.

Poetry generation is one of the common research fields in creative writing and NLG. Mcgovern and Scott (2016) presented a tool for automatic poetry generation, while Lampridis et al. (2020) implemented lyric generation machine learning models in the Greek language. Machine learning applications in NLG are a popular subject, and they have also appeared in research by Elkins and Chun (2020). They tested GPT-3 (Open AI based on machine learning projects) to pass the writer's Turing Test.

A significant observation by Kutlak et al. (2016) was a computational model of reference production, which is also known as a Referring Expression Generation (REG) algorithm. Early REG algorithms were components of dialogue systems. As it is significant for the generated content to know what the reader knows, it is significant to know what the useful and sufficient information to the reader is. Kutlak et al. (2016) attempted to solve this by presenting a computational model to produce a referring expression under uncertainly over the reader's or the hearer's model.

3. CONCLUSION AND FUTURE WORK

Natural Language Generation plays a key role in affecting the acceptance and implementation of various Natural Language Processing tasks. This study's main objective is to systematically review and analyse articles on NLG related to creative writing and story generation. The following five themes have been identified as significant findings among the reviewed articles: story generation, language models, co-creativity, poetry generation and

Doi: https://doi.org/10.54216/IJAACI.010105
Received: January 14, 2022 Accepted: May 26, 2022

machine learning. Since most of the NLG studies involved testing computational models, it is common to find that most research methods have mainly used experimental studies, problem-solving, and the development of novel solutions. Most studies have positive outcomes, which means quick advancement and progress in NLG tasks that focus on computational creativity. Most of these studies were conducted in the United States, followed by the United Kingdom, the Netherlands and Greece. The most common studies related to NLG and creative writing focused on co-creativity, co-writing, user interfaces and writing tools. In contrast, there is a clear gap in studies on foreign language translation.

When it comes to generating stories, Yang and Tiddi (2020) suggested that future work should aim to improve the coherency of the generated stories and enhance transitions between sentences. Ramos et al. (2020) focused their effort on personality-dependent models. They suggested that future work should focus on how personality-dependent models may affect a target user at the other end of an NLG system. On the other hand, Alexi et al. (2016) worked on a morphological generator, and suggested that future work should focus on the incorporation of semantic features in the database for dealing with the production of grammatical types of words. Besides, they studied the contribution of other theoretical models of morphology and generative lexical morphology in designing the algorithms and the database's structure. With respect to co-writing and co-creativity, Ippolito et al. (2019) presented an unsupervised hierarchical story infilling model, suggesting that insertion-based architectures are better suited for the infilling, and that the use of n-gram phrases instead of independent sub-words as conditioning is promising.

Booten (2019) also introduced an interactive writing system, suggesting that future work should focus on a more complex statistical approach to identifying clichés. Besides, detecting when users fall into familiar tropes is also important to consider. Booten (2019) also recommended shedding some light on a more complicated interface, allowing the writer to have more parameters to control the system.

Further analyses is required, depending on other research perspectives, as there is little known on the examination of some other NLG sub-fields in creative writing such as humour generation, in addition to the application of story generation, poem generation and co-writing in other sophisticated languages such as Arabic. Besides, some recent advancements in GPT-3 showed significant potential that can be explored in the context of creative writing.

This study used the right and the most common terminology that is related to NLG and creative writing. Many studies could have been included in this review if we used more subject-matter keywords, more time and open accessed papers. More quality papers and well-known researchers will be identified in future work. It will be a significant contribution to adapt to the reviewed language models in different languages. It can be assumed that this systematic review is the first step towards a much more considerable contribution to the field of NLG and creative writing.

Acknowledgement

This is part of a project conducted at the British University of Dubai. The author(s) declare(s) no form, sources of any funding used to conduct this research and no type of conflict of interest regarding the research and the aim of this article.

Doi: https://doi.org/10.54216/IJAACI.010105
Received: January 14, 2022 Accepted: May 26, 2022

References

1. Alexi, A., Papathanasopoulou, G. & Papakitsos, E. C. (2016). The Development of Morphological Generators and Related Issues: The Case of Modern Greek. *Linguistics Journal.*, vol. 10(1), pp. 227–237 [online]. [Accessed January 28, 2021]. Available at: https://eds.a.ebscohost.com/eds/detail/detail?vid=0&sid=33003dad-5eb4-4968-b4b6-e188a3a55665%40sessionmgr4008&bdata=JnNpdGU9ZWRzLWxpdmU%3d#AN=128269433&db=ufh.
2. Booten, K. (2019). Toward Digital Progymnasmata. *ICCC* , pp. 1–8 [online]. [Accessed January 28, 2021]. Available at: http://computationalcreativity.net/iccc2019/papers/iccc19-paper-50.pdf.
3. Calderwood, A., Qiu, V., Gero, K. I. & Chilton, L. B. (n.d.). How Novelists Use Generative Language Models: An Exploratory User Study. *IUI '20 Workshops*. Cagliari [online]. [Accessed January 28, 2021]. Available at: https://transformer.huggingface.co/doc/gpt2-large.
4. Chakrabarty, T., Muresan, S. & Peng, N. (2020). Generating similes effortlessly like a Pro: A style transfer approach for simile generation. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, pp. 6455-6469,.
5. Clark, E., Ross, A. S., Tan, C., Ji, Y. & Smith, N. A. (2018). Creative Writing with a Machine in the Loop: Case Studies on Slogans and Stories. *IUI '18: 23rd International Conference on Intelligent User Interfaces*, pp. 329–340.
6. Dale, R. & Reiter, E. (1997). Building applied natural language generation systems. *Natural Language Engineering*, vol. 3(1), pp. 57–87.
7. Elkins, K. & Chun, J. (2020). Can GPT-3 Pass a Writer's Turing Test? *Journal of Cultural Analytics*. CA: Journal of Cultural Analytics, p. 17212.
8. Grand View Research. (2019). *Natural Language Generation Market Size, Share | Industry Report 2025* [online]. [Accessed January 28, 2021]. Available at: https://www.grandviewresearch.com/industry-analysis/natural-language-generation-market.
9. Ippolito, D., Grangier, D., Callison-Burch, C. & Eck, D. (2019). Unsupervised Hierarchical Story Infilling. *Proceedings of the First Workshop on Narrative Understanding*. Association for Computational Linguistics (ACL), pp. 37–43.
10. Kazemi, D. (2020). *NaNoGenMo*. *NaNoGenMo* [online]. [Accessed January 28, 2021]. Available at: https://nanogenmo.github.io/.
11. Kutlak, R., van Deemter, K. & Mellish, C. (2016). Production of referring expressions for an unknown audience: A computational model of communal common ground. *Frontiers in Psychology*. Frontiers Research Foundation, vol. 7(AUG).
12. Lampridis, O., Kefalas, A. & Tzallas, P. (2020). Greek Lyrics Generation. *IFIP Advances in Information and Communication Technology*. Springer, pp. 445–454.
13. Manjavacas, E., Karsdorp, F., Burtenshaw, B. & Kestemont, M. (2017a). Synthetic Literature. Writing Science Fiction in a Co-Creative Process. *Proceedings of the Workshop on Computational Creativity in Natural Language Generation (CC-NLG 2017)*, pp. 29–37 [online]. [Accessed January 28, 2021]. Available at: https://www.cpnb.nl.

Doi: https://doi.org/10.54216/IJAACI.010105
Received: January 14, 2022 Accepted: May 26, 2022

14. Manjavacas, E., Karsdorp, F., Burtenshaw, B. & Kestemont, M. (2017b). Synthetic Literature: Writing Science Fiction in a Co-Creative Process. *Proceedings of the Workshop on Computational Creativity in Natural Language Generation (CC-NLG 2017)*. Association for Computational Linguistics (ACL), pp. 29–37.
15. Mcgovern, J. D. & Scott, G. (2016). EloquentRobot: A Tool for Automatic Poetry Generation. Proceedings of the Seventh ACM Conference on Bioinformatics, Computational Biology, and Health Informatics. Seattle.
16. Perera, R. & Nand, P. (2017). Recent Advances In Natural Language Generation: A Survey And Classification Of The Empirical Literature *. *Computing and Informatics*.
17. Reiter, E. & Dale, R. (2000). *Building Natural Language Generation Systems*. Cambridge [online]. [Accessed January 27, 2021]. Available at: https://books.google.ae/books?hl=en&lr=&id=qnWQU9C8bDkC&oi=fnd&pg=PP1&dq=%22Natural-language+generation%22&ots=j0-PfpXphB&sig=s8IZoPnnUFu1p3ubIA9518680iM&redir_esc=y#v=onepage&q=%22Natural-language%20generation%22&f=false.
18. Ricelli, R. M., Monteiro, D. S. & Paraboni, I. (2020). Personality-dependent content selection in natural language generation systems. *Journal of the Brazilian Computer Society*. Springer, vol. 26(1), pp. 1–21.
19. Rice, M. (2020). *Can AI Writing Software Replace Writers? Built In* [online]. [Accessed January 28, 2021]. Available at: https://builtin.com/artificial-intelligence/natural-language-generation.
20. Roemmele, M. & Gordon, A. (2018). Linguistic Features of Helpfulness in Automated Support for Creative Writing. Association for Computational Linguistics (ACL), pp. 14–19.
21. Siddharthan, A., Ponnamperuma, K., Mellish, C., Zeng, C., Heptinstall, D., Robinson, A., Benn, S. & van der Wal, R. (2019a). Blogging birds. *Communications of the ACM*. Association for Computing Machinery, vol. 62(3), pp. 68–77.
22. Siddharthan, A., Ponnamperuma, K., Mellish, C., Zeng, C., Heptinstall, D., Robinson, A., Benn, S. & van der Wal, R. (2019b). Blogging birds: Telling informative stories about the lives of birds from telemetric data. *Communications of the ACM*. Association for Computing Machinery, vol. 62(3), pp. 68–77.
23. Trăuşan-Matu, Ş. (2019). Computer-based Story Generation. An Analysis from a Phenomenological Standpoint. *International Journal of User-System Interaction*, vol. 12(1), pp. 39–53.
24. WILKE, S. & BEROV, L. (2018). Functional Unit Analysis: Framing and Aesthetics for Computational Storytelling. *7th International Workshop on Computational Creativity, Concept Invention, and General Intelligence*. TriCoLore (C3GI/ISD/SCORE). [online]. [Accessed January 28, 2021]. Available at: https://www.researchgate.net/publication/330599016_Functional_Unit_Analysis_Framing_and_Aesthetics_for_Computational_Storytelling.
25. Yang, X. & Tiddi, I. (2020). Creative Storytelling with Language Models and Knowledge Graphs. *Proceedings of the CIKM 2020 Workshops*. Galway [online]. [Accessed January 28, 2021]. Available at: https://kmitd.github.io/ilaria/.

Doi: https://doi.org/10.54216/IJAACI.010105
Received: January 14, 2022 Accepted: May 26, 2022

26. Dumay, J. & Cai, L. (2014). A review and critique of content analysis as a methodology for inquiring into IC disclosure. *Journal of Intellectual Capital*. Emerald Group Publishing Ltd., vol. 15(2), pp. 264–290.
27. Hustadt, U. (2016). *Research Methods in Computer Science* [online]. [Accessed 29 January 2021]. Available at: https://cgi.csc.liv.ac.uk/~ullrich/COMP516/notes/lect06.pdf.
28. Writefull. (2020). *Writefull. Digital Science* [online]. [Accessed 29 January 2021]. Available at: https://writefull.com/.
29. AI Writer. (2020). *AI Writer™ - The best AI Text Generator, promised. AI-Writer.com* [online]. [Accessed 29 January 2021]. Available at: http://ai-writer.com/.

Doi: https://doi.org/10.54216/IJAACI.010105
Received: January 14, 2022 Accepted: May 26, 2022