A test for a Lovelace machine

Yasser S. Arenas Rebolledo
Graduation Thesis, January 2016
Media Technology MSc program,
Leiden University Supervisor: Peter van der Putten
Leiden University Second Supervisor: Maarten Lamers

Abstract

The aim of this study is to find to what extent computers can assist humans in the creative process of writing titles. To this end, a computer tool was designed that suggests randomly selected titles to users from a pre-built corpus. This paper gives a description of both the development of the system and the tests applied to the participants. A total of 89 participants divided in two groups completed two tasks which consisted of giving a non-fixed number of titles for two pictures. One group was allowed to use a template-based system for generating titles, the other group did not use any tools. The results of the experiments show an increase in creativity of the participants that use the computational creativity tool for writing titles.

1 Introduction

Computers are everywhere. At work and at home, people use computer programs in order to draft documents, data sheets and presentations. The programs they use are tested and have a high level of reliability. They have helped professionals become better at their work: they improved their efficiency and accuracy for example.

In 1843, Ada Lovelace already envisioned that computers would assist people with scientific calculations. She also envisioned that computers would be able to assist humans in creating artwork (Toole, 1987). And indeed, tools have been developed to assist artists in their creative endeavours. Verbasizer for example is a computer program used by the musician David Bowie to help him in the song writing process (Thompson, 2007). Some tools in the field of computational creativity look to enhance human creativity, while others seek to surpass it, like software capable of generating music in style of Mozart or Bach (Cope, 2006). Those tools that seek to assist humans, resemble what Lovelace foresaw. In this paper, we will especially look at these "Lovelace machines" .

Computational theorists still debate about what can be considered as creative. In this paper, the definition used by Zhu et al. (2009) is followed. They define creativity as 'the ability to extrapolate beyond existing ideas, rules, patterns, interpretations, etc., and to generate meaningful new ones'. In the same paper parameters are defined for what the output of a creativity tool should look like:

1. The item to be measured has to be different from other existing items. If one can model existing items with a statistical model, the new item should be an “outlier”.

2. The item has to be meaningful. An item that consists of random noise might well be an outlier, but it is not of interest.

A tool that meets both criteria is Titular, a dynamic creativity tool that suggests novel titles for songs (Settles, 2010). A detailed description of a reproduction of this tool can be found in section 2, Materials.

The linguistic tool was thoroughly tested and compared to other title generating systems using Amazon Mechanical Turk (Settles, 2010). In this test, Titular's automatically generated titles scored relatively well compared to other systems on the first criterion of the output for a creativity tool. However, this experiment did not evaluate to what extent the tool actually enhanced human creativity. In an additional test, participants in a songwriter contest experimented with Titular as a support in writing and naming songs. This experiment generated some positive anecdotal evidence; the actual creative output of the participants was not assessed however. Although the combined results of the tests look promising, they leave space to assess to what extent the system actually enhances human creativity.

As such, the interest of this research is not to measure the creative output of the program in itself but to quantify the creativity of individuals that use this tool. The evaluation of the creativity of the participants was made using one of the methods described by J.P. Guilford (Guilford, 1967). In Guilford's work participants were given a plot story and then were asked to provide original titles for the plot story. This approach matches with the purpose of the tool itself. What type of work (song, story or painting) is used in the test does not influence the creative thinking of the participants, as shown in a psychological study on creativity (Chen et al, 2006) where participants were asked to give titles for paintings instead of titles for a story. The score of the titles generated by the participants was calculated according to Torrance's test for creativity (Torrance, 1966) using two of the four scales: fluency and originality.

This paper proposes an experiment to evaluate the creativity of two groups of participants to generate titles for two paintings (see section 3, Experiments). One group was allowed to use the tool and the other was not. Afterwards, the creative performance of each of the individuals was evaluated and the average creative performance of the group that had the tool at its disposal was compared to the creativity of the group that did not. The results showed that the group of participants using the tool demonstrated a significant increase in both fluency and originality. Statistical information on the results of the experiments can be found in section 4, the discussion and conclusion in section 5 and 6.

2 Materials

The first part of this section describes how the Title Generator was built that was used in the experiments. The second part provides details on the addition of interactive extensions to the program.

2.1 Title Generator

The Title Generator was built according to a similar methodology used for Titular, a text synthesis algorithm that can automatically generate song titles (Settles, 2010). The Title Generator tool uses a template based approach (Deemter et al., 2005) to obtain the structures necessary to create titles. For this study, two functionalities were added to give users more control over the tool and the output; these functionalities will be justified later in this paper. As the source code of the original Titular program could not be modified, which was necessary to add the new functions, an own version of the program was developed. In the following paragraphs, the steps to develop the Title Generator will be described.

The dataset used to feed the Title Generator with templates of titles and words was the million song dataset (Bertin-

strangers in the night
->(‘**NNS**’, ‘in’, ‘the’, ‘**NN**’)

wicked woman
-> (‘**JJ**’, ‘**NN**’)

Fig. 1 Titles converted to templates using POS-tags.

Mahieux et al., 2011). This data set consists of 1,000,000 titles of songs of different genres, artists and languages. First, the dataset was filtered for English titles and duplicate titles and words that were not part of the titles of the tracks (such as ‘featuring’, ‘live’, ‘mixed by’, etc.) were eliminated. What remained was a final list of 138,257 tracks that served as a base to extract the templates and the words to create new titles.

Subsequently, the templates of the titles were obtained by replacing every word in the real titles in the list with their respective Part-Of-Speech tag (POS-tag) using the Natural Language Toolkit (NLTK) (Bird, 2006). A full list of all the tags is available in the NLTK documentation. Examples of the tag definitions used are noun-plural (NNS), noun-singular (NN) and adjective (JJ), see Figure 1. It is important to mention that not all the words were replaced: words such as conjunctions, determiners, prepositions and pronouns remain the same. As a result of this process, a new list with 45,844 templates was generated, including a registration of how often each template occurs. The list of

VBG,"['flattering', 'forgiving', 'tuning', 'songwriting', 'bloodsucking', 'shapeshifting', 'spouting', 'arching', 'putrefying', 'waiting', 'spearing', 'deceiving',', 'asleeping', 'nutting', 'starring', 'ainging', 'dampening', 'examining', …’bugging’, …]”

Fig. 2 Word associated with POS-tag VBG.

(‘NN’,’these’,‘NNS’)
->**make** these **robotics**

('we', 'VBD', 'the', ‘NN’, ‘NN’)
-> we **loved** the **alcatraz man**

('if', 'I', 'could', ‘VB’, ‘the’, ‘NNS’)
-> if I could **heal** the **spirits**

('like', 'NN', 'in', 'the', 'NNS')
->like **isle** in the **hands**

Fig. 3 New titles generated by the program replacing POS-tags.

possible templates is a lot shorter than the original track list as many templates follow the same structure.

As a next step, the words in the titles were grouped according to their corresponding POS-tag. This resulted in a list of 20 POS-tags that each have different types of nouns, verbs, adverbs and adjectives. Figure 2 shows a few words associated with the POS-tag VBG (verb, gerund/ present participle). The words and their relative frequency were stored in a different file which contains 28,252 different words. Words with a higher occurrence have a higher chance to be selected by the Title Generator. The same is the case for the list of templates mentioned earlier.

Finally, to produce new context free grammar titles (Lari and Young, 1990), the Title Generator selects five templates from the templates list and replaces each POS-tag in the templates with a word associated with that tag from the word list. Figure 3 shows a few examples of the generated titles.

2.2 Interactive extensions

Three new functionalities were added to the Title Generator that are not included in Titular (Settles, 2010) to give the user more control over the output of the program and to provide the user with more meaningful and familiar titles. The added functionalities allow the user to keep words

('like', '**NN**', 'in', 'the', '**NNS**')

->like **turn** in the **plants**
->like **fox** in the **thieves**
->like **water** in the **cliches**
->like **space** in the **years**

Fig. 4 Different words fill a template chosen by the user.

and obtain different suggestions for a given template, using these functionalities recursively until the desired output is achieved.

The first of the functionalities (see Figure 4) provides the user with the option to click on a template of his or her liking. After selecting the template, different words fill the template. This allows the user to explore all the possibilities of a given template while keeping a structure that the user can distinguish and interpret easily. The second functionality (illustrated by Figure 5) gives the user the possibility to click on a word in the suggested titles in order to obtain new templates using that word. This option aims to inspire users to come up with more title structures as well as to discover new associations with other words. The third and last functionality

love VBP
->little **love**
['RB', 'VPB']
->the firebird that you **love**
['the', 'NN', 'that', 'you', 'VBP']
->you **love** me to steal
['you', 'VBP', 'me', 'to', 'VB']
->I **love** a life
['i', 'VBP', 'a', 'NN']
->mending **love**
['VBG', 'VBP']

Fig. 5 A selected word displayed in different templates.

faces in the **rhythm**
('NNS', 'in', 'the', 'NN')

->sings in the **rhythm**
->moves in the **rhythm**
->shepherds in the **rhythm**
->things in the **rhythm**
->stones in the **rhythm**

Fig. 6 Extension that allows user to keep a part of the generated title while the rest is replaced.

added to the Title Generator is the ability to change or maintain certain words in the suggested template while keeping the same template (see Figure 6). This allows the user to obtain titles with partial changes, only giving new suggestions for a specific part of the template. The objective of these extensions was to make the tool more interactive so that the output of the program would approximate the preference of the user more.

3 Experiments

This study aims to assess to what extent the Title Generator enhances human creativity. Therefore, an experiment was conducted to evaluate whether the tool helps users to be more creative in terms of quantity and originality.

3.1 Writing tasks

In order to assess whether the Title Generator enhances the creativity of its users, the tool was incorporated in a writing task which is based on examples from other studies that measure creativity. Most notably a study in which participants had to write down any number of titles for a short story that was given to them (Eisenberger & Rhoades, 2001) and a study where they were asked to give one title to each one of four images (Chen et al, 2006). For the purpose of this research,

the set-up of the two studies was combined.

The experiment was conducted with a total of 89 individuals who were asked to write down titles for two paintings. In total, 39 of the participants had the Title Generator at their disposal, the rest did not. The individuals were approached in the Utrecht University library and were all students at this University.

The paintings that were selected had to give the participants of the study some guidance, but should also allow them to come up with their own creative interpretations of the image. Therefore, two paintings were chosen that are figurative (they represent clear objects and people) but painted in an expressionist style. Furthermore, in order to prevent participants having seen the paintings before or associating their titles with the artist rather than with the painting, a painter was selected who is relatively unknown in the country where the experiment was run. The paintings that are used are both from the hand of the Ecuadorian painter Oswaldo Guayasamin. The first painting, “pareja en silencio” (in English: couple in silence), portrays a couple hugging each other (Figure 7). The second one, “el guitarrista” (the guitar player), portrays a guitar player (Figure 8).

Fig. 7 painting used in task 1

Fig. 8 Painting used in task 2

Participants were provided with a sheet of A4 paper containing a reduced size but color image of “pareja en silencio” (task 1) on the front side and “el guitarrista” (task 2) on the back side. Both images were followed by the instruction “please list all the possible titles you can think of for this painting” and contained a number of lines for the participants to write down the titles. The exercise was accompanied by the verbal instructions: “take all the time you need and when you are ready please return the sheet to the interviewer”.

Afterwards, the creative performance of each of the individuals was evaluated and the average creative performance of the group that had the Title Generator to its disposal was compared to the creativity of the group that did not.

3.2 Group 1: control

The objective of the first study was to establish a baseline of sole ‘human’ creativity. Hence, the participants in this group did not have the Title Generator to their disposal nor any other tools. This group served as a control group for the experiment. Their results were compared to those of group 2.

3.3 Group 2: assisted

To measure the difference in creativity with the control group, the participants of group 2 were allowed to use the Title Generator as inspiration to write down titles for the two images. Users were allowed to explore the capabilities of the Title Generator to produce titles according to their preferences. Participants were permitted to use the generated titles in two possible ways:

1. make an exact copy of the title given by the computer program.

2. modify the given title by adding or replacing words of their own inspiration.

To immediately verify the usefulness of the tool after the task, participants were asked to provide objective feedback about the Title Generator. For this a form was given to the participants to be filled out (see appendix A).

3.4 Evaluation of creative performance

The data gathered by the two studies was evaluated based on two dimensions of the evaluation for creativity, which are fluency and originality (Torrance, 1966). Fluency refers to the number of titles each participant gave for each painting. The participants received one point for every given title.

The originality was measured by establishing how often the title was mentioned by the other participants of the same group and for the same painting. If a title was given by less than 5% of the participants, a score of 1 was given to the title. If the title was given by less or equal than 1% of the participants 2 points were given for each title. A score of 0 was given if more or equal than 5% of the participants came up with the same title. To obtain a clean measure of originality for each participant, the points obtained were divided by the total fluency of each participant. It was considered whether the scores for originality should be influenced by their quality too, as an important criterion of creativity is that output should be meaningful. However, evaluating quality is a very subjective task and requires a separate research to asses whether the titles present high or low quality; hence, it was decided not to include it in this work.

4 Results

This section provides an overview of the creative performance of the group that used the Title Generator and the one that did not. They are scored on two dimensions: fluency and originality.

4.1 Fluency

As explained in the above section, the dimension fluency refers to the number of titles given by the user to the paintings. The results show that for the first task, the mean fluency score for participants using the Title Generator (group 2) is greater than the number of titles given by the group that did not have the Title Generator at its disposal (group 1). The chart in Figure 9 shows the difference in the means for the number of titles in both groups for the painting *couple in silence* (task 1).

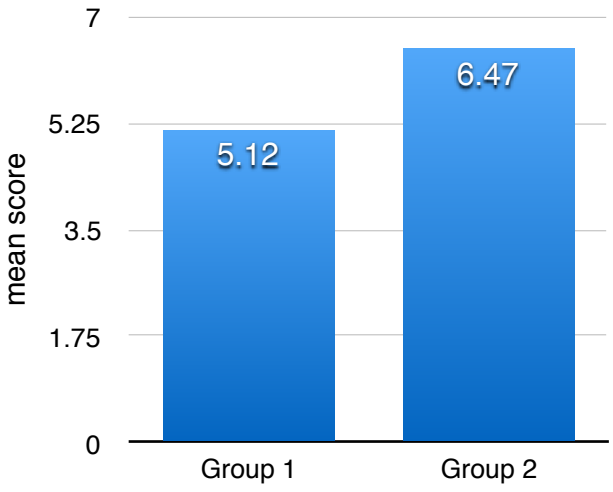

Fig. 9 The mean score for fluency in group 2, using the Title Generator, is greater than that of group 1 in task 1.

	group 1	group 2
n	50	38
mean	5.12	6.47
median	5	5.5
sd	2.38	3.48
min	1	1
max	14	20

Table 1 statistical measurements for fluency for task 1.

	group 1	group 2
n	50	39
mean	4.62	6.74
median	4	6
sd	2.41	3.83
min	2	1
max	16	19

Table 2 statistical measurements for fluency for task 2.

More statistical measurements on the fluency of both groups in task 1 can be found in Table 1. A t-test shows that the difference in the mean fluency score between the two groups is significant ($p < 0.044$).

Similar results were found for the painting *the guitar player* (task 2). The mean score for fluency in group 1 was significantly lower than that of group 2 as can be seen in the chart in Figure 10. Additional statistical results on the fluency of both groups in task 2 are shown in Table 2. A t-test shows that the difference is again significant ($p < 0.003$).

4.2 Originality

The originality is based on how often each title was mentioned for the same painting by other participants in the group. For the first task, the mean score in originality is greater in the group that used the Title Generator. The following chart (Figure 11) shows the difference in the mean score for originality in group 1 and 2 for the painting *couple in silence* (task 1). Extra statistical measures on the originality of both groups in task 1 are shown in Table 3. A t-test shows that this difference is significant ($p < 1.79E-08$).

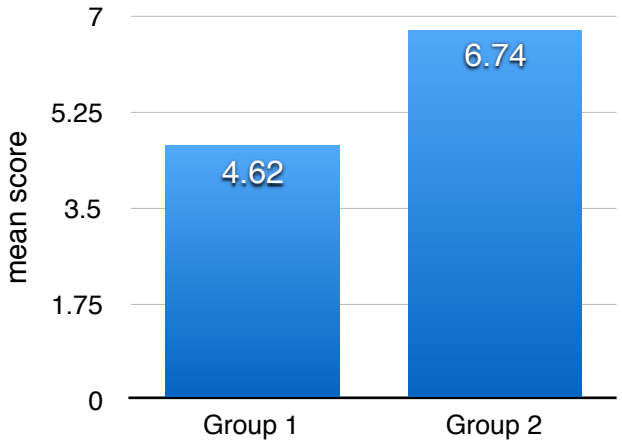

Fig. 10 The mean score for fluency in group 2, using the Title Generator, is greater than that of group 1 in task 2.

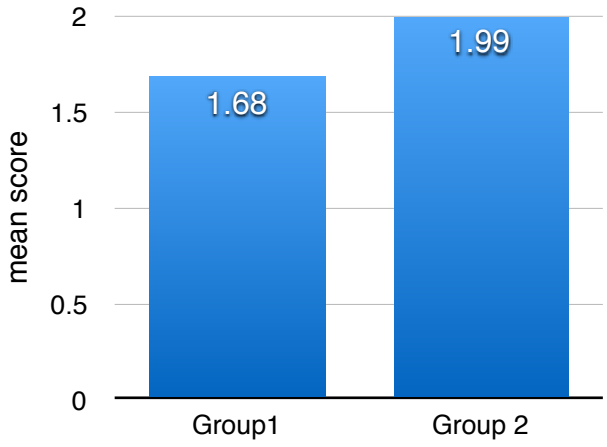

Fig. 11 The mean score for originality in group 2, using the Title Generator, is greater than that of group 1 in task 1.

	group 1	group 2
n	50	39
mean	1.68	1.99
median	1.71	2
sd	0.31	0.05
min	1.00	1.75
max	2.00	2

Table 3 Statistical measurements for originality for task 1.

	group 1	group 2
n	50	38
mean	1.55	1.99
median	1.60	2
sd	0.47	0.03
min	0.25	1.8
max	2.00	2

Table 4 Statistical measurements for originality in task 2

Almost identical results were found for the painting *the guitar player* (task 2). As Figure 12 shows, the mean score for originality in group 1 was significantly lower than that of group 2. Additional statistical measures for originality in task 2 comparing the two studies are shown in Table 4. A t-test shows that the difference in their means is significant ($p<9.49E-09$).

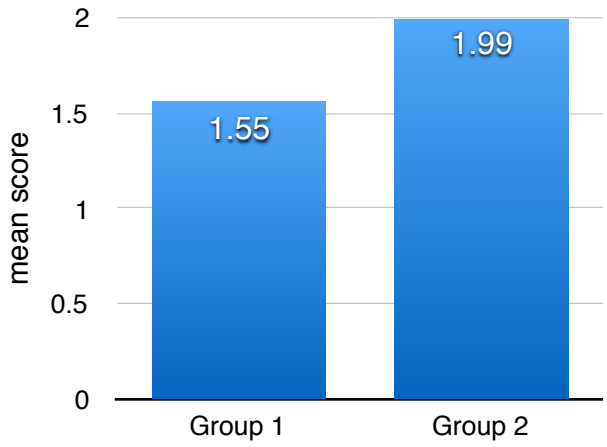

Fig. 12 The mean score for originality in group 2, using the Title Generator, is greater than that of group 1 in task

The most common responses given by group 1 in task 1 are shown in Table 5. The high occurrence of the same titles reduces the mean of originality, as can be seen in Figure 11. The same can be observed for group 1 in the second task (Table 6), where many participants again came up with identical titles.

title	number of participants that mentioned title
love	18%
embrace	14%
big hands	10%
the kiss	10%
the hug	10%

Table 5 Most common responses in group 1 for task 1.

title	number of participants that mentioned title
music	6%
the guitar	14%
love for music	6%
lonely	6%
guitar hero	6%

Table 6 Most common responses in group 1 for task 2.

title	number of participants that mentioned title
the hug	6%
glorious love	1%
wolves of night	1%
there is no day like you	1%
love the cool existance	1%

Table 7 Most common responses in group 2 for task 1.

The Tables 7 and 8 show that in group 2

title	number of participants that mentioned title
guitar man	6%
far from home	1%
for whom the life is silent	1%
blues of the outcast	1%
wild repetition	1%

Table 8 Most common responses in group 2 for task 2.

there is almost no repetition of titles in both tasks. In both studies, only one title was mentioned by more than one participant. The most common titles in this second group are “the hug” and “guitar man” for respectively test 1 and 2.

As explained in section 3.4, participants were asked to provide feedback about the Title Generator after they completed both tests. Table 9 shows the results of the survey.

Question	yes
Did you feel the Title Generator helped you to come up with more titles?	82%
Did you feel the Title Generator helped you come up with better titles?	59%
I used some of the words that came up in the generator	66%
I used (parts of) the syntax that the generator suggested	35%
Other: it gave me inspiration	41%

Table 9 Objective feedback from group 2

5 Discussion

The participants assisted by the creativity tool had an increase in the number of their responses. Their fluency is 26.3% greater for task 1 and 45% for task 2. This is also reflected by the feedback: 82% of the participants in group 2 indicated that the Title Generator helped them to come up with more titles. The data suggests that the Title Generator also helped the participants to come up with more original titles: the group that used the Title Generator was on average 18.45% and 28% more original for task 1 and task 2 respectively than those that did not use the tool. The survey shows that 58% of the participants using the Title Generator found that the tool helped them to think of better titles. Although there was a significant difference between the group that used the Title Generator and the one that did not for both fluency and originality, the tool seemed especially helpful to enhance the participants’ fluency.

As for how the Title Generator assisted the participants: those that said that the tool helped them to come up with more or

better titles, indicated they used words displayed by the generator or used part of the titles. Under “other”, a large number of participants (41%) indicated that the tool somehow inspired them (it helped them think out of the box, made them be more creative, etc). The participants were not given a fixed time to finish both tasks as it was deemed important that the participants would concentrate on the creative process, rather than focusing on finishing on time. It was observed that the participants of the second group took more time to finish the task; the fact that the tasks entertained them longer seems to support their claim that they did not run out of inspiration as soon as the participants in the first group. This observation could also be attributed to the ‘novelty bias’: participants might have spent more time on the task because of the novelty of the tool, and therefore showed higher scores on fluency. More experiments are needed to test which of the two assumptions holds true.

The majority of the titles generated by the first group, can be seen as more literal descriptions of the painting. The titles refer to the figures on the painting (“big hands”, “music”) or to concepts that were commonly seen as being related to the painting (“love”, “lonely”). The group that had to their disposal the Title Generator came up with more figurative titles. Next to more literal descriptions such as “the hug, free interpretations such as “wolves of night” were generated by the second group; instead of the more common concept of “lonely”, the more freely associated “far from home” was used for example. Nonetheless, the most ‘obvious’ titles for both paintings (“the hug” for the first painting and “the guitar” for the second painting) were still the most mentioned responses in group 2 as well. This seems to indicate that the participants did not just merely copy the suggestions of the Title Generator but also relied on their own creativity.

6 Conclusion and Future work

Ada Lovelace wrote about the possible collaboration between humans and technology, and how both science and art can benefit from working with computers. The aim of this research was to assess whether Titular, a tool developed to assist artists, actually enhances their creativity. An adapted tool, the Title Generator, was used in experiments to answer the research question. The results of the experiments seem to suggest that in this case, indeed, the collaboration between computer and man enhances human creativity. The results showed that participants in a creative writing test present more fluency and originality when they use the Title Generator. Four t-tests furthermore showed that the difference between the group that used the Title Generator and the one that did not, was significant.

The difference in fluency was most significant between the two groups. Participants indicated that both the words as well as the templates of the tool gave them inspiration to think of new titles. Although less significant, the group that used the tool was also clearly most original. With respect to the originality, the participants that used the Title Generator were able to come up with more figurative titles for the paintings, as opposed to the more literal descriptions of the participants in the other group. This seems to suggest that the Title Generator helped them to think more out of the box and generate titles that went beyond the obvious. The aspect of generating (meaningful) outliers that users are able to interpret, is an important characteristic for creativity tools.

For this study new adaptations were made to the original Titular program in order to give the user more control over the program, and enhance interaction. In further studies a measurement of the usability of these extensions could be an interesting subject of research.

Furthermore, those that used the tool and completed a survey on if and how it helped them were asked for possible areas of improvement. Several participants suggested that a possibility to obtain titles based on a seed-word or other personalised data as an input to the program could be a valuable extension. Such an addition raises interesting questions for follow up research; for example, what will be the difference in creativity using a personalised vs a non-personalised title generator?

References

Bertin-Mahieux, T., Ellis, D.P., Whitman, B. and Lamere, P., 2011. The million song dataset. In *ISMIR 2011: Proceedings of the 12th International Society for Music Information Retrieval Conference.*

Bird, S., 2006. NLTK: the natural language toolkit. In *Proceedings of the COLING/ACL on Interactive presentation sessions. Association for Computational Linguistics.*

Chen, C., 2006. Boundless creativity: Evidence for the domain generality of individual differences in creativity. In *The Journal of Creative Behavior.*

Cope, D., 1987. Experiments in Music Intelligence. In *Proceedings of the International Computer Music Conference, San Francisco: Computer Music Assn.*

Eisenberger, R. and Rhoades, L., 2001. Incremental effects of reward on creativity. In *Journal of personality and social psychology.*

Guilford, J.P., 1967. The Nature of Human Intelligence.

Lari, K. and Young, S.J., 1990. The estimation of stochastic context-free grammars using the inside-outside algorithm. In *Computer Speech and Language.*

Settles, B., 2010. Computational creativity tools for songwriters. In *Proceedings of the NAACL HLT 2010 Second Workshop on Computational Approaches to Linguistic Creativity. Association for Computational Linguistics.*

Torrance, E. P., 1966. The Torrance Tests of Creative Thinking-Norms-Technical Manual Research Edition-Verbal Tests, Forms A and B-Figural Tests, Forms A and B. Princeton, NJ: Personnel Press.

Thompson, D., 2007. Hallo Spaceboy: The Rebirth of David Bowie. ECW Press.

Toole, B., 1987. Poetical Science. In *The Byron Journal.* In The Byron Journal 15: 55–65

Van Deemter, K., Krahmer, E. and Theune, M., 2005. Real versus template-based natural language generation: A false opposition?. In *Computational Linguistics.*

Wallach, M. A. and Kogan, N., 1965. Modes of thinking in young children: A study of the creativity- intelligence distinction Holt, Rinehart, & Winston , New York.

Zhu X., Xu Z., and Khot T., 2009. How creative is your writing? In *Proceedings of the Workshop on Computational Approaches to Linguistic Creativity (CALC).*

Appendix - Survey after to ask objective feedback after study 2:

Did you feel the title generator helped you to come up with more titles? y/n

Did you feel the title generator helped you come up with better titles? y/n

If you answered yes to any of the above two questions, how did the title generator help you come up with more and/or better titles? Tick the box:

- ☐ I used some of the words that came up in the generator
- ☐ I used (parts of) the syntax that the generator suggested
- ☐ Other: _____

How could the title generator help you come up with more/better titles?

_____

_____

_____