The Propaganda Machine: Generating Biased Reports about Risk Games

Rafael Dulfer
University of Twente
Enschede, The Netherlands
r.m.dulfer@student.utwente.nl

Lorenzo Gatti
Human Media Interaction group
University of Twente
Enschede, The Netherlands
l.gatti@utwente.nl

***Abstract*—In this work we present a system that generates reports for a game of Risk. These reports, far from being "neutral", aim instead at mimicking propaganda, and try to influence the opinions of the readers about the performance of a Risk player. The system, while limited in scope, was able to persuade some test subjects in a qualitative evaluation, hinting at the abilities that a more sophisticated system might have.**

***Index Terms*—Natural language generation, persuasion, propaganda, board game**

I. Introduction

Persuasive text has been a long-standing topic of research in the Natural Language Generation (NLG) community. While there is ample literature on the computational treatment of many types of persuasive text (ranging from works based on theories of persuasion, e.g. [1] and [2], down to systems for advertisement generation [3]), little consideration has been given to propaganda, the persuasive text *par excellence* (for the difference between propaganda and persuasion, see [4]). The development of a real NLG system for propaganda is ethically troublesome, but a "toy example" may prove useful to study both the way propaganda works, and potential strategies to mitigate its effectiveness. In this paper, we describe the "Propaganda Machine", a small template-based system that aims at biasing the reader's perception of the outcomes of a Risk game. A preliminary evaluation indicates that even simple NLG systems might have a powerful effect on readers, and paves the way for further research on the topic.

II. Related Works

Most early works on persuasive NLG focused on argumentation, i.e. the generation of a coherent sequence of arguments supported by logic (see [5] for a review). Emotional persuasion (e.g. [6]) has also been studied, although to a lesser extent. Both approaches are often used in the context of behavior change support systems, where there is a clear pragmatic goal. The Propaganda Machine is also using a mixture of both: it adds fallacious claims on top of real facts (e.g. stating that deploying more "troops" in a region will ensure the future victory), while using exaggerated wording and mitigating negative events.

In general, biased reporting and other malicious uses of technology are becoming more important in NLG and NLP research. For example, automatic propaganda detection has been subject of a recent survey [7] and SemEval task [8]. Concerning NLG, the GROVER system [9] is the most relevant to this work, since it can generate "fake news" articles. GROVER, however does not base its news on "real" data, but fabricates facts starting from a user-provided headline. The Propaganda Machine, instead, is data-based, and only manufactures "explanations" and "motivations" for the facts.

The system here presented uses the events of a game of Risk as source material for the generation. Gervás [10] proposes a computational model for "storifying" the events of a game of chess, considering each piece as an actor with an incomplete view of the board and finding narrative threads that could describe the movement of the pieces; while the domain of chess might include even elements of suspense [11], these works are focusing on the content determination step, without producing a linguistic output (as opposed to [12], where the goal is to simply generating linguistic explanation for the piece movements, without specific narrative qualities). Also connected to our work is [13], where a system is using the ruleset of a role-playing game to simulate the interaction between two humans (the "Master of Ceremonies", i.e. the storyteller, and a player) and an NLG module converts them into text. A more common use case of NLG in the domain of games is, however, the generation of text for the players, i.e. varied or immersive game content that a human will read while playing, such as the text produced by in-game agents [14], [15] that the user could encounter in-game.

III. The Propaganda Machine

While deep learning systems are the state of the art in NLG [16], we developed a template-based data-to-text system since it is simpler to set up and document, and its inherent limitations (e.g. being bound to a specific dataset, no possibility of easily learning a new domain or topic) make it useless for real propagandist purposes. In this way, the research is kept to an ethically-acceptable minimum viable product.

The Propaganda Machine generates reports on "fictional war data" based on a Risk game. This fits the "toy example" approach taken for the system, and also helps further bind the system to a toy domain. The data itself comes from the open-

978-1-6654-3886-5/21/$31.00 ©2021 IEEE

source Risk clone "Domination" [17]. Risk is a classic board game in which its players attempt to take over the world[1].

A game of Risk has 6 actions that can occur: *Fortification*: new units are placed on a country. *Attack*: a player attacks another player from country A to country B. *Move*: a player moves units from country A to country B. *Get a card*: a player gains a card if they conquered a country this turn, could later be traded for more units. *Trade cards*: a player can trade 3 cards of similar type to gain more units. *Complete a mission*: a player completes a specific goal, this action possibly wins them the game.

We extended the Domination original source so that it exports all these actions to a JSON file, and let two AI players complete a full game of Risk.

In the context of this Risk game, the final goal of the system then becomes convincing the reader that a losing player is actually winning.

The Propaganda Machine and the extended Domination code are available at https://github.com/Rafaeltheraven/The-Propaganda-Machine.

A. Content Selection

An average game of Risk exported this way consists of about 490 events. As Risk is a turn-based game, the first step is thus dividing these events by turn. This organization brings most of the relevant data together and allows us to distinguish between actions performed *by* a certain player and actions performed *against* this player.

We also decided to group related data together: given a single player's turn, the most interesting events will always be related to a specific war. A war is defined as a set of attacks from one country A to another country B. We can then conclude that any previous action also involving country A will be relevant to this war. This is the first step of relevancy. The second step is to consider whether a war is continuing. If the player is able to use country A to conquer country B, then not only are all actions involving country A relevant, but also all actions involving country B. It is possible that country B is now used to invade country C. A continuing war like this will keep being grouped recursively, until the attacks stop. In this way, a single turn could have multiple wars, all of which have related events being grouped together (see Figure 1). Grouping these events allows us to make a more cohesive story. A single war and its consequences can be described in a paragraph, only for the next paragraph to then report the next war.

As a result of this content selection and grouping, the final output of the system will describe the events in the game turn by turn, war by war, keeping all relevant information together to allow for a more cohesive narrative flow.

While this method can, in theory, generate texts for the whole game, such reports would be overly long and particularly boring to read. Since the goal of the system is to alter people's perception of the Risk game, we decided to focus directly on the central part of a match, where things are most exciting, the winner will start to emerge, but is not yet completely obvious. The center of the match was defined as simply being $turns/2$. From this point, we chose to report on 4 turns, as it proved to be long enough to show variation, but not so much as to become boring.

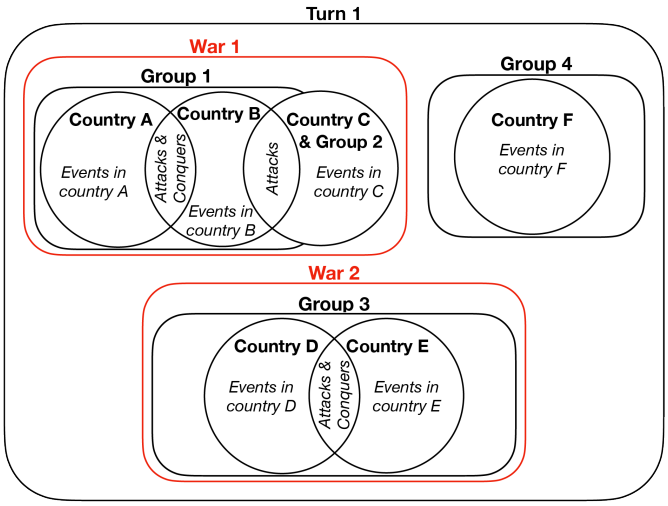

Fig. 1. In a single turn, events are grouped together in different "wars"

B. The Templates

The templates take the data generated and selected in the previous steps and create full sentences from them. The system can create two types of reports: a neutral report and a biased/propagandist report. The templates follow the default Python String Formatter format, i.e. any string of characters between brackets are interpreted as variables.

The basic sentences. The writing style of the template sentences was inspired by research on propaganda [18] and real examples from the German Propaganda Archive [19]. The propagandist report is built up from various basic sentences which follow a flow defined by the grouping described in III-A. The templates are then split into two different types, negative and positive. Simple heuristics are used to decide whether an event is positive (e.g. winning a country or an attack, receiving a card) or negative (unsuccessful attacks, losing a country, etc.).

The Propaganda Machine then picks a template from the list of templates depending on the type of event. If the event was not an attack, we can simply pick a random template from the list. If the event was an attack, however, we have two more attributes to determine. The first one is the style of the attack, of which there are three types: *steamroll*, when our army is big and their army is small; *underdog*, when their army is big and our army is small; *struggle*, when both armies are equally matched. Then, one of these four outcomes for the attack is determined: *win*, if our army wins a country; *loss*, if our army loses a country; *defend*, if our army defends a country;

[1]It is worth noting that most countries in the Risk map do not follow the borders of actual countries (e.g., the U.S. is divided in "Western U.S." and "Eastern U.S."), further distancing the game from reality and sidestepping ethical issues related to borders and nationalism.

progress, when progress is made in the war, but neither army prevailed. Given the available information, the system can pick a template from the list that fits the attack event (e.g., for an "underdog win" event, the template could be "*Like David to Goliath our brave soldiers marched into what {otherplayer} thought was assured destruction, but {currplayer} knows no fear and swiftly took down the giant in {country2}!*").

Connecting the sentences. The above system does not make for a fully coherent text yet, as it still lacks any sort of cohesion. This is added to the text by inserting connective sentences. Connective sentences serves no pragmatic purpose, but instead connect separate events together. It is these lines which allow the text to become a cohesive narrative whole.

The connective sentences function on a number of groupings which correspond to those described in section III-A. Just as with the event templates, there are positive and negative connective texts. Below this level, there are the *introduction* and *continuing* texts. An *introduction* text serves as introduction for the entire text corresponding to this turn. A turn can be split up into several paragraphs, one paragraph for each group of war data. The *continuing* texts look at the previous groups of data, see whether they were generally positive or negative, and connect the two paragraphs together.

The second layer of connective text comes in the form of connecting *related events* together. As described previously, all events related to a "war" can be grouped together. We can connect these events into a single paragraph by using more linking sentences.

Most of these templates are used to connect various "attack" events together. *Opening* contains lines that start a paragraph, *conclusion* those that end paragraphs. *reason* is used to connect a specific attack event to other type of events like fortifications or movement events. *Intermittent* is for additional flavor text regardless of event type, and *continuing* helps connect a continuing advance from country A to country B.

The results of this process can be seen in the "positive" text of Figure 2.

The neutral report. Differently from the propagandist report, the neutral report is meant to show all data in a factual way. It goes through every action in the game sequentially and describes them. Its purpose is to be used in the evaluation, to present readers with an unbiased point of view, in contrast with the highly biased propagandist reports.

Because of the simple nature of the reports, the templates are also quite simple: every event type has a single line dedicated to it, and when this event is encountered the line is selected and filled in.

These templates generate text such as the following: "*Mongolia has attacked Romania in Eastern United States from Central America. Eastern United States had 5 units before and 5 units afterwards. Central America had 10 units before and 8 units afterwards.*".

IV. Evaluation

The evaluation aimed at testing how well the propagandist texts were able to convince the reader that the losing player was actually winning.

The qualitative evaluation concerned 4 texts, all generated by the system. The game itself had two AI players, Romania and Mongolia, which both attempted to control the entire world, resulting in Mongolia being the eventual winner. As such, the goal of the evaluation became to try to convince the readers that Romania was actually winning. The texts presented to the subjects were, in varying orders, a neutral report, a biased report for Romania, a biased report for Mongolia and a background report. The background report (see Figure 3) is a special report created for the evaluation, which gave readers more info about the state of the game. Twelve consenting test subjects were presented these texts and interviewed about their opinions.

A. The Interviews

Subjects were invited to a research about computer-generated texts based on a Risk game, and led to believe the evaluation was about text quality. This was done to avoid influencing subjects into reading more critically than they usually would, and also to mimic real propaganda, as propaganda seldom presents itself as such.

Subjects were given a biased report and instructed to read it completely. Then, they would be asked some deflective open questions about quality and language use. They would then be presented with the question: "Who do you think is performing best?". Subjects would finally be interviewed on why they think a certain player is doing better.

After this first set of questions, the subject would be presented with a second text. This served as an opportunity to see how conflicting information might cause the subject to change their mind. After the second text, subjects would again be asked which player was performing the best and to elaborate on their decision. This final discussion marked the end of the interview, after which subjects were told the true purpose of the research[2].

The evaluation actually consisted of multiple sets of interviews, each seen by multiple subjects, differing in the presented reports and their order.

The first 4 subjects were presented with the background report at first, then the Romania report and finally the neutral report. This was to check how easy it would be to influence people that had a full overview of the state of the world and of the facts. In this round people could generally realize that Mongolia was winning.

The second round was the same, but without the background text. This better reflects a real-world situation in which not everything is known about the state of the world. This round saw people more frequently call Romania the winner, even when presented with objective facts from the neutral report.

The third round dispensed with the neutral report. Instead, 2 subjects were first shown the Romanian report and then followed with the Mongolian report. This round evaluated how

[2]The evaluation procedure has been approved by the Ethics committee of our University.

The glory of god has smiled upon us again! Word has reached from the Eastern United States front. It took only 1 hour to make Western United States great again! We killed all enemy soldiers with ease. The weakling Mongolians have no chance against our brave soldiers. This was made possible because Our great strategists have decided to deploy more troops to Eastern United States. This will surely bring us ever closer to victory! We are strategically moving our troops in Eastern United States to Western United States. Eastern United States will be kept safe by the brave men staying behind, while Western United States will become stronger and safer than ever.

Fig. 2. Positive Introductory Paragraph

Background

These reports concern the nations of Mongolia and Romania. Both nations are trying to take control of all countries in the world. Before the events in these reports, Mongolia has the following countries: Argentina, Japan, Southern Europe, Brazil, Middle East, Siam, India, Afghanistan, China, Ukraine, Western Europe, Peru, Quebec, Ontario, Greenland, North West Territory, Alberta, Alaska, Kamchatka, Yakutsk, Siberia, Mongolia, Irkutsk, Indonesia, Western United States, Western Australia, North Africa, New Guinea, Eastern Australia, Egypt, Venezuela, Ural and Central America, while Romania has the following countries: Congo, South Africa, East Africa, Iceland, Great Britain, Madagascar, Scandinavia, Eastern United States and Northern Europe.

World Map

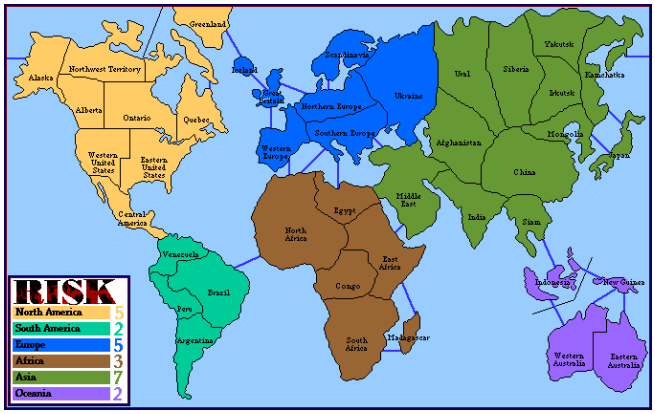

Fig. 3. Background report used for the evaluation

conflicting propaganda would affect the reader. While people would at first generally point to Romania as the winner, after being shown the Mongolian text they would usually settle on Mongolia as the final overall winner.

The last round was a reversal of the third: 2 subjects were shown the Mongolian report, then the Romanian one. This round had mixed results.

B. Results

One out of three interviewees incorrectly pointed to Romania as the player who did best overall. Of the people who were only shown the Romanian and the neutral text, only one in four saw through the propaganda and realized Mongolia was the winner. Some of the interviewees pointed out the propagandist nature of the texts, but this did not consistently affect their perception of the state of the game.

The neutral text was not able to convince any of the Romania voters that Romania was not doing best. While it made some subjects doubt their initial conviction, they usually found it too confusing and boring to read to really change their opinions.

Subjects frequently cited the abundance of repetition in the texts to be detrimental, either causing them to get annoyed by the text or, in the worst case, feel like they were being lied to. Response to the propaganda was quite varied. Some correctly pointed out the propaganda and therefore went with the exact opposite of what the propaganda was trying to convince them of, while others did not notice the propaganda at all or even directly referenced the propagandist lines as reasons why one side might be winning ("Mongolia is fortifying a lot, I think they might be afraid of Romania.").

V. Conclusion

The propaganda turned out to work surprisingly well. While only a minority of those interviewed specifically pointed at Romania as the winner, those who did believed in the propaganda quite strongly.

It is worth noting that the Propaganda Machine is limited in ways that real-world ones would not be. A "real" system could selectively show information, it could appeal to common symbols and it could even be tailored to a specific population. If we take the variation which most closely resembles real-world propaganda scenarios (no full knowledge of the world, only shown one side of propaganda), then three in four people fell for texts that were quite simple and blatant in their propagandist contents. One could imagine that a more sophisticated machine, with less repetition, more professionally-written templates and more real-world virtues to appeal to, could work even better. Such a system can be created (and possibly already is in the disguise of neutral journalism [20]).

Our initial experiment shows that countering propaganda is not easy: generally, the neutral text had no effect, possibly due to its verbosity or limited appeal. The background text was the best at showing readers that Romania was losing, regardless of propaganda. In the real world, however, such omnipresent knowledge of the state of the world is impossible. The text that was second-best at convincing readers Romania was actually losing was the Mongolia text, but fighting propaganda with propaganda, while it does have historical precedent, is hardly the most moral method. A good option could be finding a middle ground between the neutral text and the background text. The former was an "information dump", causing readers to get overloaded, bored, to generally ignore it. The latter was easy to read and understand, but it required an omniscient

knowledge of the world which is impossible in the real world. If the raw data from the neutral text were to be presented in a more digestible, but still objective style, it could hopefully help people see through the propaganda. More research and a thorough extensive evaluation is of course needed to get quantitative results and to be able to better understand the effects of such propagandist texts.

Apart from its usage as a case-study, with the addition of more templates (and after adapting to games other than Risk), the Propaganda Machine could also be useful in the videogame industry. It could provide an additional immersive element for strategy games, adding color and realism by providing messages that could appear in fictional newspaper for the player's own faction or for the opponent; or it might be a more integral part of the game dynamics, where the propagandist nature of its message could try to steer the player's course of action towards a certain direction, as do the pre-scripted messages in "Papers, Please".

REFERENCES

[1] L. Gatti, M. Guerini, O. Stock, and C. Strapparava, "Sentiment variations in text for persuasion technology," in *Proceedings of the 9th International Conference on Persuasive Technology (PERSUASIVE 2014)*. Springer, 2014, pp. 106–117.

[2] G. Carenini and J. D. Moore, "Generating and evaluating evaluative arguments," *Artificial Intelligence*, vol. 170, no. 11, pp. 925–952, 2006.

[3] V. Munigala, A. Mishra, S. G. Tamilselvam, S. Khare, R. Dasgupta, and A. Sankaran, "PersuAIDE! An adaptive persuasive text generation system for fashion domain," in *Companion Proceedings of the The Web Conference 2018 (WWW)*, 2018, pp. 335–342.

[4] G. S. Jowett and V. O'Donnell, *Propaganda & persuasion*. Sage publications, 2018.

[5] T. J. M. Bench-Capon and P. E. Dunne, "Argumentation in artificial intelligence," *Artificial intelligence*, vol. 171, no. 10-15, pp. 619–641, 2007.

[6] M. Miceli, F. de Rosis, and I. Poggi, "Emotional and non-emotional persuasion," *Applied Artificial Intelligence*, vol. 20, no. 10, pp. 849–879, 2006.

[7] G. D. S. Martino, S. Cresci, A. Barrón-Cedeño, S. Yu, R. D. Pietro, and P. Nakov, "A survey on computational propaganda detection," in *Proceedings of the 29th International Joint Conference on Artificial Intelligence (IJCAI-20)*, 2020.

[8] G. Da San Martino, A. Barrón-Cedeno, H. Wachsmuth, R. Petrov, and P. Nakov, "SemEval-2020 task 11: Detection of propaganda techniques in news articles," in *Proceedings of the 14th Workshop on Semantic Evaluation (SemEval)*, 2020, pp. 1377–1414.

[9] R. Zellers, A. Holtzman, H. Rashkin, Y. Bisk, A. Farhadi, F. Roesner, and Y. Choi, "Defending against neural fake news," in *Advances in Neural Information Processing Systems 32 (NeurIPS 2019)*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché Buc, E. Fox, and R. Garnett, Eds. Curran Associates, Inc., 2019, pp. 9054–9065. [Online]. Available: http://papers.nips.cc/paper/9106-defending-against-neural-fake-news.pdf

[10] P. Gervás, "Targeted storyfying: Creating stories about particular events." in *Proceedings of the 9th International Conference on Computational Creativity (ICCC '18)*, 2018, pp. 232–239.

[11] R. Doust and P. Gervás, "Content determination for chess as a source for suspenseful narratives," in *Proceedings of the 3rd Workshop on Computational Creativity in Natural Language Generation (CC-NLG 2018)*, 2018, pp. 26–33.

[12] J. Kowalski, Ł. Żarczyński, and A. Kisielewicz, "Evaluating chess-like games using generated natural language descriptions," in *Proceedings of the 15th International Conference on Advances in Computer Games (ACG 2017)*, M. H. Winands, H. J. van den Herik, and W. A. Kosters, Eds. Cham: Springer International Publishing, 2017, pp. 127–139.

[13] A. Tapscott, C. León, and P. Gervás, "Generating stories using role-playing games and simulated human-like conversations," in *Proceedings of the 3rd Workshop on Computational Creativity in Natural Language Generation (CC-NLG 2018)*, 2018, pp. 34–42.

[14] C. R. Strong, M. Mehta, K. Mishra, A. Jones, and A. Ram, "Emotionally driven natural language generation for personality rich characters in interactive games," in *Proceedings of the 3rd AAAI Conference on Artificial Intelligence for Interactive Digital Entertainment (AIIDE-07)*, 2007.

[15] U. Ehsan, P. Tambwekar, L. Chan, B. Harrison, and M. O. Riedl, "Learning to generate natural language rationales for game playing agents," in *Joint Proceedings of the AIIDE 2018 Workshops (AIIDE-WS 2018)*, 2018.

[16] A. Gatt and E. Krahmer, "Survey of the state of the art in natural language generation: Core tasks, applications and evaluation," *Journal of Artificial Intelligence Research*, vol. 61, pp. 1–64, jan 2018.

[17] yura.net, "Domination," 2020. [Online]. Available: https://domination.sourceforge.net/

[18] H. D. Lasswell, "The Theory of Political Propaganda," *American Political Science Review*, vol. 21, no. 3, pp. 627–631, aug 1927.

[19] R. Bytwerk, "German propaganda archive," 1998. [Online]. Available: https://research.calvin.edu/german-propaganda-archive/

[20] A. Graefe, "Guide to Automated Journalism," *Tow Center for Digital Journalism Report*, 2016.