Twitter rape threats and the Discourse of Online Misogyny (DOOM): using corpus-assisted community analysis (COCOA) to detect abusive online discourse communities

Mark McGlashan
Lancaster University
m.mcglashan@lancaster.ac.uk

Claire Hardaker
Lancaster University
c.hardaker@lancaster.ac.uk

1 Introduction

In July 2013, Caroline Criado-Perez successfully campaigned to have a woman appear on an English banknote, and was subsequently inundated with extreme misogynistic abuse on Twitter. Over the following days, this escalated, and more prominent women were sent death and bomb threats. Whilst legislative bodies came under intense pressure to handle the issue, there is a lack of research into such behaviour, making evidence-based solutions difficult. As a result of an ESRC urgency grant dedicated to investigating this case, this presentation outlines the project and some findings with regards to the methodological background to finding rape threat trolls and identifying the networks that they form, as well as some suggestions for future directions and projects.

2 Data and sampling

The work focuses on a database of interactions with the @CCriadoPerez Twitter account that occurred over a period of 3 months in 2013. The sample was restricted to only tweets which were sent to (i.e. mentioned) or from (i.e. tweeted from) the @CCriadoPerez account during the period 25 June 2013 to 25 September 2013.

The sample was collected using Datasift (www.datasift.com), a commercial service that provides access to 'social data' from sources such as Twitter and Tumblr. The 'historics' functionality allows users, unlike other services, to search for and easily collect millions of historical online posts.

Our sampling frame returned 76,235 interactions involving @CCriadoPerez: 67,129 mentions of the account and 9,106 tweets from the account. Converted into a corpus, the language data in these tweets (excluding URLs, hashtags and mentions) amounted to a total of 1,014,222 words, which we call the Caroline Criado-Perez Corpus (CCPC).

3 Methods: Corpus-assisted Community Analysis (COCOA)

Corpus-assisted Community Analysis is a multimethodological approach to the study of online discourse communities that combines methods from Discourse Analysis, Corpus Linguistics and Social Network Analysis.

3.1 Corpus-driven Discourse Analysis

Drawing predominantly on the work of Baker (2006), we took a corpus-driven approach to analysing discourses in the language contained in the CCPC. The CCPC was explored and analysed using the AntConc corpus analysis tool (Anthony 2014). By taking a corpus-driven approach the analysis of linguistic units like lexis and n-grams, we were able to gain an overall view of frequent topics and discourses that were present in conversations contextualized by a period of sustained online abuse focused on an individual online user.

3.2 Social Network Analysis

At the beginning of the research, we were informed by Caroline Criado-Perez, the target of the online abuse in question, about a number of threating and offensive online communications that she had received (and which initiated this research). In informing us, she disclosed the screen names of all of the accounts that she was aware of that were sending her abuse. As such, we began the research with a seed list of account names of abusive users.

Using these names, we were able to search the database of 76,235 tweets for all of the interactions in which those abusive users identified by Criado-Perez occurred (either as someone who tweets or is mentioned) and to populate a larger sample of abusive users which we categorise into two groups: high-risk users (n=61) and low-risk users (n=147). High-risk users exhibited behaviours such as, intent to menace, harassment, and sexual aggression. Low-risk users tended to make misogynistic or generally insulting remarks but were not considered threatening.

4 Corpus-assisted Community Analysis (COCOA): implementation and findings

Subsets of the database and CCPC were created based on finding from the implementation of methods discussed in 3.1 and 3.2 with regard to high-risk and low-risk users: a CCP high-risk corpus, CCP low-risk corpus, and CCP no-risk corpus. The CCP no-risk corpus is comprised of tweets from users not identified as being abusive.

4.1 Linguistic findings

Using keyword analysis, the language found in the CCP high-risk and CCP low-risk corpora was compared to that found in the CCP no-risk corpus which highlighted some of the key discourses in the talk of 'risky users', including:

- Rape: *rape, raep, raping*
- Misogyny: *cunt, bitch*
- Homophobia: *faggot, gay*
- Racism/anti-Semitism: *nigger, jew*
- Genitalia/anatomy: *pussy, penis, fucking, ass, cock*

Moreover, we found that the overall response by users of Twitter to the abuse received by Criado-Perez was to condemn abuse and to contest online abuse (especially misogynistic abuse).

4.2 Applying linguistic findings in detecting discourse communities

After having found what language was *key* about the CCP high-risk and CCP low-risk corpora, we were also interested in if 'risky' users associated, and how they associated if they did. We found that 'risky' users frequently associated with each other in directed networks (cf. Scott 2013) as well as through ambient affiliation (Zappavigna 2012). We assessed directed networks as existing when 'risky' users mention other 'risky' users in their tweets. Ambient affiliation (which can also be understood in terms of undirected networks (cf. Scott 2013)) was assessed as being instances in which 'risky' users talk about the same things regardless of whether they are known to each other or not.

We found that not only do 'risky' users mention each other (ergo, talking to one another and forming directed networks) but also collectively engage in the use of abusive language (relying on discourses outlined in 4.1) when targeting Criado-Perez.

5 Conclusions

We applied a combination of methods from Corpus Linguistics, Discourse Analysis and Social Network Analysis in order to find a community of users on Twitter who communally engage in the use of a range of offensive discourses in order to enact abuse online.

Corpus Linguistics enabled the detection of a number of discourses which were significant in the population of Twitter users that partook in a spate of online abuse.

Social Network Analysis enabled to us to populate a large sample of abusive users based on a smaller group of users identified as being abusive and to explore the networks and interactions in which 'risky' users were involved.

Our proposed combination of methods for Corpus-assisted Community Analysis (COCOA) gave us a framework for combining Corpus Linguistic and Social Network Analytical approaches to the analysis of online communities and brought together the findings of two different forms of analysis to provide a more complex and complete overview of the discursive and networking behaviour of 'risky' online users during a period of prolonged abuse in the context of Twitter.

Acknowledgements

This work was supported by the Economic and Social Research Council [grant number ES/L008874/1].

We would like to thank: Steve Wattam for his tireless help, engineering wizardry and endless cynicism; Uday Avadhanam for initiating him into the (horribly complex) world of R; Tony McEnery for all his help and patience.

References

Anthony, L. 2014. AntConc (Version 3.4.3.) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.antlab.sci.waseda.ac.jp/

Baker, P. 2006. *Using Corpora in Discourse Analysis*. London: Continuum.

Scott, J. 2013. *Social Network Analysis*. 3rd Ed. London: Sage.

Zappavigna, M. 2012. *Discourse of Twitter and social media*. London: Continuum.