AUTHOR QUERY FORM

ELSEVIER	**Journal:** PRAGMA **Article Number:** 4144	**Please e-mail your responses and any corrections to:** **E-mail:** corrections.esch@elsevier.thomsondigital.com

Dear Author,

Please check your proof carefully and mark all corrections at the appropriate place in the proof (e.g., by using on-screen annotation in the PDF file) or compile them in a separate list. Note: if you opt to annotate the file with software other than Adobe Reader then please also highlight the appropriate place in the PDF file. To ensure fast publication of your paper please return your corrections within 48 hours.

For correction or revision of any artwork, please consult http://www.elsevier.com/artworkinstructions.

Any queries or remarks that have arisen during the processing of your manuscript are listed below and highlighted by flags in the proof. Click on the 'Q' link to go to the location in the proof.

Location in article	Query / Remark: click on the Q link to go Please insert your reply or correction at the corresponding line in the proof
	The reference given here is cited in the text but is missing from the reference list – please make the list complete or remove the reference from the text: Cicourel (1969), Banks (2010), Vamialis (2013). Please complete and update the reference given here (preferably with a DOI if the publication data are not known): Hardaker (forthcoming). For references to articles that are to be included in the same (special) issue, please add the words 'this issue' wherever this occurs in the list and, if appropriate, in the text.
Q1	Please confirm that given names and surnames have been identified correctly.
Q2	Please check the telephone number of the corresponding author, and correct if necessary.
Q3	Please provide a maximum of 6 keywords.
Q4	"Your article is registered as a regular item and is being processed for inclusion in a regular issue of the journal. If this is NOT correct and your article belongs to a Special Issue/Collection please contact m.renaud@elsevier.com immediately prior to returning your corrections.".
Q5	Ref. Vamialis (2013) is cited in the text but not provided in the reference list. Please provide it in the reference list or delete this citation from the text.
Q6	Ref. Banks (2010) is cited in the text but not provided in the reference list. Please provide it in the reference list or delete this citation from the text.
Q7	Please check the placement of Tables 1, 5 and 9, and correct if necessary.
Q8	Ref. Cicourel (1969) is cited in the text but not provided in the reference list. Please provide it in the reference list or delete this citation from the text.
Q9	One or more sponsor names and the sponsor country identifier may have been edited to a standard format that enables better searching and identification of your article. Please check and correct if necessary.
Q10	Please update Ref. Hardaker (forthcoming).

Location in article	Query / Remark: click on the Q link to go Please insert your reply or correction at the corresponding line in the proof
	Please check this box or indicate your approval if you have no corrections to make to the PDF file ☐

Thank you for your assistance.

Available online at www.sciencedirect.com

ScienceDirect

Journal of Pragmatics xxx (2015) xxx–xxx

journal of PRAGMATICS

www.elsevier.com/locate/pragma

Highlights

Journal of Pragmatics: xxx (2015) pp. xxx–xxx

''Real men don't hate women'': Twitter rape threats and group identity

Claire Hardaker*, Mark McGlashan

- Analysis of a corpus of tweets.
- Rape threats trigger interesting (re)negotiation of masculinity (e.g. ''real men'').
- High- and low-risk communities use similar discourses.

Available online at www.sciencedirect.com

ScienceDirect

Journal of Pragmatics xxx (2015) xxx–xxx

journal of PRAGMATICS

www.elsevier.com/locate/pragma

''Real men don't hate women'': Twitter rape threats and group identity

Claire Hardaker *, Mark McGlashan

Lancaster University, UK

Received 26 October 2014; received in revised form 22 October 2015; accepted 10 November 2015

Abstract

On 24th July 2013, it was announced that feminist campaigner Caroline Criado-Perez's petition to the Bank of England to have Elizabeth Fry's image on the UK's £5 note replaced with the image of another woman was successful. The petition challenged the Bank of England's original plan to replace Fry with Winston Churchill, which would have meant that no woman aside from the Queen would be represented on any UK banknote. Following this, Criado-Perez was subjected to sustained misogynistic abuse on Twitter, a microblogging social network, including threats of rape and death. This paper investigates this increasingly prominent phenomenon of rape threats made via social networks. Specifically, we investigate the sustained period of abuse directed towards the Twitter account of feminist campaigner and journalist, Caroline Criado-Perez. We then turn our attention to the formation of online discourse communities as they respond to and participate in forms of extreme online misogyny on Twitter. We take a corpus of 76,275 tweets collected during a three month period in which the events occurred (July to September 2013), which comprise 912,901 words. We then employ an interdisciplinary approach to the analysis of language in the context of this social network. Our approach combines quantitative approaches from the fields of corpus linguistics to detect emerging discourse communities, and then qualitative approaches from discourse analysis to analyse how these communities construct their identities.

1. Introduction

This paper investigates an increasingly prominent phenomenon: rape threats made via social networks. Specifically, we investigate a sustained period of abuse directed towards the Twitter account of feminist campaigner and journalist, Caroline Criado-Perez. We then turn our attention to the formation of online discourse communities as they respond to and participate in forms of extreme online misogyny on Twitter. The abuse followed Criado-Perez's petition which challenged the Bank of England's decision to remove the image of Elizabeth Fry from the £5 note and replace it with that of Winston Churchill. The premise of the petition was to maintain the representation of influential women on British currency, since the appearance of men only could be deemed a ''damaging message that no woman has done anything important enough to appear [on our banknotes]'' (Criado-Perez, 2013). The petition was successful and the Bank of England announced on the 24th of July 2013 that author Jane Austen's image will appear on the new £10 note issued in 2016.

Following the petition, Criado-Perez began receiving an influx of abuse through her Twitter account (@CCriadoPerez), including threats of rape and murder, which were malicious and numerous enough to warrant police intervention. These

* Corresponding author at: Department of Linguistics and English Language, Lancaster University, Lancaster LA1 4YT, UK. Tel.: +44 1524 593212.
E-mail address: c.hardaker@lancaster.ac.uk (C. Hardaker).

http://dx.doi.org/10.1016/j.pragma.2015.11.005
0378-2166/

threats subsequently escalated to involve bomb threats of other prominent female figures, including Colleen Nolan, MP Stella Creasy and Professor Mary Beard. Following police and journalistic investigations, Twitter users Isabella Sorley, John Nimmo, and Peter Nunn who had each sent abuse to Criado-Perez, were eventually prosecuted and given custodial sentences. However, as shown below, there were many others who sent extreme and sustained abuse to Criado-Perez, yet faced no legal redress. The lack of consequence is due to many contextual and legal factors, including, but not limited to, complications arising from trans-national jurisdiction; inadequate legislation (e.g. the UK's Communications Act 2003); inadequate provision by internet service providers and online platforms; investigative bodies lacking the skills and/or resources to investigate new forms of illegal online behaviour, especially when combined with the ease with which users can remain anonymous online; and the sheer amount of abusive online behaviour that would overwhelm the legal system if every qualifying case were prosecuted. We return to some of these issues below.

In this paper, we address two key issues: (1) the language surrounding sexual aggression on Twitter, and (2) the emergence and construction of communities in response to that sexually aggressive language.

2. Computer-mediated communities

Computer-mediated communication (CMC) refers to human interactions occurring through the use of devices such as computers, tablets, and smartphones using formats including email, text messages, and tweets. Although we recognise the multimodal nature of many forms of web-based interaction, we focus here on primarily textual forms of CMC "involving typed words that are read on digital screens" (Herring and Stoerger, 2014:570). From this, we analyse interactions mediated through the social networking microblog, Twitter.

> Linguistic scholarship in the area of CMC is now well established. It began with descriptive accounts of CMC as it differed from other forms of linguistic communication but progressed swiftly onto analyses of politeness, conversational turn-taking, and sociolinguistic accounts of dialect, gender, social status, etc. and their influence on language use in CMC. (Herring et al., 2013)

As Herring suggests above, linguists are increasingly turning to social media and networking platforms such as Facebook and Twitter, since these can provide massive amounts of publicly and freely accessible, organically occurring, easily downloaded language data. When we turn to Twitter specifically, we find that it facilitates many kinds of interaction, and that it is used for a wide range of purposes, such as keeping in touch with friends, sharing multimedia, consuming news, advertising cottage industries, engaging with voters, and gathering real-time consumer feedback. As a public-facing social network (unlike social networks designed for private interaction, e.g. Facebook), Twitter provides a space for debate, humour, updates, news, products, gossip, and more besides.

The result of this is that online networks offer many beneficial and unique opportunities, such as education, companionship, and current affairs news. However, users may also come into contact with (or become engaged in) behaviours that pose risks to their personal wellbeing, safety, and security. Issues such as online grooming, cyberharassment, predation, e-fraud and so forth have become a real online threat (see Hardaker, forthcoming),[1] but have also transgressed into the offline world. Reports of suicides linked to cyberbullying and harassment are on the rise, and it is these latter types of antisocial online behaviour—behaviour that poses a risk to others (i.e. 'risky behaviours') that this paper is interested in.

2.1. Online and offline identities

Within academia, offline identity has received considerable attention in fields as diverse as gender, im/politeness, sociolinguistics, and pragmatics (e.g. Boxer and Cortés-Conde, 1997; Cameron, 1997; Edwards, 1998; Holmes, 1997; Mullany, 2007; Terkourafi, 2005; Verschueren, 2004). However, identity may well be an analytic fiction (Simon, 2004). It is not a 'thing', nor a purely cognitive phenomenon. Instead, just as dancing is a dynamic physical process that only becomes apparent when undertaken, identity is a dynamic behavioural, socio-psychological enactment carried out through relational interaction with others (O'Brien, 1999:78).

Identity is sometimes simplistically discussed in terms of two (artificially dualistic) categories: *individual identity*, or one's self-definition as a person in one's own right, and *collective identity*, or one's self-definition as a person in relation to one's group memberships. These categories help to define each other, however:

[1] Because this paper interests itself with antisocial online behaviour, the focus is on the negative side of social networks. It is worth noting, however, that these same sites have enabled extraordinary acts of kindness, charity, and selflessness. The Internet does not cause users to behave kindly or cruelly. It simply facilitates their own choice of behaviour.

Please cite this article in press as: Hardaker, C., McGlashan, M., "Real men don't hate women": Twitter rape threats and group identity. Journal of Pragmatics (2015), http://dx.doi.org/10.1016/j.pragma.2015.11.005

> The same self-aspect (e.g. German) can provide the basis for a collective identity at one time ('We, the Germans'), whereas at another time it may be construed as a constituent or element of one's individual identity ('I am a psychologist, male, German, have brown eyes and so forth'). In the first case the particular self-aspect defines a social category of which oneself is one member among others, whereas in the other case it is one feature among several other features of oneself, the ensemble of which constitutes one's individual identity. (Simon, 2004:54)

Bucholtz and Hall (2005) offer a far more nuanced approach to identity by drawing on research from social psychology (e.g. Giles et al., 1991; Meyerhoff, 1996; Tajfel and Turner, 1986), linguistic anthropology (e.g. Ochs, 1992; Silverstein, 1976, 1979, 1985), and sociolinguistics (e.g. Eckert and Rickford, 2001; Le Page and Tabouret-Keller, 1985; Mendoza-Denton, 2002). From this, they determine that,

> [i]dentity does not emerge at a single analytic level—whether vowel quality, turn shape, code choice, or ideological structure—but operates at multiple levels simultaneously. Our own approach privileges the interactional level, because it is in interaction that all these resources gain social meaning. (Bucholtz and Hall, 2005:586)

Identity is a conscious and unconscious patchwork of what the individual conveys (e.g. a troll tweets ''did you forget how to sammich?'' at a feminist), what others ascribe to the troll (e.g. that they are conveying outdated, patriarchal notions of women), and the result of interactional negotiations (e.g. multiple other individuals agree that the offending user is trolling) (Bucholtz and Hall, 2005:605–607).

Overall, for Bucholtz and Hall (2005), identity is produced intersubjectively and across multiple dimensions, rather than individually. It also emerges and circulates in interaction rather than simply being assigned a priori (2005:587). The work of Bucholtz and Hall, amongst others, demonstrates the rich and growing body of research into offline identity, however, online identity research, especially surrounding mutability, is still catching up. Face-to-face, judgements about others may be made instantly, based on appearance, behaviour, and speech, but via computer-mediated communication (CMC) users have far more control over self-presentation:

> The potential for constructing alternative identities is one of the most salient features of Internet use. In face-to-face interaction restrictions are placed on the identity a person is able or permitted to construct for themselves at that particular point in time; for example, people cannot instantly change their physical appearance at will. However, as Reid (1994) notes, the anonymity and physical separation of cyberspace enables social experimentation, as well as explorations of identity and self. (Baker, 2001)

Users can invent and explore identities that they would struggle to enact convincingly, if at all, offline. As discussed throughout, however, other users seem to automatically equate discrepancies between online and offline identities as perniciously motivated attempts at deception. This returns to the issue that whilst academic research may discuss theoretical perspectives of the mutability and multifaceted nature of identity, for lay users, the interpretation may be far more simplistic.

2.2. *Anonymity and disinhibition*

Anonymity has been a facet of published content since before the inception of the printing press. Authors have long concealed their identities to express unpopular opinions or make available a text that represents a form of dissent. However, anonymity in the context of CMC presents ethical and legal quandaries, and, like identity, the notion of anonymity is also not clear-cut.

In simple terms, anonymity is generally understood as a state of being unidentifiable:

> One has anonymity or is anonymous when others are unable to relate a given feature of the person to other characteristics. (Wallace, 1999:24)

Full anonymity—becoming *unknown* in the sense that any traces of a person's possible known identity, including name, location, age and so forth cannot be related to them—occupies the most extreme point on a cline between full identity disclosure through to full anonymity (Zarsky, 2004:1340). The anonymity that CMC can facilitate is noted as being both potentially beneficial and detrimental to individuals and society, however, our focus here is on those that abuse anonymity for the purposes of causing others online distress without repercussion.

Anonymity can foster a sense of impunity, loss of self-awareness, attitudinal polarisation, and a likelihood of acting upon normally inhibited impulses—an effect known as deindividuation (Siegel et al., 1986). Indeed, group members may not be ''seen or paid attention to as individuals'' by users (Festinger et al., 1952:382), but instead perceived of as an homogenous mass, in turn weakening the user's perceptions of both individuality and of personal responsibility and liability (Diener, 1979). Additionally, users may experience a sense of disinhibition such that they become willing to express opinions online that they would never voice if they knew that those opinions could be attributed to them offline

Please cite this article in press as: Hardaker, C., McGlashan, M., ''Real men don't hate women'': Twitter rape threats and group identity. Journal of Pragmatics (2015), http://dx.doi.org/10.1016/j.pragma.2015.11.005

Table 1
Interaction types on Twitter.

Interaction type	Function
Tweet	An online post made by a Twitter user.
Mention	A includes B's username in their tweet, e.g. ''Hello @CorpusSocialSci!'' B is notified of this.
Retweet	A re-posts B's tweet, so that A's followers can see it. B is notified of this. Note retweets can expand the tweet's audience far beyond that originally intended.

(Vamialis, 2013:32). And psychologically, users may give less consideration to the recipient's feelings. This, according to Douglas and McGarty (2001:399), is manifested in behaviours like flaming and trolling. As described by Vinagre:

> Sometimes people share very personal things about themselves. They reveal secret emotions, fears, wishes. They show unusual acts of kindness and generosity, sometimes going out of their way to help others. We may call this benign disinhibition. However, the disinhibition is not always so salutary. We witness rude language, harsh criticisms, anger, hatred, even threats. Or people visit the dark underworld of the Internet—places of pornography, crime, and violence—territory they would never explore in the real world. We may call this toxic disinhibition. (Vinagre, 2008:321)

Further, this high degree of anonymity within CMC can offer far more control over one's self-presentation than face-to-face. As such, the possibility of deception is greatly increased, whether intentional or accidental, or self- or other-imposed (Preece, 2000; Rheingold, 1993; Spears and Lea, 1992). When we add to all of this our ability to reach a diverse worldwide audience comprised many thousands of cultures, it is little surprise that online conflict is commonplace (Baker, 2001). Q6 Indeed, when we consider escalated, criminal forms of online conflict,

> . . .the anonymity and mobility afforded by the Internet has made harassment and expressions of hate effortless in a landscape that is abstract and beyond the realms of traditional law enforcement. (Banks, 2010:238)

3. Data and method

This study examines a corpus of Twitter data that involves interactions of the Twitter account of Caroline Criado-Perez (@CCriadoPerez). Q7 The sample is made up of three kinds of interactions made possible by the Twitter platform.[2] These are shown in Table 1.

The sample spans 92 days of activity, from midnight 25/06/13 to midnight 25/09/13 inclusive. The period was selected by identifying the date that Criado-Perez first highlights an instance of abuse directed towards her (25/07/2013) regarding her successful petition:

Example 1: ''Tweet zero''.

User	Date/Time	Tweet
JackRiley92	25/07/13 15:35	@CCriadoPerez are you the sad bitch that's running a campaign to have more women on banknotes???

The tweet initially creates no further interactions (replies, retweets, favourites, etc.) until Criado-Perez replies to it the following day:

Example 2: Criado-Perez response.

User	Date/Time	Tweet
CCriadoPerez	25/07/13 12:00	Are you the sad twerp who takes time out of his day to track down strangers to abuse them? Enjoy your life I guess. . . @JackRiley92

[2] Interactions are limited to 140 UTF-8 characters, but can also include pictures, hashtags, email addresses, and hyperlinks. Due to wordcount limitations, these other features are not considered in this paper.

Please cite this article in press as: Hardaker, C., McGlashan, M., ''Real men don't hate women'': Twitter rape threats and group identity. Journal of Pragmatics (2015), http://dx.doi.org/10.1016/j.pragma.2015.11.005

Table 2
Sampling criteria for each Twitter interaction type.

Interaction type	Function
Tweet	Any tweets from @CCriadoPerez
Mention	Any tweets in which @CCriadoPerez is mentioned
Retweeting	Any retweets by @CCriadoPerez
Retweeted	Any @CCriadoPerez tweets that are retweeted

Table 3
Size of the CPTMC (sizes by number of tweets).

Interaction type/subcorpus	Month 1	Month 2	Month 3	Totals
Mentioned (@CCriadoPerez account is tweeted by another account)	5166	53,768	8,195	67,129
Tweeting (@CCriadoPerez account tweets another account)	2746	4646	1714	9106
Totals	**7912**	**58,414**	**9909**	**76,235**

Within the events of the abuse sent to Criado-Perez, the tweet by @JackRiley92 effectively stands as ''tweet zero'' (from the medical parlance of ''patient zero''—the first individual infected with a contagion that becomes an epidemic). Extrapolating outwards from this, a sample was taken for a full calendar month prior to this date to examine whether there was a history of abuse in the short term and for two full calendar months following this date to investigate how the abuse unfolded. Aside from dates, additional sampling criteria (see Table 2) were used to capture all instances of direct interaction occurring in relation to the @CCriadoPerez account, and this resulted in the Criado-Perez Complete Corpus (or CPCC).

For the purposes of this paper, less direct forms of interaction such as retweets and favourites (where a user marks a posted tweet as a favourite) were excluded from the CPCC. The results of this sampling procedure yielded the Criado-Perez Tweets & Mentions Corpus (henceforth, CPTMC) totalling 76,235 tweets. These tweets were divided into four subcorpora as detailed in Table 3.

For every tweet made on Twitter, metadata is recorded which contains a number of attributes—or properties—enabling a range of possibilities for analysis. These attributes are as follows:

Account level	
Screen Name	A user-defined unique identifier, e.g. @DrClaireH, @Mark_McGlashan, etc.
Username	A user-defined name associated with the screen name, e.g. Claire Hardaker, Mark McGlashan, etc.
Description	A short, optional biography

Tweet level	
Date/Time	The data and time that a tweet was posted ('sent')
Text	The content of a tweet
Geo	The geographical location from which a tweet is sent (NB. unreliable)
Hashtags	A list of all of the hashtags included in a tweet
Links	A list of all of the webpage links included in a tweet
Mentions	A list of all of the screen names a user has included ('mentioned') in a tweet
Friends Count	The total number of users that an account follows at the time a tweet is sent
Statuses Count	The total number of tweets a user has sent at the time a tweet is sent

The focus of this analysis is on the **Text** attribute (though where relevant, data from other attributes has been retrieved throughout the analysis). To construct the CPTMC from the CPCC, the Text attribute was isolated, stripped of all

Please cite this article in press as: Hardaker, C., McGlashan, M., ''Real men don't hate women'': Twitter rape threats and group identity. Journal of Pragmatics (2015), http://dx.doi.org/10.1016/j.pragma.2015.11.005

hashtags, links, and mentions, and made readable for use with a concordance tool. This left a corpus of 76,235 tweets, totalling 1,014,222 words, and for the purposes of this study, we used AntConc version 3.4.2m.

3.1. Corpus linguistics and discourse analysis

Corpus linguistics (CL) can be viewed as a methodological approach or set of procedures oriented towards the study of language (Baker, 2014:7; McEnery and Hardie, 2012:1), in particular large collections of language data, or *corpora*. Although often misperceived as being ''a purely quantitative approach'' to linguistic analysis (Baker, 2014:7), increasingly sophisticated methodological 'synergies' (Baker et al., 2008) drawing on CL methods are being formulated and formalised to address research agendas in traditionally qualitative fields such as gender and language (Baker, 2014) and stylistics (Mahlberg, 2013).

With regards to discourse analytic (DA) approaches to the study of language—through which discourses are argued to be 'social practices' informed by ideology—language is first and foremost a 'way in' to observing and analysing the ideologies (such as racism, sexism, feminism, patriotism) that inform its use. Hybrid methodologies for DA including CL are becoming increasingly formalised in approaches such as corpus-assisted discourse studies (or, CADS) and corpora are recognised as a useful source of data for the purposes of *triangulation* (Cicourel, 1969). This can involve testing hypotheses or comparing findings from qualitative analysis of a particular language variety against quantified observations in reference corpora or other comparable specialised corpora (Baker, 2006:15–17). Discourse studies can employ CL to uncover systematic linguistic practices that realise the structural relationships between ideology and language (Baker et al., 2008).

CL is beginning to show that the study of corpora is more than just about statistics and quantitative generalisation. Combined CL and discourse analytical approaches have successfully offered qualitative insights into large amounts of language data, such as in the study of Islamophobia (Baker, 2010; Baker et al., 2008, 2013).

There is no single way to perform CL analysis for the purposes of DA; however, there are several analytical approaches common to all implementations of CL. Here, we focus on frequency, collocation/n-grams, and keywords. Frequency, a fundamental in CL research, is ''a simple tallying of the number of instances of something that occur in a corpus'' (McEnery and Hardie, 2012:49), and can be used to infer how frequently language recurs in a corpus. Unusually high or low recurrence can be of intrinsic interest, however, as Baker argues, whilst useful, the functionality of frequency counts is limited:

> Their main use is in directing the reader towards aspects of a corpus or text which occur often and therefore may or may not show evidence of the author making a specific lexical choice over others. (Baker, 2006:68)

In other words, frequent repetition may not give insight into the kinds of discourses that exist in a corpus. For this, a DA approach requires *context*. Collocation analysis is one method of analysing linguistic context and meaning. The notion of collocation, denotes the idea that important aspects of the meaning of a word (or other linguistic unit) are not contained within the word itself, considered in isolation, but rather subsist in the characteristic [linguistic] associations that the word participates in. (McEnery and Hardie, 2012:123)

Generally, these ''characteristic associations'' refer, in their broadest sense, to ''two or more words which have a tendency to be used together'' (Cantos Gómez, 2013:196). Some forms of collocation are so strong and stable that they become what are referred to as n-grams or lexical bundles (Biber et al., 2004).

Finally, keyword analysis is performed by comparing a frequency wordlist generated from one corpus against a frequency wordlist of another corpus, allowing the observation of words that are statistically most and least frequent. Such words are referred to as positive or negative keywords and, unlike a frequency wordlist, positive keywords allow the analysis of linguistic saliency rather than simple frequency (Baker, 2006:125).

3.2. Ethical considerations

There are a number of ethical (and potentially copyright-based) considerations to be made when dealing with collecting and analysing data from social media, including tweets. Traditionally, ethnographic research has preferred to make participants anonymous to protect their identity. However, Twitter's privacy policy states that:

> Our Services are primarily designed to help you share information with the world. Most of the information you provide us is information you are asking us to make public. This includes not only the messages you Tweet and the metadata provided with Tweets, such as when you Tweeted, but also the lists you create, the people you follow, the Tweets you mark as favorites or Retweet, and many other bits of information that result from your use of the Services. We may use this information to customize the content we show you, including ads. Our default is almost always to make the information you provide public for as long as you do not delete it from Twitter, but we generally

Please cite this article in press as: Hardaker, C., McGlashan, M., ''Real men don't hate women'': Twitter rape threats and group identity. Journal of Pragmatics (2015), http://dx.doi.org/10.1016/j.pragma.2015.11.005

Table 4
Top 20 most frequent lexical words.

Rank	Word	Freq	Rank	Word	Freq
1	Twitter	4616	**11**	know	2386
2	abuse	4465	**12**	support	1921
3	women	4309	**13**	woman	1661
4	people	3712	**14**	right	1655
5	threats	3435	**15**	hope	1614
6	think	3374	**16**	thanks	1586
7	rape	3248	**17**	trolls	1491
8	good	2785	**18**	sorry	1377
9	men	2536	**19**	time	1377
10	thank	2389	**20**	love	1360

Table 5
Most frequent topics/discursive strategies.

Topic/discursive strategy	Lexical items
(Sexual) aggression	abuse, rape, threats, trolls
Gender	men, women, woman
Mental processes	hope, know, love, think
Politeness markers	sorry, thank, thanks

> give you settings to make the information more private if you want. Our Services broadly and instantly disseminate your public information to a wide range of users, customers, and services. For instance, your public user profile information and public Tweets are immediately delivered via SMS and our APIs to our partners and other third parties, including search engines, developers, and publishers that integrate Twitter content into their services, and institutions such as universities and public health agencies that analyze the information for trends and insights. When you share information or content like photos, videos, and links via the Services, you should think carefully about what you are making public. (Twitter, 2015)

As such, Twitter users are informed of the instant and broad nature of the dissemination of any public tweets. Additionally, anonymising accounts would contravene Twitter's policies on displaying their data in static publications, which specify that in static and offline publications, tweets should show the name, username, and unmodified text.

4. The language of rape threats and identity construction

In the analysis, we implement methods from corpus linguistics to outline frequent topics of conversation occurring in the corpus. Whilst the findings from this analysis show that several topics and discursive/rhetorical strategies are highly frequent within the corpus, we focus primarily on talk relating to (sexually) aggressive behaviours.

We begin our analysis by examining frequent features in the language of the CPTMC through examining a frequency wordlist. A frequency wordlist shows the total number of times each unique lexical item occurs within a corpus (cf. McEnery and Hardie, 2012:243). In our initial results, the most frequent features, as with many corpora, were grammatical or 'function' words such as determiners and prepositions (Baker, 2006:53).

Since these can obscure—at least on a surface level—discourses that might be of interest, we excluded all word-classes but nouns, verbs, and adjectives, leaving results that could give us ''a better idea about of *discourses* within the corpus'' (Baker, 2006:54).[3] The results of our lexical wordlists were as follows:

The frequent lexical items in Table 4 reveal a number of broadly identifiable topics (or discursive strategies) within the corpus that can be summarised in Table 5.

Due to limitations of space, we focus on the topics of (sexual) aggression and gender, as well as their intersections.

[3] Although frequency wordlists reveal the *most common* lexical features of a corpus, they can obscure infrequent but discursively interesting features. However, we use frequency here as an initial way of delving into the corpus to enable deeper interpretations of more infrequent linguistic and discursive phenomena at later stages of the analysis.

Please cite this article in press as: Hardaker, C., McGlashan, M., ''Real men don't hate women'': Twitter rape threats and group identity. Journal of Pragmatics (2015), http://dx.doi.org/10.1016/j.pragma.2015.11.005

Table 6
Collocates of (sexual) aggression lexis.

Rank	ABUSE collocates				RAPE collocates				THREATS collocates				TROLLS collocates			
	Freq	Freq L	Freq R	Collocate	Freq	Freq L	Freq R	Collocate	Freq	Freq L	Freq R	Collocate	Freq	Freq L	Freq R	Collocate
1	745	470	275	twitter	1826	167	1659	threats	1826	1659	167	rape	154	99	55	twitter
2	388	188	200	threats	692	346	346	rape	404	153	251	twitter	73	66	7	feed
3	369	333	36	report	321	145	176	twitter	388	200	188	abuse	63	18	45	threats
4	362	17	345	button	257	53	204	threat	366	349	17	death	63	54	9	against
5	271	110	161	getting	243	112	131	abuse	233	59	174	against	60	34	26	abuse
6	243	131	112	rape	240	65	175	death	159	35	124	violence	54	8	46	rape
7	233	193	40	online	171	93	78	women	143	51	92	women	50	25	25	trolls
8	206	70	136	women	151	141	10	threatening	136	97	39	getting	46	22	24	people
9	189	27	162	received	147	141	6	threatened	132	80	52	people	45	18	27	women
10	164	146	18	vile	130	50	80	against	123	41	82	made	44	24	20	stop
11	160	76	84	people	117	86	31	people	117	61	56	received	42	33	9	internet
12	154	50	104	receiving	117	92	25	men	114	57	57	threats	41	37	4	ignore
13	148	101	47	against	100	36	64	jews	113	86	27	receiving	40	17	23	think
14	120	60	60	abuse	88	83	5	threaten	100	66	34	men	40	15	25	good
15	116	96	20	response	87	43	44	think	97	47	50	think	36	21	15	let
16	109	83	26	petition	86	23	63	violence	93	36	57	police	35	13	22	men
17	102	52	50	support	81	65	16	getting	93	36	57	making	34	18	16	support
18	102	54	48	men	76	63	13	receiving	92	63	29	online	33	29	4	feeding
19	101	74	27	stop	71	49	22	received	85	69	16	vile	32	27	5	taking
20	101	9	92	sent	69	42	27	woman	82	75	7	violent	32	4	28	need

4.1. (Sexual) aggression and gender

A collocation analysis of each of the frequent terms that make up the topic of (sexual) aggression—*abuse, rape, threats*, and *trolls*—was implemented to assess the meanings of these words as they occurred *in context* and how they shaped/were shaped by words with which they co-occurred. This was done by using the collocation function in AntConc and employs the Mutual Information (MI) statistical measure. Although other measures exist (log-likelihood, *z*-score), we draw on MI as it assesses both how closely words associate (by measuring frequency of co-occurrence) but also how strong those associations are (by measuring the likelihood that those two words occur together versus in isolation) (cf. Cantos Gómez, 2013:204–208).

The corpus was searched for each of the (sexually) aggressive terms, with a specification set to return only collocates occurring within a span of five words to either side of the search terms. The results of this are given in Table 6.

The collocates for each of these terms were then aggregated to observe collocates that were consistent—suggesting a stability in ways of talking about (sexual) aggression—across all terms found in the (sexual) aggression topic. At this point, we further distinguish between terms relating to (sexually) aggressive behaviours (*abuse, rape, threats*) and group nominations (*trolls*) and make some observations regarding the construction of (sexually) aggressive behaviours as well as aggressive groups. Table 7 shows collocates that occurred consistently frequently with (sexually) aggressive behaviours (*abuse, rape, threats*) and group nominations (*trolls*):

Table 7
Collocates consistently co-occurring consistently with aggressive behaviours and groups.

Rank	Collocate
1	Twitter
2	Threats
3	Rape
4	Women
5	People
6	Against
7	Abuse
8	Men

Please cite this article in press as: Hardaker, C., McGlashan, M., "Real men don't hate women": Twitter rape threats and group identity. Journal of Pragmatics (2015), http://dx.doi.org/10.1016/j.pragma.2015.11.005

Table 8
Collocates consistently occurring only with aggressive behaviours.

Rank	Collocate
1	Getting
2	Received
3	Receiving

Furthermore, some collocates occurred consistently and uniquely frequently with aggressive behaviours (*abuse, rape*, *threats*):

Further examination of Table 6 revealed that terms concerning gender, which were also some of the most frequent in the entire CPTMC corpus (*men*, *women*), are also frequent collocates of all terms of (sexual) aggression.

When *women* collocated with terms of (sexual) aggression, the surrounding discussion appeared to highlight issues concerning women as being the targets of a variety of forms of abuse or threats. When occurring alongside mentions of *women*, *abuse* and *threats* both predominantly occurred as nouns attributed to a particular class of abuse (*online abuse* and *sexist abuse*) or threat (*rape threats*). Moreover, throughout the CPTMC, *abuse* and *threats* occurred frequently alongside other nouns. Some classes were specific to *abuse* (e.g. *domestic abuse, child abuse, gendered abuse*) and *threats* (e.g. *bomb threats, death threats*) but some were shared (e.g. *criminal threats/abuse, cyber abuse/threats*). Adjectives expressing evaluation such as *awful, cowardly, disgraceful, despicable, graphic, hateful,* and *horrendous* were also prominent collocations, indicating the kinds discourse prosodies that may have been triggered when *abuse* and *threats* occurred as a collocate of *women*.

When talked about in relation to *threats* and *abuse*, *women* occurred as the grammatical target of abuse/threats, as indicated by the collocates shown in Table 8; *women received* or were *receiving* abuse/threats. *Getting* was used most frequently to talk about *getting women on banknotes*. Meanwhile, the grammatical actor (the one performing the abusive/threatening action) is typically absent or implied, therefore placing emphasis on the goal of those material processes (i.e. *rape threats* and *abuse*) and the recipient of those threats rather than the perpetrator. As such, there appears to be an intersection in the CPTMC corpus of the frequent topics of (sexual) aggression and gender with regards to women being consistently framed as the victims/targets of (sexual) aggression.

Whilst the construction of women seems largely clear cut—women are the receivers of abuse—the construction of men is more contested. As a collocate of *threats*, *men* were typically constructed as the makers and senders of rape threats, often with @CCriadoPerez named as the target:

Example 3.

User	Date/Time	Tweet
TyronWilson	2013-07-26 21:51:42	Can't believe there are men tweeting rape threats at @CCriadoPerez for working to get a woman on banknotes..some people need to get lives.
SimonTurkas	2013-07-30 12:40:04	How insecure some men must be to send threats to @CCriadoPerez simply because a female will appear on a banknote! #SHOUTINGBACK

The same was also true when *men* collocated with *abuse*:

Example 4.

User	Date/Time	Tweet
RealHumptyB	2013-08-05 07:35:58	Misguided men who abuse women on Twitter, pls read http://t.co/f6A0F9LQ0q on @CCriadoPerez, new force for women, democracy & modern England.
RFoXXy	2013-07-31 20:21:33	@CCriadoPerez Can't understand y so many men r sending abuse/threats 2 u. Who r these men? How can they have these attitudes towards women!?

Most interestingly, there were threads of contestation concerning the construction of a ''valid'' form of masculinity. Some users argued, for instance, that ''men don't get rape threats'' or that they are subjected to far less abuse than women online:

Please cite this article in press as: Hardaker, C., McGlashan, M., ''Real men don't hate women'': Twitter rape threats and group identity. Journal of Pragmatics (2015), http://dx.doi.org/10.1016/j.pragma.2015.11.005

Example 5.

User	Date/Time	Tweet
rugcernie	2013-07-27 21:25:47	I can answer that, @UltimationEE. Men don't get rape threats, ever! @CCriadoPerez
nonklatink	2013-08-10 11:19:14	@CCriadoPerez: ''maybe men don't get abuse just for being men with opinions.'' True. I upbraided many trolls before getting abusive replies.

Meanwhile, another community of users worked to explicitly construct a form of masculinity that they considered valid – specifically the identity of a ''real man''. These users deemed that being a *real man* was incompatible with abusive and threatening behaviour towards women. In other words, claims of *real men* as a legitimate form of gendered identity required the absence of gender-based (sexual) aggression:

Example 6.

User	Date/Time	Tweet
pasionflower	2013-07-27 10:33:23	@EverydaySexism @CCriadoPerez Real men don't rape.
Adamali03	2013-07-28 12:53:19	@CCriadoPerez #rape – real men protect and love the women in their lives.
theopenfire	2013-07-27 08:54:47	Support @CCriadoPerez, surely no place for this. Real men don't hate women. http://t.co/n7SuHcVKCc
ryangriffin89	2013-07-28 15:30:10	@CCriadoPerez real men are on your side!
pbagnall	2013-07-29 11:57:20	@CCriadoPerez abuse on twitter isn't free speech. It suppresses free speech. Real men welcome women's voices, cause we're not scared of them
will_seeman	2013-08-05 15:55:49	@CCriadoPerez 1) Appalled at threats you've had and wanted to say so. All real men should speak out against it.

This points to just two possible constructions of different gendered identities based on (sexually) aggressive behaviour. In short, given a context of increased focus on (sexual) aggression, throughout the events captured in this corpus, the positioning of men and women and the constructions of gender identities relative to (sexual) aggression was being contested, developed, and defined.

4.2. The language of rape threats: different discourse communities

One of the strongest collocations in the CPTMC, the n-gram *rape threats*, occurs 1419 times in total, accounting for 43.69% of all 3248 instances of *rape*. Although *rape* may semantically imply a form of behaviour, when talked about in the corpus, *rape* is frequently positioned as being primarily a form of threat. *Rape* also collocates very frequently with other *threat* lemma—''a group of wordforms that are related by being inflectional forms of the same base word'' (McEnery and Hardie, 2012:245)—including *threat, threats, threatening, threatened, threaten*. This suggests a stable discourse prosody in which the semantics of rape are conflated with that of threat.

Here, we are interested in whether communities form around particular discourses, and whether (newly) distinguishable communities share in the production of certain discourses. We focus on constructions of rape and how different discourse communities form and construct themselves through shared linguistic practices and discourse vis-à-vis their discursive constructions of rape. We study three broad groups of Twitter users identified in the CPTMC corpus: *high-risk, low-risk*, and *no-risk*.

High-risk users were defined as Twitter profiles that contained evidence of: intent to cause fear of (sexual) harm; harassment; and potentially illegal behaviour. **Low-risk** users were defined as Twitter profiles that contained evidence of: offensive material; insults; ridicule; no (linguistic) evidence of intent to cause fear or threat of (sexual) harm; and spamming (as opposed to harassment). **No-risk** users were defined as Twitter profiles that contained evidence none of the above.

A number of abusive users were pre-identified by Criado-Perez during the period covered within the data-sampling period. To track and identify more abusive users and their communicative networks, two methods of manual identification were employed. Users were identified through observing both directed connections (where a user mentions another in their tweet) and undirected or ''ambient'' connections whereby users might ''simply be speaking about the same topic at

Please cite this article in press as: Hardaker, C., McGlashan, M., ''Real men don't hate women'': Twitter rape threats and group identity. Journal of Pragmatics (2015), http://dx.doi.org/10.1016/j.pragma.2015.11.005

the same time'' (Zappavigna, 2014:211). Both methods involved manual interpretation of the content of tweets and classification of users.

For example, @Beccas43 (see Example 7) was identified as low-risk due to their use of *rape* to ridicule more prevalent discourses in the corpus. *Rape* in the tweets of @Beccas43 is decontextualized from the prevalent discourse mentioned and repeated so as to appear banal and inoffensive (which might, in itself, be offensive).

Example 7.

User	Date/Time	Tweet
Beccas43	26/07/13 17:17:48	.@CCriadoPerez RAPE RAPE RAPE RAPE RAPE RAPE RAPE RAPE RAPE RAPE JEWS JEWS JEWS RAPE RAPE RAPE RAPE RAPE RAPE RAPE RAPE RAPE RAPE JEWS JEWS
Beccas43	26/07/13 17:18:20	@CCriadoPerez RAPE RAPE RAPE RAPE RAPE RAPE RAPE RAPE RAPE RAPE JEWS JEWS JEWS RAPE RAPE RAPE RAPE RAPE RAPE RAPE RAPE RAPE RAPE JEWS JEWS
Beccas43	26/07/13 17:18:52	,@CCriadoPerez RAPE RAPE RAPE RAPE RAPE RAPE RAPE RAPE RAPE RAPE JEWS JEWS JEWS RAPE RAPE RAPE RAPE RAPE RAPE RAPE RAPE RAPE RAPE JEWS JEWS

Alongside *rape* is the word *jews*, and a further search revealed that *jews* was often used by others as a term of ridicule to cause offence.

Example 8.

User	Date/Time	Tweet
n1k__nak	26/07/13 17:23:56	@Beccas43 @CCriadoPerez HUEHUEHUEHUEHUHEUEHU I HEART RAPE, RAPING, RAPING JEWS RTC ETC LOLOLO
oqoco	28/07/13 13:35:50	@TrueCrimeUK @CCriadoPerez If Jesus couldn't stop the Jews from persecuting him, U def won't stop others who are bent on hating you're view
JonathanMayor	28/07/13 17:48:46	@darko_marco @CCriadoPerez If you say a la Hitler ''Jews = bad'' you cannot then use religious freedom as an argument.
FreiheitSecAnon	28/07/13 21:44:15	Hey.@CCriadoPerez followers! Go suck a giant nigger Jews cock. Might help you out a bit. Lulz
howardvaan	29/07/13 13:21:22	@quinnnorton @CCriadoPerez.. Or some of the US accounts talk about Jews, etc. (racist conspiracy nonsense).
FintanOToolbox	01/08/13 15:10:08	@Pimlids @CCriadoPerez @twitter Should black people fight for the right to join the KKK? Should Jews fight for the right to join the Nazis?
kingtytankhamen	11/08/13 23:04:44	''My greatest dream is a world without jews.'' - @CCriadoPerez

Example 8 shows that user @n1k__nak connects with user @Beccas43 (classified as low-risk) in both undirected and direct ways—they both talk about *jews* in a manner that could be seen as low-risk (the intention of its use appears to be ridicule) but @n1k__nak also directly connects by mentioning @Becca43. Others in Example 8, however, affiliate in an undirected way when talking about *jews*—they just happen to talk about the same topic in different ways—but neither directly or indirectly connects to known users who might be low- or high-risk.

Through repeating this process—following numerous directed and undirected connections—a total of 208 'risky' users were detected (147 low-risk, sixty-one high-risk). Three separate subcorpora were created from the tweets of each user group, named CPTMC no-risk, CPTMC low-risk, and CPTMC high-risk. A keyword analysis was then conduced whereby both the CPTMC low-risk and CPTMC high-risk corpora were compared against the CPTMC no-risk corpus in order to assess differences in discourse between the user groups and to assess whether different discourse communities exist (Table 9).

Several frequent keywords were shared by low- and high-risk users in the CPTMC suggesting an interface between language and discourse with regards to sexual violence (*rape, raep*) and misogynistic insults (*bitch, cunt*) that may be characteristic of risky users engaged in making or talking about rape threats. However, whilst this mutual interest in similar lexis may indicate that they are part of a wider discourse community, differences between the groups also exist.

Please cite this article in press as: Hardaker, C., McGlashan, M., ''Real men don't hate women'': Twitter rape threats and group identity. Journal of Pragmatics (2015), http://dx.doi.org/10.1016/j.pragma.2015.11.005

Table 9
Top 10 keywords in the CPTMC low-risk and high-risk sub-corpora.

Rank	Low-risk keywords			High-risk keywords		
	Freq	Keyness	Keyword	Freq	Keyness	Keyword
1	41	195.117	bitch	45	147.763	lol
2	35	164.722	cunt	24	123.965	bitch
3	152	149.871	rape	23	121.966	cunt
4	17	122.816	jews	12	101.802	raep
5	13	83.777	pussy	9	87.576	loool
6	48	83.618	fuck	13	82.556	raping
7	10	67.572	cuz	11	79.651	nigger
8	18	62.827	penis	105	77.087	your
9	40	62.543	internet	10	74.380	faggot
10	9	62.377	raep	74	68.023	Rape

A significant feature in the talk of the low-risk discourse community was *internet*, including attempts to define it as a discreet social space with its own particular rules, regulations, and realities which might be challenging and unpleasant, but not illegal. Example 9 shows @kingtytan parodying the grievances of @CCriadoPerez—that rape threats sent using the internet are wrong—alongside an exaggerated and caricatured version of radical feminist rhetoric.

Example 9.

User	**Date/Time**	**Tweet**
Kingtytan	2013-07-26 17:17:19	.@CCriadoPerez SAYING MEAN THINGS ON THE INTERNET IS ILLEGAL #KILLALLMEN #DIECISSCUM

The positioning of @kingtytan is therefore intentionally deceptive and meant to discredit not only @CCriadoPerez's claims and arguments concerning online abuse, but also her identification as a feminist. Meanwhile, key in the tweets of high-risk users is the verb *raping*:

Example 10.

User	**Date/Time**	**Tweet**
Lord0Lulz	2013-07-27 17:52:02	@kingtytan @SultanOfPing @CCriadoPerez Some women just need a good raping every now and again I guess:-/

In short, the high-risk users appeared to breach numerous UK laws regarding threat, harassment, and obscenity, however the low-risk users' employment of sarcasm, insult, and mockery should not be automatically discounted as causing no damage. However, whilst an especially interesting, the current wordcount does not permit full investigation of this particular aspect.

5. Conclusions

We started out this paper with two particular aims. The first was to investigate the language surrounding sexual aggression on Twitter, and within our corpus, the discourse of abuse focussed particularly on rape. Within this discourse, we found that the discussion focussed on this behaviour as a threat, and arguably as a misogynistic weapon utilised to control the discourse of women online. Women were predominantly the target of these threats (both literally and grammatically) whilst the discourses surrounding men and rape involved the construction of ''real'' masculinity as one that categorically excludes the use of threatening or violent behaviour towards women.

This moves us into the second issue, namely the emergence and construction of discourse communities in response to that sexually aggressive language. However, before moving into possible answers, perhaps the most crucial issue here was how cleanly and neatly different ''communities'' or ''groups'' can be identified, especially when dealing with a highly fluid, fast-moving environment like Twitter populated by users who may coalesce around a topic or user and engage in transient interactions for a mere matter of seconds before moving on. Indeed, terms like ''community'' or ''group'' seem far

Please cite this article in press as: Hardaker, C., McGlashan, M., ''Real men don't hate women'': Twitter rape threats and group identity. Journal of Pragmatics (2015), http://dx.doi.org/10.1016/j.pragma.2015.11.005

too strong for a collection of people who may have no further connection to each other than to have tweeted the same target with either support or abuse. The very notion, here, of a ''community'' or ''group'' is therefore problematic even before we move into issues such as determining where boundaries between groups lie.

Notwithstanding this particularly troublesome issue, a larger, nebulous group emerged from the analysis, and within this, it was possible to identify a smaller network of low-risk users (those who tweeted insults and sarcasm), and a smaller-still network of low- and high-risk users (those who tweeted threats, harassment, and even breached any number of UK laws). It would be easy to automatically discount the low-risk users from their place in the larger network, however, it is worth considering that similarities between the discourses shared by these groups could facilitate a user's gradual escalation from low-risk (unpleasant) through to high-risk (illegal) online interaction, possibly without even being quite aware of that gradual shift. Indeed, both the low- and high-risk abusers coalesced not only around the discussion of rape, but also of misogyny, racism, and homophobia.

Whilst anonymity enables individuals to freely exchange ideas and opinions that, expressed otherwise, could irrevocably damage their reputation or cause them personal harm (Vamialis, 2013:32), it can also be used as a shield from behind which to offend, attack, defame, and harass others, whilst protecting the assailant from easy identification and subsequent social or legal reprisals. At the same time, social networks have proliferated, diversified, and evolved at a pace which has drastically outstripped the laws developed to govern them, leaving targets of online attacks in the difficult position of breaking new ground when attempting to prevent and prosecute criminally offensive online behaviour. Similarly, the lack of research into this domain means that empirical, evidence-based updates to that legislation are extremely difficult, and it is in light of this shortage that this paper seeks to make its contribution.

Funding

Q9 This work was supported by the Economic and Social Research Council [grant number ES/L008874/1].

References

Baker, P., 2001. Moral panic and alternative identity construction in Usenet. J. Comput.-Mediat. Commun. 7 (1) http://jcmc.indiana.edu/vol7/issue1/baker.html (08/12/2009)
Baker, P., 2006. Using Corpora in Discourse Analysis. Continuum, London.
Baker, P., 2010. Representations of Islam in British broadsheet and tabloid newspapers 1999–2005. J. Lang. Polit. 9 (2), 310–338.
Baker, P., 2014. Using Corpora to Analyze Gender. Bloomsbury, London.
Baker, P., Gabrielatos, C., Khosravinik, M., Krzyzanowski, M., McEnery, T., Wodak, R., 2008. A Useful Methodological Synergy? Combining Critical Discourse Analysis and Corpus Linguistics to Examine Discourses of Refugees and Asylum Seekers in the UK Press. Sage Publications.
Baker, P., Gabrielatos, C., McEnery, T., 2013. Discourse Analysis and Media Attitudes. Cambridge University Press, Cambridge.
Biber, D., Conrad, S., Cortes, V., 2004. If you look at . . .: Lexical Bundles in University Teaching and Textbooks. Appl. Linguist. 25 (3), 371–405.
Boxer, D., Cortés-Conde, F., 1997. From bonding to biting: conversational joking and identity display. J. Pragmat. 27, 275–294.
Bucholtz, M., Hall, K., 2005. Identity and interaction: a sociocultural linguistic approach. Discourse Stud. 7 (4–5), 585–614.
Cameron, D., 1997. Performing gender identity: young men's talk and the construction of heterosexual masculinity. In: Johnson, S., Meinhof, U.H. (Eds.), Language and Masculinity. Blackwell, Oxford, pp. 86–107.
Cantos Gómez, P., 2013. Statistical Methods in Language and Linguistic Research. Equinox, Sheffield.
Criado-Perez, C., 2013. We Need Women on British Banknotes. Available from: http://www.change.org/en-GB/petitions/we-need-women-on-british-banknotes (accessed 10.08.13)
Diener, E., 1979. Deindividuation, self-awareness, and disinhibition. J. Pers. Soc. Psychol. 37 (7), 1160–1171.
Douglas, K.M., McGarty, C., 2001. Identifiability and self-presentation: computer-mediated communication and intergroup interaction. Br. J. Soc. Psychol. 40 (3), 399–416.
Eckert, P., Rickford, J.R., 2001. Style and Sociolinguistic Variation. Cambridge University Press, Cambridge.
Edwards, D., 1998. The relevant thing about her: social identity categories in use. In: Antaki, C., Widdicombe, S. (Eds.), Identities in Talk. Sage, London.
Festinger, L., Pepitone, A., Newcomb, T., 1952. Some consequences of de-individuation in a group. J. Abnorm. Soc. Psychol. 47 (2, Suppl.), 382–389.
Giles, H., Coupland, J., Coupland, N. (Eds.), 1991. Contexts of Accommodation: Developments in Applied Sociolinguistics. Cambridge University Press, Cambridge.
Q10 Hardaker, C. (forthcoming). *The Antisocial Network*. London: Palgrave.
Herring, S.C., Stoerger, S., 2014. Gender and (a)nonymity in computer-mediated communication. In: Ehrlich, S., Meyerhoff, M., Holmes, J. (Eds.), The Handbook of Language, Gender, and Sexuality. Oxford, pp. 567–586.
Herring, S.C., Stein, D., Virtanen, T., 2013. Introduction to the pragmatics of computer-mediated communication. In: Herring, S.C., Stein, D., Virtanen, T. (Eds.), Pragmatics of Computer-Mediated Communication. Mouton De Gruyter, Berlin, pp. 3–32.
Holmes, J., 1997. Women, language and identity. J. Sociolinguist. 1 (2), 195–223.
Le Page, R.B., Tabouret-Keller, A., 1985. Acts of Identity: Creole-based Approaches to Language and Ethnicity. Cambridge University Press, Cambridge.
Mahlberg, M., 2013. Corpus analysis of literary texts. In: Chapelle, C.A. (Ed.), The Encyclopedia of Applied Linguistics. Oxford.

Please cite this article in press as: Hardaker, C., McGlashan, M., ''Real men don't hate women'': Twitter rape threats and group identity. Journal of Pragmatics (2015), http://dx.doi.org/10.1016/j.pragma.2015.11.005

McEnery, T., Hardie, A., 2012. Corpus Linguistics. Cambridge University Press, Cambridge.
Mendoza-Denton, N., 2002. Language and identity. In: Chambers, J.K., Trudgill, P., Schilling-Estes, N. (Eds.), The Handbook of Language Variation and Change. Blackwell, Oxford, pp. 475–499.
Meyerhoff, M., 1996. Dealing with gender identity as a sociolinguistic variable. In: Bergvall, V.L., Bing, J.M., Freed, A.F. (Eds.), Rethinking Language and Gender Research: Theory and Practice. Longman, London, pp. 202–227.
Mullany, L.J., 2007. ''Stop hassling me!'': impoliteness, power and gender identity in the professional workplace In: Bousfield, D., Locher, M.A. (Eds.), Impoliteness in Language. Mouton de Gruyter, Berlin.
O'Brien, J., 1999. Writing in the body. Gender (re)production in online interaction. In: Smith, M.A., Kollock, P. (Eds.), Communities in Cyberspace. Routledge, London, pp. 76–104.
Ochs, E., 1992. Indexing gender. In: Duranti, A., Goodwin, C. (Eds.), Rethinking Context: Language as an Interactive Phenomenon. Cambridge University Press, Cambridge, pp. 335–358.
Preece, J., 2000. Online Communities: Designing Usability, Supporting Sociability. John Wiley, Chichester, UK.
Reid, E. (1994). Cultural formations in text-based virtual realities. (Unpublished MA thesis). Melbourne, Australia: University of Melbourne, http://www.aluluei.com/cult-form.htm.
Rheingold, H., 1993. The Virtual Community: Homesteading on the Electronic Frontier. Addison-Wesley, Reading, MA.
Siegel, J., Dubrovsky, V.J., Kiesler, S., McGuire, T.W., 1986. Group processes in computer-mediated communication. Organ. Behav. Hum. Decis. Process. 37 (2), 157–187.
Silverstein, M., 1976. Shifters, linguistic categories, and cultural description. In: Basso, K.H., Selby, H.A. (Eds.), Meaning in Anthropology. University of New Mexico Press, Albuquerque, pp. 11–55.
Silverstein, M., 1979. Language structure and linguistic ideology. In: Clyne, P.R., Hanks, W.F., Hofbauer, C.L. (Eds.), The Elements: A Parasession on Linguistic Units and Levels. Chicago Linguistic Society, Chicago, pp. 193–247.
Silverstein, M., 1985. Language and the culture of gender: at the intersection of structure, usage, and ideology. In: Mertz, E., Parmentier, R.J. (Eds.), Semiotic Mediation: Sociocultural and Psychological Perspectives. Academic Press, Orlando, FL, pp. 219–259.
Simon, B., 2004. Identity in Modern Society. A Social Psychological Perspective. Blackwell, Oxford.
Spears, R., Lea, M., 1992. Social influence and the influence of the 'social' in computer-mediated communication. In: Lea, M. (Ed.), Contexts of Computer-Mediated Communication. Harvester Wheatsheaf, New York, pp. 30–65.
Tajfel, H., Turner, J., 1986. The social identity theory of intergroup behavior. In: Worchel, S., Austin, W.G. (Eds.), Psychology of Intergroup Relations. Nelson-Hall, Chicago, pp. 7–24.
Terkourafi, M., 2005. Identity and semantic change: aspects of T/V usage in Greek Cypriot. J. Hist. Pragmat. 6 (2), 283–306.
Twitter, 2015. 'Twitter Privacy Policy.' Twitter no Version 9 (18 May 2015). https://twitter.com/privacy
Verschueren, J., 2004. Identity as denial of diversity. In: Brisard, F., Meeuwis, M., Vandenabeele, B. (Eds.), Seduction, Community, Speech. Benjamins, Amsterdam, pp. 171–181.
Vinagre, M., 2008. Politeness strategies in collaborative e-mail exchanges. Comput. Educ. 50 (3), 1022–1036.
Wallace, K.A., 1999. Anonymity. Ethics Inf. Technol. 1, 23–35.
Zappavigna, M., 2014. Enacting identity in microblogging through ambient affiliation. Discourse Commun. 8 (2), 209–228.
Zarsky, T. Z. (2004). Thinking Outside the Box: Considering Transparency, Anonymity, and Pseudonymity as Overall Solutions to the Problems of Information Privacy in the Internet Society. (Unpublished PhD Thesis). Columbia Law School.

Claire Hardaker is a Lecturer in Corpus Linguistics at Lancaster University, and Principal Investigator on the ESRC funded project, ''Twitter rape threats and the discourse of online misogyny'' (ES/L008874/1). Her research focuses on deception, aggression, and manipulation, particularly in an online context. In particular, she has focussed on trolling, and she is currently writing a monograph entitled *The Antisocial Network* for Palgrave.

Mark McGlashan is a Senior Research Associate at the ESRC Centre for Corpus Approaches to Social Science (CASS) based in Lancaster University's Department of Linguistics and English Language where he is also a PhD candidate in Applied Linguistics. As a Senior Research Associate, his research focuses on discourses of online misogyny and the development of methods for detecting and analysing cases of online rape threats largely in the context of Twitter. His PhD research focuses on multimodal representations of same-sex parent families in children's picturebooks. His main research interests are gender and sexuality, critical discourse analysis, multimodality, corpus linguistics, and social network analysis.

Please cite this article in press as: Hardaker, C., McGlashan, M., ''Real men don't hate women'': Twitter rape threats and group identity. Journal of Pragmatics (2015), http://dx.doi.org/10.1016/j.pragma.2015.11.005