ISABELLE VAN DER VEGT, BENNETT KLEINBERG & PAUL GILL

LINGUISTIC THREAT ASSESSMENT: CHALLENGES AND OPPORTUNITIES

Large-scale linguistic analysis may help security practitioners in making sense of violent and extreme communications.

Threat assessment generally involves the process of gathering information after a threat has been made to understand the risk of violence posed by a person. Usually, a threat will have been uttered in the form of verbal or written language. Nowadays, security professionals are confronted with assessing violent and extreme language on a large scale online. In light of these developments, we have been examining the application of computational linguistics to the study of grievance-fuelled targeted violence, including terrorism and mass murder. We call this approach 'linguistic threat assessment', in which our focus lies upon its computational implementation. This article highlights our main findings and the challenges and opportunities of this approach.

LINGUISTIC AREAS OF INTEREST

In our application of computational linguistics methods to the understanding of grievance-fuelled communications, we interviewed thirteen threat assessment professionals (with an average 18 years of experience) about their approach to anonymous threatening communications. Participants all read the same anonymous threat letter and subsequently discussed how they would assess the case. Although practice differed greatly between professionals — such as the cues paid attention to and the conclusions drawn from it — the responses in which linguistic information was used for assessment could be summarised as belonging to one of three areas of language, namely:

1. **Linguistic content:** *what* are people writing, i.e., in terms of word frequencies.
2. **Linguistic style:** *how* are people writing, i.e., in terms of grammar.
3. **Linguistic trajectories:** how does content and style develop over time.

Consequently, we leveraged these different areas of language for the study of grievance-fuelled communications. For example, we examined linguistic style in a study on abuse directed at politicians to infer gender gender, age, and personality traits based on language use in written abuse. Although we discovered some interesting gender and personality differences in the way participants wrote, the error margins for determining these traits based on language use alone were large, which means actionable predictions are difficult.

COMPUTATIONAL LINGUISTICS

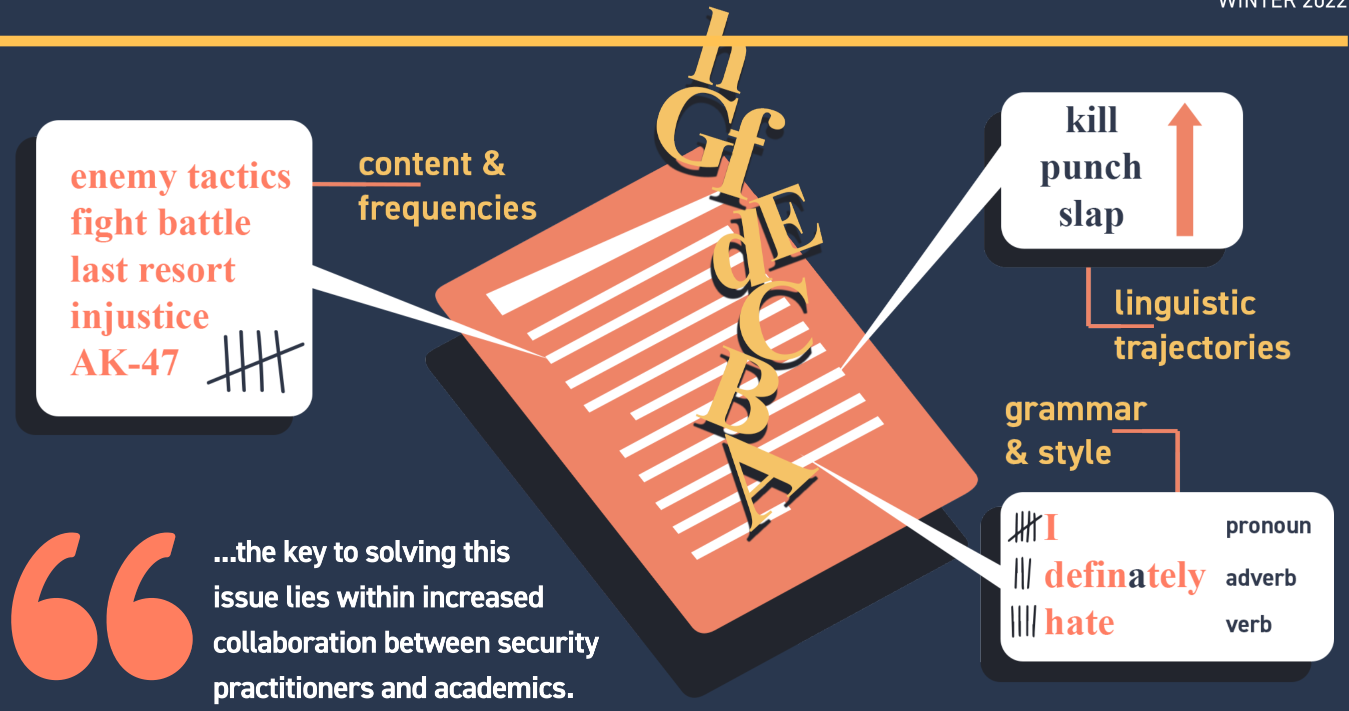
The branch of linguistics in which the techniques of computer science are applied to the analysis and synthesis of language and speech. *Oxford English Dictionary.*

We have also demonstrated the utility of measuring language over time (i.e., linguistic trajectories) to assess the effects of external events on an extremist group and the evolution of language on a far-right forum. Of particular relevance to security professionals is perhaps our study on the development of the 'Grievance Dictionary', which puts emphasis on the linguistic content (i.e., *what* someone conveyed).

THE GRIEVANCE DICTIONARY

The Grievance Dictionary is a tool specifically developed to analyse grievance-fuelled and/or threatening language at scale. It makes use of word frequencies to measure different (psychological) concepts in text. It is similar to the LIWC dictionary, which can measure a wide variety of psychological (e.g., friendship, sadness) and linguistic concepts (e.g., pronouns, swear words), but is specifically focussed on grievance-fuelled communications. Again, we started with consulting expert threat assessors (similar sample as stated above) and asked what they look for in a text when they assess a potential threat of violence.

From that expert exercise, we established 22 categories that make up the Grievance Dictionary, which includes categories such as weapons, murder, desperation, and planning. Next, we generated wordlists representative for each category and tested their validity using an online rating task, in which 2,318 participants on crowdsourcing platform Prolific assessed the 'goodness of fit' of 20,502 words for these categories. In applying the dictionary to measure the aforementioned 22 concepts, we saw marked differences between different text samples. For instance, we saw that lone-actor terrorist texts scored higher on all but one measure (especially murder, soldier, and weaponry) when compared to right-wing extremist forum posts. The only category on which these samples did not differ, was our measure of loneliness.



> ...the key to solving this issue lies within increased collaboration between security practitioners and academics.

These first analyses using the Grievance Dictionary demonstrate how it can be used to analyse large volumes of text, for instance in the case of a lengthy manifesto or an entire forum. In essence, these large volumes of text are condensed down into 22 comprehensible measures that are relevant to security professionals or researchers dealing with grievance-fueled violence. These measures can subsequently be integrated into a broader assessment of an individual or group of individuals, or can be used for research purposes in which different types of authors (e.g., different ideologies, violent vs. non-violent) are compared on Grievance Dictionary measures.

CHALLENGES AND OPPORTUNITIES

One challenging issue within the field of linguistic threat assessment is access to data. Targeted violence is a low base rate phenomenon, and the number of cases where the perpetrator produced linguistic material related to an incident will be even smaller. It is common procedure within this field to make use of lone-actor terrorist manifestos to better understand violent language use, as it is known these authors committed an act of violence. However, the sample size of lone-actor terrorist manifestos is small (our database counts approximately 25).

These manifestos are often compared to a larger sample of neutral, non-violent texts to assess linguistic differences. In doing so one of the main questions within this field remains unanswered, (which is what we are perhaps most interested in discovering), namely, which linguistic markers set apart a violent text written by an individual with violent intent, from an individual without such intent. That is, we want to know what — linguistically — sets apart the actualisers from the non-actualisers. Are there specific Grievance Dictionary categories that significantly differ between these groups? At present, we do not know because we do not have the data to study these questions.

When using extremist forum data, we simply do not know whether the individuals behind a post were in fact violence actualisers or not. In other words, the ground truth behind the data is not available to us. One notable recent initiative includes the use of a former extremist in order to identify the violent from the non-violent extremists on a forum. However, apart from this one paper, we believe the key to solving this issue lies within increased collaboration between security practitioners and academics . We expect that police or security practitioner databases contain a multitude of communications, which were initially seen as violent or extreme, and subsequently did or did not lead to violence.

Linguistic analysis of such data will be incredibly valuable for our understanding of (possible) links between violent language and behaviour. By sharing data, we can continue to increase our understanding of violent language and thereby further the field of linguistic threat assessment.

Dr Isabelle van der Vegt is an honorary research associate at the Department of Security and Crime Science at University College London and a scientific project manager at the Research and Documentation Centre for the Dutch Ministry of Justice and Security.

Bennett Kleinberg is an assistant professor at the Department of Methodology and Statistics at Tilburg University and an honorary associate professor at the Department of Security and Crime Science at University College London.

Paul Gill is Professor of Security & Crime Science at University College London.