Developing an Natural Language Processing Pipeline for Conflict Narrative Detection

Stephen Anning
spa1e17@soton.ac.uk
University of Southampton, UK

George Konstantinidis
g.konstantinidis@soton.ac.uk
University of Southampton, UK

Craig Webber
c.webber@soton.ac.uk
University of Southampton, UK

CCS CONCEPTS

• **Computing methodologies** → **Natural language processing**;
• **Applied computing** → **Law**; **Sociology**.

KEYWORDS

Peace Studies, Natural Language Processing, NLP, Cultural Violence, Self-Other Gradient, Hate Speech, Interdisciplinarity

ACM Reference Format:
Stephen Anning, George Konstantinidis, and Craig Webber. 2020. Developing an Natural Language Processing Pipeline for Conflict Narrative Detection. In . ACM, New York, NY, USA, 2 pages.

1 INTRODUCTION

Guided by theories of violence from Peace Studies, this research proposes a natural language processing (NLP) pipeline for "Conflict Narrative Detection". Related to this research is "Hate Speech Detection" that seeks to detect abusive language in online platforms. A 2019 review of this field by the Alan Turing Institute, however, finds, "at present, the data, tools, processes and systems needed to effectively and accurately monitor online abuse are not fully available and the field is beset with terminological, methodological, legal and theoretical challenges" [8]. Accordingly, to address these challenges, we seek to reconceptualise Hate Speech Detection as Conflict Narrative Detection with an NLP pipeline.

In conceptualising Conflict Narrative Detection, we are guided by sociological theory for technological design. By conflict narrative, we mean a narrative produced by an orator who intends to legitimise violence against another individual or group. Accordingly, we use "cultural violence" as a guiding theory, which seeks to understand the processes of violence legitimisation. In addressing the theoretical and methodological challenges of Hate Speech Detection, we propose a novel methodology for detecting and measuring cultural violence in a conflict narrative. This methodology then provides a framework for the proposed NLP pipeline. Accordingly the contributions of this research are as follows:

1. Propose the interdisciplinary theory of cultural violence for analysing the extent of violence legitimisation in a conflict narrative.
2. Propose a novel methodology for detecting and measuring cultural violence in natural language.
3. Make available reproducible experiments for assessing general-purpose and state-of-the-art NLP technologies [1].
4. Present a markup schema for modelling social groups in natural language.
5. Identify necessary modifications to existing technologies for NLP pipeline design.

To the best of our knowledge, this research constitutes the first attempt to use cultural violence for designing and developing an NLP pipeline. As such, this sociotechnical approach supports and continues interdisciplinary research carried out within Web Science.

2 CULTURAL VIOLENCE

The theory of cultural violence was first proposed in a 1990 paper by the Peace Researcher, Johan Galtung and is defined as, "those aspects of culture, the symbolic sphere of our existence – exemplified by religion and ideology, language and art, empirical science and formal science (logic and mathematics) – that can be used to justify or legitimise direct or structural violence" [1]. Direct violence refers to visible physical force, while structural violence refers to latent experiences of harm such as the imposition of economic measures resulting in poverty. Galtung proposes that direct or structural violence are legitimised "by changing the moral colour of an act from red/wrong to green/right or at least to yellow/acceptable" [1] by creating what Galtung describes as a "Self-Other gradient".

The Self-Other gradient refers to how an orator may use each aspect of culture to elevate a "Chosen People" (Self) and other those deemed "lower down the scale of worthiness" (Other) [1]. For example, with the phrase, "God bless America" George Bush uses a religious aspect of culture to elevate his ingroup of America with divine status. Also invoking religion is the collective noun, "axis of evil", he uses to collectively other his outgroup of Iran, Iraq and North Korea. Drawing upon social identity theory, we refer to the Self as an orator's ingroup, and the Other as their outgroup, with the distance between each known as "inter-group differentiation" [7]. What follows is a general hypothesis where the greater the differentiation between groups created by the Self-Other gradient, the more legitimate acts of violence against an outgroup becomes.

Cultural violence provides a unifying idea to enable interdisciplinary research using the six aspects of culture as a framework. From the discipline of evolutionary psychology, Martin presents a novel perspective to explain the role of religion and ideology in group dynamics [6]. He finds they serve to moderate violence in intra-group relations by solving five group problems. We use these five problems as a theoretical framework for the markup schema presented in this research. "Art" as representations of religion and ideology is found within the discipline of semiotics and is used by Kingdon to analyse the violent narratives of ISIS [5]. The role of

[1] GitHub Repo - CNA Pipeline

, ,

"empirical science" as status grievances suggests economic literature such as "Relative Deprivation" in "Why Men Rebel" by Ted Gurr [2]. The basis for "formal science" is from behavioural economics pioneered by Daniel Kahneman and Amos Tversky [4] to explain how people use religion and ideology to pre- or post-rationalise status grievances. Cultural violence tends to be an implicit feature of these disciplines, nevertheless, Galtung's broader ideas from peace research are often explicitly used. In addressing the theoretical challenges of Hate Speech Detection identified by the Alan Turing institute, cultural violence is an established interdisciplinary theory with a precedence derived from several academic disciplines.

3 METHODOLOGY

Derived from the theory of cultural violence, we have developed the following methodology for detecting a conflict narrative by measuring the Self-Other gradient. Measuring this gradient means firstly marking up the linguistic representations of named entities, religion and ideology in a text. The lexical relationship between each is then used to detect feature phrases that identify groups, elevation and othering statements. Finally, these phrases become qualitative units of a measurement schema to measure the Self-Other gradient. Four objectives summarise this process:

Obj 0. Pre-processing of text
Obj 1. Detect the ingroup and outgroup of an orator's text
Obj 2. Detect and classify feature phrases as either "ingroup elevation" or "outgroup othering".
Obj 3. Apply a schema to measure Self-Other gradient.

4 GROUP MARKUP SCHEMA

In the pre-processing of text, we employ named concept recognition. For marking up concepts in a text we have developed a theory-driven schema using "the five group problems" from Martin's "Why We Fight" as a framework. As shown in the GitHub repository [2], we use these problems as a theoretical framework to categorise synonymous concepts that represent aspects of culture for different group types. For example, the concept of *enemy*, synonymous with *invader* and *occupier*, is categorised by the military attribute of *adversary*. Having marked up relevant concepts in a text, NLP is then used to determine their lexical relationship with named entities for identifying the narrative's feature phrases.

5 TECHNICAL CHALLENGE

In developing this methodology, we tested general-purpose and state-of-the-art sentiment analysis and topic modelling technologies. As representations of conflict narratives, we curated a dataset comprising Hitler's *Mein Kampf* along with political speeches from George Bush and Osama bin Laden during the "War on Terror". In how he advocated for non-violent change, speeches from Martin Luther King provide control data. Surprisingly, we find neither existing general-purpose nor state-of-the-art NLP technologies are able to differentiate between Hitler's *Mein Kampf* and Luther King's *I Have a Dream*, which represent extremes of negative and positive sentiment [3]. Such technologies do not seem fit for the purpose of conflict narrative detection.

A general theme emerging from these experiments is a tension between the employment of statistical and pattern-based NLP technologies. The field of NLP began in the 1950s using pattern-based systems and has since evolved to employ sophisticated statistical methods requiring large datasets to generate statistically significant results [4]. Statistical method is appealing as it enables rapid development to scale. Nevertheless, our experiments show how quantitative outputs generated by these technologies - regardless of technical sophistication - tend to produce uncertain meaning.

While being more labour intensive to create, we find pattern-based methods are necessary for conflict narrative detection. Tests using Hearst patterns, developed by Prof Marti Hearst, have shown the potential for objective 1 of the methodology [5]. Hearst patterns enable algorithms, "to discover a hyponymic lexical relationship between two or more noun phrases in a naturally-occurring text" [3]. Hyponymic relations are said to exist when one noun phrase categorises another in a parent-child relation linked by predicate phrases. For example, the phrase, "America has enemies, such as Al Qaeda and the Taliban", the hypernym, *enemy* is linked to the hyponyms *Al Qaeda* and *the Taliban* by the predicate phrase *such as*. Moreover, the markup schema provides the means to annotate such hyponyms of the hypernym *enemy* as an outgroup. In contrast to quantitative outputs, such qualitative methods have the potential to provide more meaningful results.

6 CONCLUSION

In developing an NLP pipeline for Conflict Narrative Detection, a synthesis between social and technical aspects of NLP forms the core of our interdisciplinarity and is matched with a guiding philosophy beginning with first principles before employing advanced techniques or technologies. In doing so, we present a new methodology for detecting conflict narratives in natural language and a series of reproducible experiments to determine suitable NLP technologies for this methodology. A consistent theme emerging from our experimentation is the need to complement statistical method with pattern-based approaches to meaningfully process natural language. Accordingly, in the detection of online abuse, our ambition is to augment the quantitative aspects of existing NLP technologies with qualitative outputs that represent the patterns of language used to legitimise violence.

REFERENCES

[1] Johan Galtung. 1990. Cultural Violence. *Journal of Peace Research* 27, 3 (1990), 291–305.
[2] Ted Robert Gurr. 2015. *Why Men Rebel.* Routledge.
[3] Marti A Hearst. 1992. Automatic Acquisition of Hyponyms From Large Text Corpora. In *Proceedings of the 14th conference on Computational linguistics-Volume 2.* Association for Computational Linguistics, 539–545.
[4] Daniel Kahneman. 2011. *Thinking, Fast and Slow.* Macmillan.
[5] Ashton Rebecca Kingdon. 2017. Seductions of the Caliphate: A Cultural Criminological Analysis of Online Islamic State Propaganda. *Internet Journal of Criminology* (2017).
[6] Mike Martin. 2018. *Why We Fight.* Oxford University Press.
[7] Henri Tajfel, John C Turner, William G Austin, and Stephen Worchel. 1979. An Integrative Theory of Intergroup Conflict. *Organizational identity* 56 (1979), 65.
[8] Bertie Vidgen, Helen Margetts, and Alex Harris. 2019. *How much online abuse is there? A systematic review of evidence for the UK Policy Briefing – Full Report.* Technical Report. The Alan Turing Institute, London.

[2] Experiment 3 - Developing the Schema of Group Ideologies
[3] Experiment 1 - Testing NLP Technologies for Hate Speech Detection
[4] Goldberg, Yoav (2019) Keynote Speech, spaCy IRL, Berlin
[5] Experiment 7 - Dependency Matcher