City, University of London Institutional Repository

A Cognitive Theoretical Approach of Rhetorical News Analysis

Ishrat Rahman **Sami**[1], Dr Tony **Russell-Rose**[1] and Prof. Larisa **Soldatova**[1]

[1]*Goldsmiths, University of London, New Cross, London, SE14 6NW*

Abstract

The storytelling narrative is the key to conveying an author's opinion and argument about a specific topic to intended readers. A good narrative not only conveys the underlying message but also leads readers to a better conceptual understanding of the discussed topic. Stories play a vital role in understanding through their chronological style of reporting. Similarly, to gain readers' attention from beginning to end, news agencies generally adopt an inverted pyramid structure where a story starts with stating the most important material. The facts of news are encapsulated in five basic questions *Who*, *Where*, *What*, *When* and *Why* which are fundamental for any news readers' understanding. Distributions of the categorical facts of the news correlate to the answers to *What*, *When*, *Where* and *Who* questions and the answer to *Why* is correlated to the authors' argumentation and evidence. In this paper, we presented a theory of mapping *5Ws* and *Aristotle's Rhetoric* into the format of Joseph Campbell's *The Hero's Journey* as a structural story template to assist in automatic understanding of the structure of news and evaluated the approach via cognitive reading and writing user experiment tasks.

Keywords

Narrative Representation, Storytelling, Visualization

1. Introduction

Authors preserve their rhetoric, creativity and knowledge in stories. News writing falls under the genre of storytelling [1]. The story planning of the news material before writing aids speed, accuracy and influence via controlled information flow [2]. Throughout centuries, ***Aristotle's Rhetoric*** has guided writers to create effective communication using ***Ethos***, ***Logos*** and ***Pathos*** [3] [4]. ***Ethos*** is the art of establishing authority on a document considering state-of-the-art knowledge regarding its topic. ***Logos*** is building logical argumentation to explain the authors' viewpoint of the topic. ***Pathos*** is an attempt to persuade the readers emotionally toward the intended goal to consider required actions. It is the rhetoric that brings cognitive structure to natural language documents[3]. Following this path of persuasive writing, professional authors use various structural story templates to plan the core message that needs to be conveyed to audiences. One such popular and successful structural story template of the 20th century is Joseph Campbell's ***"The Hero's Journey"*** [5]. It involves the main character going on an adventure, facing challenges, learning a lesson and winning with the new found knowledge
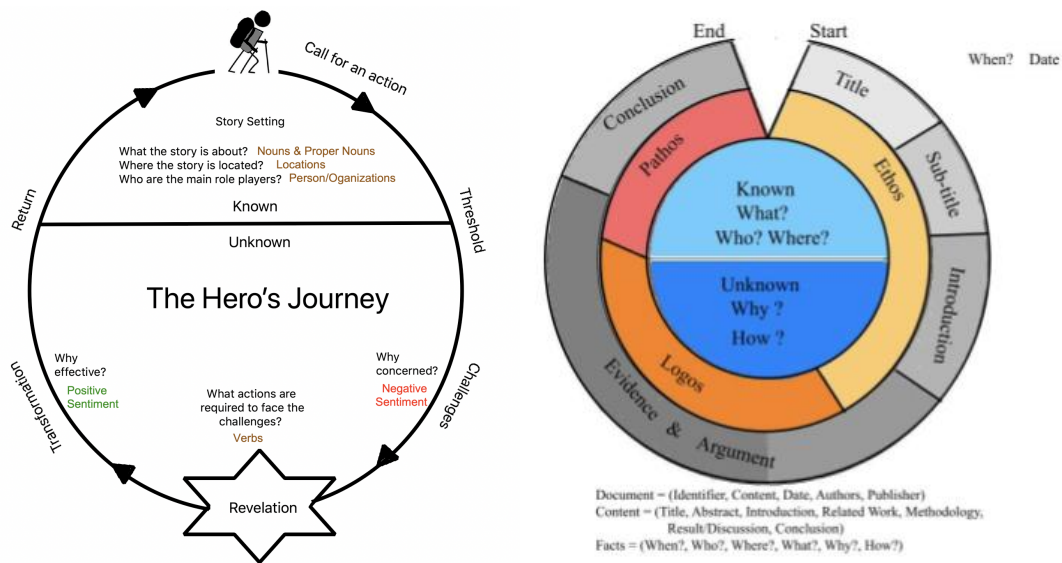
(a) *5Ws* mapping into *The Hero's Journey* template. (b)Sequential author's rhetorical mapping.

Figure 1: Distribution of the author's rhetorical view throughout various sections of news.

before returning home as presented in Figure 1 (a). To win readers' attention, a journalist must expose the **Who**, **Where**, **What**, **When** and **Why** of a news story consciously [6]. Missing any of the *5Ws* is referred to as **holes** in journalism [6]. This article theoretically mapped *5Ws* and *Aristotle's Rhetoric* into *The Hero's Journey* to experiment with the potential of story plan extraction as shown in Figure 1 (b). Based on this theoretical mapping we extracted story words from the news and visualized the words. The visualization was put into a user behavioural experiment to understand the cognitive richness of the proposed representation. This paper presents our cognitive reading and writing task-based experiment on 32 participants of university students, staff and teachers. It also compared text-based tasks against visualization-based tasks.

2. Related Work

Considering the speed and amount of new information along with the rise of fake news, fact-checking the news content and its rhetoric has become a big challenge during news analysis. To automatically predict the veracity of claims in news, researchers have been using techniques based on natural language processing, machine learning, knowledge representation and databases [7]. The authors of fake news aim to excite the sentiments of the readers towards the intended goal [8]. Therefore, to determine the polarity and strength of sentiments expressed in fake news, various knowledge-based, context-based and content-based sentiment analysis approaches are used to detect fake news [8]. Some news sentiment analysis systems assign scores indicating positive or negative opinions to each topic in the corpus using statistical analysis on sentiment cues [9]. Some systems use deep learning methodologies of recurrent
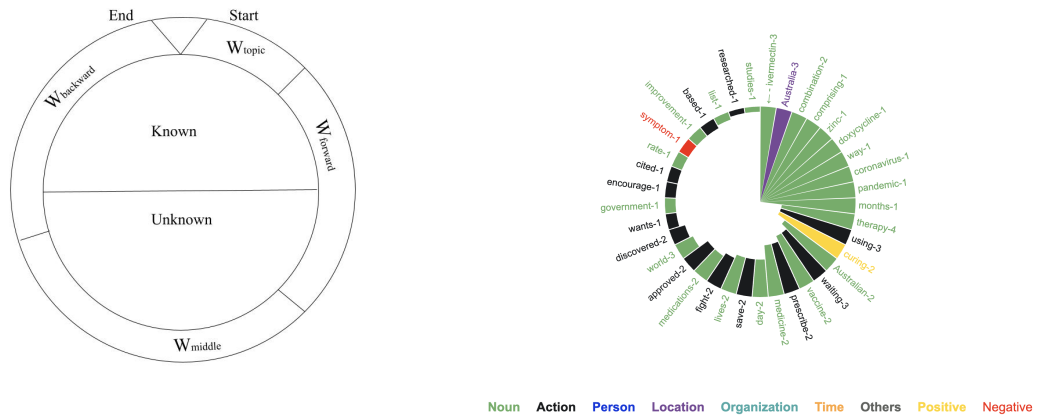
(a) Story words mapping based on the word categories (b) Sample story plan visualization of a sample news

Figure 2: Mapping and visualizing story words into the author's rhetorical mapping theory.

neural networks with long short-term memory units for forecasting from news archives [10]. News, articles and books convey authors's rhetoric of the author about the topics discussed in the document. In this paper, we employed story templates based on terms extraction to understand the rhetoric of the author and validated it using user experiments. This can be used in rhetoric-based fact-checking in future studies.

3. Theory

News, articles and books are a rich communication medium that preserves authors' rhetoric through **Ethos**, **Logos** and **Pathos** which triggers readers' cognitive thinking for learning. The facts of a document are encapsulated in six basic questions **Who**, **Where**, **What**, **When**, **Why** and **How**. In Figure 1 (b), we proposed a theoretical distribution of rhetorical questions along the sequential structure of news using **The Hero's Journey** template. Following this template diagram, we visualized the automatically extracted authors' plan in a clockwise circular pattern using the D3 circular bar plot [11] as shown in Figure 2 (b). The concept of known and unknown is also inspired by this template. This is a factual distribution encapsulating the answer to categorical facts of the news **What**, **When**, **Where** and **Who**. The answer to **Why** and **How** is correlated to the authors' argumentation and evidence. We also mapped **Aristotle's Rhetoric** into this format for the news as shown in Figure 2 (a). To establish **Ethos**, at the beginning of the document, from title to introduction, the author explains the topic/**What**, the problem area/**Where** and the main characters/**Who** related to the story. To establish **Logos** in the body of the document, the authors attempt to explain **Why** and **How**. For enriching/educating the reader, in the conclusion the author attempts to establish **Pathos** by stating the current situation of the event and by addressing future issues. Information about **When** of news is related to its date attribute.

Table 1
Scale of cognition

Criterion	Scale of cognition
Who	0-2 where 0 = wrongly understood ... 2 = well understood
Where	0-2 where 0 = wrongly understood ... 2 = well understood
What	0-2 where 0 = wrongly understood ... 2 = well understood
When	0-1 where 0 = wrongly understood, 1 = understood
Why	0-2 where 0 = wrongly understood ... 2 = well understood
Is summary interpretation true	0-1 where 0 = false, 1 = true
Quality of summary	1-5 where 0 = poor ... 5 = well written

We used an archive of pharmaceutical news from a website for analysis. The information extraction philosophy from a news document for this demo is based on our skimming technique. Word is the atomic unit of processing. This technique processes all sentences from top to bottom for extracting story words. We split news into M (5) blocks along document length and focus on multiple block appearances of selected words. Story words are extracted from the news based on four categories.

- W_{topic}: Most frequent N words that appear in all blocks.
- $W_{forward}$: Words that have the highest forward position weight. If a word appears earlier (based on sentence position) in the document it gets a higher weight.
- W_{middle}: Words that appear in more than M / 2 blocks.
- $W_{backward}$: Words that have the highest backward position weight. If a word appears later(based on sentence position) in the document it gets a higher weight.

Figure 2 (a) displays how the four categories of words are assembled for visualization using a circular bar chart and Figure 2 (b) demonstrated an example news.

The radial bars show the forward weight of the story words measured from the centre. We have classified the positive and negative nouns based on two static lists of words. The categorical classification of persons and locations came from the categorical information provided by Google Knowledge Graph API [12].

4. Cognitive reading and writing experiment

The theoretical mapping is evaluated via an online participation-based controlled experiment. We scored understanding factors of readers' cognition based on comprehension tasks using a homogeneous group of 32 participants. We followed the within-group experiment design. Each participant was given four comprehension tasks. The order of the tasks was generated using ***Latine Square Design***. Each participant responded to these questions: "***When*** did the incident take place?", "***Who*** are the main character(s)/role player(s) of the story?", "***Where*** did the story take place?", "***Why*** is the story important?", "Write a ***summary*** of the story in a few sentences" and "***Ease of comprehension***". We invited 3 academic reviewers to blindly score the comprehension tasks. Apart from ease of the task and task time scores, the rest of

Table 2
Experiment results

Criterion	Text(mean)	Visualization(mean)	P-value	Hypothesis testing with p = 0.05
Who	1.31	1.03	0.0175619	Reject null hypothesis
Where	1.80	1.47	0.0000929	Reject null hypothesis
What	1.63	1.53	0.2633649	Can't reject null hypothesis
When	0.66	0.55	0.1820127	Can't reject null hypothesis
Why	1.31	1.28	0.7512205	Can't reject null hypothesis
Is summary interpretation true	0.95	0.78	0.0019375	Reject null hypothesis
Quality of summary	3.03	2.39	0.0001664	Reject null hypothesis
Completion time	8.53 minutes	7.35 minutes	0.0182290	Reject null hypothesis
Ease	3.84	2.68	0.0000003	Reject null hypothesis

the questions were scored by the reviewers based on model answers and following the scales in Table 1. We performed a paired t-test on the average scores of text-based comprehension tasks against the average scores of visualization-based comprehension tasks. The result is reported in Table 2. According to Table 2 we have achieved 95% confidence in the reported result on all criteria apart from **What**, **When** and **Why**. The result demonstrates that the current state of representation is providing close cognition scores. This reveals the fact that within the context of visualization-as-summarization, the mapping offers a benefit due to the more compact representation.

5. Conclusion

An automatic understanding of authors' rhetoric can be extremely useful for comprehensive tasks like abstract summarization or strategic story plan visualization during the learning and teaching process. Topic models [13] help us to analyze the news based on **What** the news is about. Name entity recognition [14] systems and classifiers [15] can help us to analyze news based on **Where** and **Who**. News timeline helps us to analyze the **When**. Story plan templates like **Aristotle's Rhetoric**, **The Hero's Journey** and **5Ws** can aid extraction of the main story words for analyzing news by **Why** and **How** along with **Who**, **Where**, **When**, **What**. The evaluation of our theoretical mapping demonstrated a close human understanding of the compressed representation when compared to the whole text task results.

References

[1] A. Russell, Networked: A contemporary history of news in transition, Polity, 2011.

[2] M. L. M. L. Spencer, News Writing The Gathering, Handling and Writing of News Stories, Project Gutenberg, 2007.

[3] C. Rapp, Aristotle's Rhetoric, in: E. N. Zalta (Ed.), The Stanford Encyclopedia of Philosophy, spring 2010 ed., Metaphysics Research Lab, Stanford University, 2010. URL: https://plato.stanford.edu/archives/spr2010/entries/aristotle-rhetoric/.

[4] M. Meyer, Aristotle's Rhetoric, Topoi 31 (2012) 249–252. doi:10.1007/s11245-012-9132-0, place: Dordrecht Publisher: Springer Nature BV.

[5] Y. Cao, R. Klamma, M. Jarke, The Hero's Journey - Template-Based Storytelling for Ubiquitous Multimedia Management, Journal of Multimedia 6 (2011) 156–169. doi:`10.4304/jmm.6.2.156-169`.

[6] A. McKane, News writing, SAGE, London, 2006.

[7] Z. Guo, M. Schlichtkrull, A. Vlachos, A Survey on Automated Fact-Checking, Transactions of the Association for Computational Linguistics 10 (2022) 178–206. doi:`10.1162/tacl_a_00454`.

[8] M. A. Alonso, D. Vilares, C. Gómez-Rodríguez, J. Vilares, Sentiment analysis for fake news detection, Electronics 10 (2021) 1348.

[9] N. Godbole, M. Srinivasaiah, S. Skiena, Large-scale sentiment analysis for news and blogs., Icwsm 7 (2007) 219–222.

[10] W. Souma, I. Vodenska, H. Aoyama, Enhanced news sentiment analysis using deep learning methods, Journal of Computational Social Science 2 (2019) 33–46.

[11] Y. Holtz, Double circular barplot in d3.js, 2018. URL: https://www.d3-graph-gallery.com/graph/circular_barplot_double.html.

[12] Google Knowledge Graph Search API, 2021. URL: https://developers.google.com/knowledge-graph.

[13] W. Mu, K. H. Lim, J. Liu, S. Karunasekera, L. Falzon, A. Harwood, A clustering-based topic model using word networks and word embeddings, Journal of big data 9 (2022) 1–38.

[14] J. Li, A. Sun, J. Han, C. Li, A survey on deep learning for named entity recognition, IEEE transactions on knowledge and data engineering 34 (2022) 50–70.

[15] B. Muthu, S. Cb, P. M. Kumar, S. N. Kadry, C.-H. Hsu, O. Sanjuan, R. G. Crespo, A framework for extractive text summarization based on deep learning modified neural network classifier, ACM transactions on Asian and low-resource language information processing 20 (2021) 1–20.