

Trigger warning: the development and popularization of new forms of censorship in social media
and microblogging platforms

Sara Nummenpää
Pro Gradu Thesis
English Philology
Faculty of Humanities
University of Oulu
Spring 2017

Table of Contents

1	Introduction	1
2	Theoretical background	3
2.1	Computer-mediated Communication	3
2.2	CMC on the Internet	5
2.2.1	Modes of CMC	7
2.2.2	The World Wide Web	8
2.2.3	Web 2.0	9
2.2.4	Technological determination	12
2.2.5	Freedom of expression	13
2.2.6	Community, social effects, and large scale communication	14
2.3	Censorship	15
2.4	Gatekeeping and content warnings in mainstream media	17
2.5	Censorship and psychology	18
2.6	Self-censorship and social media	19
2.6.1	Public and private self-censorship	20
2.6.2	Microblogging as a mediated social practice	21
3	Data	23
3.1	Tumblr as a microblogging platform	23
3.2	Metatags	25
3.3	Browser extensions	25

4	Methodology	26
4.1	Frequency and content analysis	26
4.2	Corpus analysis	28
5	Analysis	29
5.1	Development of metatags in censorship practices	29
5.2	Comparative frequency analysis	32
5.2.1.	#Emetophobia	33
5.2.2.	#Agoraphobia	35
5.2.3.	#Arachnophobia	36
5.3	Trigger warnings in use	39
5.3.1.	#Emetophobia tw	39
5.3.2.	#Agoraphobia tw	42
5.3.3.	#Arachnophobia tw	43
5.4	The arachnophobia tag variants	46
5.4.1.	#Arachnophobia cw, #cw arachnophobia, #arachnophobia tw, and #tw arachnophobia	46
5.4.2.	Word frequency analysis	50
5.4.3.	Clusters and concordances	57
5.4.4.	Concordance plots	61
6	Conclusion	64
	List of References	66

Appendix A. 100 most common words in the marked arachnophobia tags.

Appendix B. Normalized frequency of 100 most common words in the marked arachnophobia tags.

Appendix C. Normalized frequency of 100 most common words in the Corpus of Contemporary American English.

1 Introduction

Within the past few years there has been a noticeable shift in the nature of content production and distribution on social media websites. With the spread of mental health awareness, as well as the development of browser plug-ins which enable the blacklisting of keywords, the use of content warnings on personal blog posts has become increasingly frequent. Moreover, increased social awareness among the general public has brought to light a number of questions concerning the numerous facets of political correctness, freedom of expression, and personal image management that are involved in active social networking. These elements have converged in a widespread practice of innovative content classification and censorship in certain social spheres, which is produced on a self-reflective basis. In other words, this particular brand of censorship is self-imposed, and the social culture of these online social spheres has developed to accommodate and perpetuate its practice. The intent of this paper is to examine the development and popularization of these new forms of censorship as they are used on microblogging platforms.

Particular attention will be paid to the usage of trigger warnings on Tumblr.com due in part to the format, which enables the classification and censorship of content through a system of metatags, as well as to the site culture, which has helped popularize both censorship and self-censorship in social media. The nature of this censorship will be described as it appears within the framework of the website, and the relationship between the agents of censorship and the content beings censored will be scrutinized. The concept of censorship in these spheres will be defined and grounded in a contemporary context, and the nature of this practice will be analyzed in the dynamic environment in which it presents itself. The cross-sections of linguistics and psychology which lie at the heart of these developments will be considered, with a particular focus on the executive functions of language taking on an active role in cooperative censorship in social media. This study will focus on describing the form which these metatags take and their relationship to the text with which they are associated, as well as the popularization of their usage and how it has evolved.

As these developments are rather recent, there has not been much research done concerning these specific phenomena. Naturally, there is much to be considered in order to form a comprehensive understanding of the nature and usage of these new forms of censorship on the Internet. However, since the phenomenon is new, this study will focus on a diachronic descriptive approach in order to define the practice and its development on Tumblr. The phenomenon as it appears must first be approached from a descriptive standpoint, with attention paid to the specifics of practice as illustrated by a particular set of data. Once these phenomena are better understood, only then can the general nature of their usage be determined. Furthermore, due to the personal, wide-spread, and collaborative nature of the internet, the popularization of new forms of censorship in social media may reflect other social developments in these demographics that are most active on microblogging platforms. By considering these developments and the role of the Internet in their conception and popularization, they can be placed within an appropriate context for further research. The study of these practices offers an insight into a rising dynamic social culture and how this culture is perpetuated by modern technologies.

The following section presents an involved synthesis of the history of censorship in media, the tension between censorship practices and concepts of freedom of expression, content warnings in mainstream media, and the psychological significance of censorship. Section 3 gives an overview of the data, including an explanation of how Tumblr posts function and the mechanics involved in censorship. Section 4 then explains the format and methodology of the study, as well as the processes of data collection and analysis. Section 5 includes an in-depth analysis of the multiple different forms these censorship practices take on Tumblr.com, examined first through a broad survey of frequency and content, then followed by a closer look at the lexical content of one group of metatags in particular. Finally, section 6 presents the conclusions of this study and a discussion of its implications, including a suggestion of how to proceed with further research.

2 Theoretical background

This section will begin with a closer look at computer-mediated communication and how it functions on the Internet. This will be followed by an overview of the practice of censorship in the context of both private and public communication and social interaction. The role of gatekeeping as well as the usage of content warnings in mainstream media will then be further examined, so as to establish a point of comparison when making the shift to cooperative censorship in more personal forms of communication. It must be noted that since the forms of censorship that this study will examine are a relatively recent phenomenon, there is very little written on the topic itself. This study must therefore rely on research relevant to more traditional forms censorship in order to establish a baseline from which to begin analysis.

The psychological implications of censorship will also be examined in order to present a more comprehensive understanding of the more practical functions of censorship in a social context. However, this will remain somewhat limited, so as to avoid the risk of unwarranted extrapolation in a field not directly related to this study. Finally, this section will look at what forms of self-censorship are present in various social media networks and the roles this censorship plays in the dynamic social exchanges taking place on microblogging platforms.

2.1 Computer-mediated communication

In order to better understand the phenomena at hand, its function must first be considered in the context of its form and mode. As it takes place solely through the Internet, naturally it falls under the classification of computer-mediated communication. Computer-mediated communication, hereafter referred to as CMC, is, by its most basic definition, communication which takes place between humans via the instrumentality of computers (Herring, 1996, p. 1). However, its exact parameters still remain somewhat enigmatic; a number of researchers have attempted to more

accurately describe CMC, and have produced a number of similar definitions of varying degrees of precision. December (1997) has described it as the following:

Computer Mediated Communication is a process of human communication via computers, involving people, situated in particular contexts, engaging in processes to shape media for a variety of purposes.

December presents a carefully considered, but vague definition, expanding upon Herring's interpretation to situate CMC within the context of its roles and its function in society with a focus on the actuation of its purpose. In other words, CMC involves human communication through a variety of cooperative processes to engage with different forms of media in order to create meaning. Conversely, Santoro (1995; as cited in Thurlow, 2004) has attempted to more clearly describe the different modes of CMC within the context of human communication:

At its broadest, CMC can encompass virtually all computer uses including such diverse applications as statistical analysis programs, remote-sensing systems, and financial modelling programs, all fit within the concept of human communication (p. 15).

While December focuses on the *what*, Santoro's more concrete applications engage with the *how*. Though these different formats do not immediately seem to involve more typical forms of human communication, they speak to the enigmatic nature of what communication is and how it is performed. However, for the sake of this study focus will mainly be placed on text-based CMC on the Internet. This can take two forms: synchronous CMC, where the message is received immediately, and asynchronous CMC, where it is received at a later point in time. This study will be looking at the latter.

To further understand the nature of CMC, its history and development must be addressed. Essentially, CMC began when the first electronic digital computer was built, or at the very least in the early 1960s, when the first prototype emails were exchanged. However, the field has experienced an exponential expansion only within the last twenty years, as computers made the shift from being highly technical, expensive machines to highly accessible and personal communicative tools (Thurlow et al., 2004, p. 14). Nowadays, text-based CMC is engaged in daily

by millions of people worldwide. This phenomenal growth of CMC has encouraged rampant speculation of its implications within the global socio-cultural environment, coinciding in particular with the advent of the Internet. The particular role of the Internet in the context of CMC will be further discussed in Section 2.2.

According to Herring (1996), the discussion which has arisen due to the spread of CMC has mainly been concerned with:

[T]he democratic and anarchic possibilities inherent in widespread use of a networked medium which allows anyone with access to speak out more or less anonymously, and which is not as yet subject to any centralized authority or control" (p. 1).

It is this question of the possibilities and repercussions, which are tied to the very nature of CMC, that has dominated the discussion of CMC on the Internet, but the availability of factual information and research on the matter has remained slim; at the turn of the century, little language-focused research on new phenomena had been published, and Herring insisted that "there remains a need for descriptive and empirical research on computer-mediated interaction" (ibid.). Indeed, the challenge for emergent phenomena and media environments has always been their need for descriptive analysis before more sophisticated analyses can be executed (Virtanen et al., 2013, p. 15). In order to conduct such descriptive research within this dynamic and multifaceted environment, the role of the Internet in CMC must first be discussed.

2.2 CMC on the Internet

In the words of Herring (2002), "the Internet brought millions of people online, and what they did mostly was communicate" (p. 109). Indeed, the opportunities provided by increasingly widespread access to computers combined with advancements in Internet connectivity has resulted in a veritable explosion of CMC taking place through the Internet. This communication takes place

through a variety of different modes, including email, newsgroups, forums, and chat; and in a variety of different social contexts, including but not limited to professional, political, recreational, and commercial (Herring, 2002, pp. 109–110). This multifaceted nature of CMC on the Internet has only diversified as new developments in computer technology have taken place. The history of these developments will be further discussed in Sections 2.2.2 and 2.2.3, but the very nature of Internet communication and information spread has ensured that these changes have resulted in a lasting impact on the foundations of what is considered to be modern day CMC research.

Indeed, in the *Annual Review of Information Science and Technology*, Herring (2002) states that “the Internet increasingly defines CMC by providing the context within which many, if not most, CMC applications operate” (p. 110). Nowadays, the Internet has become inextricable from the majority of the CMC taking place, as its role has become central to the large scale functioning and usage of computers as communicational tools. The special affordances that this relationship allows have led to a unique dynamic between computer and communicator, one which remains fluid and ever evolving as computer science develops new ways of interacting through the Internet.

The following sections will look at some of the features of CMC on the Internet, its implications and effects, as well as some of the supporting theories which help define the field of study. Section 2.2.1 will discuss the different modes of CMC, followed by Section 2.2.2 and 2.2.3 which discuss the history and development of CMC on the Internet in two different phases: the World Wide Web and Web 2.0. Section 2.2.4 introduces the question of the role of technological determination in the development and practice of CMC, and Section 2.2.5 considers the issue of freedom of expression in the context of CMC. This is followed by an overview of its influence on community formation and large scale discussions in Section 2.2.6.

2.2.1 Modes of CMC

According to Herring (2002), “perhaps the most important cumulative finding of Internet research over the past fifteen years is that computer-mediated communication varies according to the technologies on which it is based, and according to its contexts of use” (p. 111). The factors which demonstrate this variance range from temporal constraints (i.e. synchronous vs asynchronous communication), to communicative channels, and other features of the messaging platform through which the communication is taking place. However, these characteristics are not the sole determining factor in the classification of CMC. It must be taken into account that differences in user demographics (e.g., gender, age, race, education, etc.) can result in vastly different communication styles, even within the same system (Burkhalter, 1999, p. 63). Additionally, situational variables such as the purpose of communication, participant structures, and the topic of communication greatly influence online communication.

Ultimately, the combination of a variety of different factors dictates the resultant communication style, and in conjunction with the characteristics of the computer environment, the convergence of these elements allows the CMC to be situated within a recognizable mode. A mode is essentially a genre of CMC which combines messaging protocols with the social and cultural practices that have evolved around their use (Murray, 1988). In other words, the mode is determined by both the features and constraints of a messaging platform in coordination with the ways in which they are utilized. Due to the very nature of online communication, the emergence of the different modes of CMC is inherently tied to history of the Internet. From the ARPANET, the predecessor of the modern day Internet, to the Usenet and the World Wide Web, CMC has developed alongside the technology, utilizing its different features and capabilities to expand the reaches of human communication. A brief overview of this shared history can be seen in Figure 2.1.

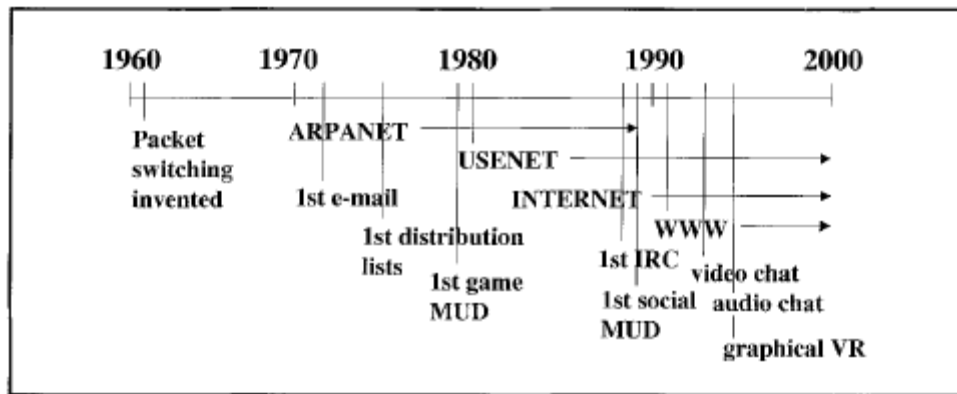


Figure 2.1. The co-evolution of the Internet and CMC (Herring, 2002, p. 114).

As shown above, the most influential developments of CMC can be overlaid on top of a concise timeline of the history of the Internet with striking coincidence. Beginning with ARPANET and the 1st email in the 1970s, to the introduction of the World Wide Web and video and audio chat in the 1990s, Internet technology and CMC have developed symbiotically, with the form and function of the former informing the evolution of the latter.

2.2.2 The World Wide Web

The advent of the World Wide Web brought with it a number of developments in the format and usage of CMC. First conceptualized in 1989 by Tim Berners-Lee, a scientist of the CERN (European Organization for Nuclear Research), the World Wide Web was implemented on the Internet in 1991 (Herring, 2002, p. 125). The goal of this venture was to create a cooperative shared information space “through which large numbers of people and machines could communicate” (Berners-Lee, 1996; as cited in Herring, 2002, pp. 125–126). Indeed, in 1993 when the Mosaic graphical Web browser was introduced the Web finally became widely accessible by the public, and consequently, businesses. This subsequent commercial involvement spurred further development of the Web as an interactional space and marketing medium (Goggin, 2000). By the year 2000, there were over five million Web sites in existence and Web browsing had become the

main application of the Internet, second only to email in popularity (Herring, 2002, p. 126; Pastore, 2000).

In essence the Web is a meta-mode of CMC: it subsumes and integrates all other modes of CMC (Soukup, 2000; as cited in Herring, 2002, p. 126). The Web serves as a link to chat interfaces, discussion lists, and email applications in addition to other Web pages, as well as to different communication media like text, graphics, audio, and video (ibid.). In this sense, the characteristics associated with Web communication are broad and multifaceted in nature, which poses the question: to what extent is this communication interpersonal? With the intrusion of capitalism the Web had become a place to make commercial transactions, and Web pages used for other purposes were also less conducive to interpersonal interaction than other modes, like email. Indeed, as Wakeford (2000) has pointed out, Web-based communication is less interactive than other modes such as chat; Web pages often communicate meaning through various nontextual means, such as graphics, page layout, and web design, to which a user cannot respond in kind. Ultimately, it was these limitations that formed the foundation for the development and expansion of the Web as a communicative network, as exemplified by Web 2.0.

2.2.3 Web 2.0

The late 1990s saw a stark shift towards more dynamic websites, geared specifically towards frequent communication. These websites were designed so that their content could be updated frequently and allowed users to interact with the content by more immediate means, through chat boxes, comment sections, and the like (Virtanen et al., 2013, p. 12). These trends laid the foundations for what came to be called Web 2.0. Though highly debated, the term Web 2.0 is typically understood to refer to a set of popular web-based platforms “characterized by social interaction and user-generated content” (ibid.). As Web 2.0 set out to solve the limitations of the World Wide Web, Web sites adapted to the rising pressure and desire for easier communication and became far more conducive of interpersonal interaction. This gave way to the rise of blogging

sites, social media networks, and editable catalogues of information (i.e. Wikipedia and its progeny).

Web 1.0	Web 2.0
Personal websites	Blogging
Publishing	Participation
Britannica online	Wikipedia
Content management systems	Wikis
Stickiness	Syndication
Directories (taxonomies)	Tagging (folksonomies)

Figure 2.2. Web 1.0 vs. Web 2.0 phenomena (adapted by Virtanen et al., 2013, p. 13; from O’Reilly, 2005).

Figure 2.2 above highlights the evolution of different Web phenomena from the shift from Web 1.0, or the World Wide Web, to Web 2.0. Personal websites were overtaken by personal blogs, Britannica online was overshadowed by the cooperative encyclopedia site Wikipedia; overall there was a general shift from content existing as singular, independent points to cooperatively produced, shareable content (i.e. Stickiness vs. Syndication). This shift in the very nature of the Internet warranted a new classification: Web 2.0.

The term itself was first introduced in 2004, when it was far more aspirational than descriptive, and its meaning has fluctuated with the developments in Web usage and design (Virtanen et al., 2013, p. 13) Accordingly, today Web 2.0 refers to “the changing trends in, and new uses of, web technology and web design, especially involving participatory information sharing; user-generated content; an ethic of collaboration; and use of the web as a social platform” (Virtanen et al., 2013, p. 12). This concept has come to encompass a large portion of modern Web sites and applications, and as they have become accessible from not only desktop computers and laptops, but also smart phones and tablets, the presence and relevance of Web 2.0 in daily life has increased accordingly. The use of the Web as a social platform is particularly visible as social media has become a large staple of popular culture, as people engage in CMC through social media sites such as Facebook,

Twitter, and Tumblr. This study is concerned primarily with these occurrences of CMC, and the developments which have arisen from such usage.

On the other hand, as with the first iteration of the World Wide Web, the websites of Web 2.0 come with their own set of limitations. This is reflected in how users engage with the site and the developments of these adaptations can be catalogued through the discourse taking place. Virtanen et al. (2013) has addressed these phenomena as follows:

Web 2.0 discourse includes user adaptations to circumvent the constraints of Web 2.0 environments: for example, interactive uses of @ and #, as well as retweeting, on Twitter [...] and performed interactivity on what are, in essence, monologic blogs (p. 14).

This study will focus on these phenomena, specifically on those related to the different interactive usages of the hashtag (#) on Tumblr, which have developed as a relatively new and site-specific convention. That being said, Web 2.0 discourse phenomena can be classified into three groups: phenomena familiar from older CMC modes, phenomena which adapt to Web 2.0 environments, and new or emergent phenomena that did not exist prior to the introduction of Web 2.0 (ibid.). Familiar discourse phenomena include modes such as email, chat, and forums, whereas adapted phenomena include status updates (Lee, 2011), quoting or retweeting, and blogging (Myers, 2010). On the other hand, emergent discourse phenomena are ones which either did not exist or did not rise to any level of public awareness before the development of Web 2.0. In most cases, many of these phenomena cannot be definitively labeled as entirely new, as they can possibly have online or offline antecedents (Virtanen et al., 2013, p. 14). These new phenomena are somewhat difficult to discuss, as they are yet relatively unstudied and unresearched, as a descriptive analysis generally must predate any deeper analyses. Nevertheless, this study is primarily concerned with emergent phenomena.

2.2.4 Technological determination

The nature of these discourse phenomena in the context of CMC on the Internet is multifaceted, and their origin can be problematic to trace. The factors which influenced their formation can be difficult to define; with this in mind, Herring (1996) has posed the following question:

To what extent does the computer medium alter human interaction, and to what extent do people simply map their existing patterns of behavior onto communication in the new medium? (p. 4).

Indeed, it is difficult to determine in which direction influence flows when human and computer interact so intimately, and it is with this same question that the issue of technological determination is concerned. According to Virtanen et al. (2013), “new expressive needs and forms arise from and adapt to specific conditions of the new medium” (p. 9). However, the question still remains: to what extent do these expressive forms arise from the conditions of the medium, and to what extent do existing conditions adapt to them?

Making the distinction between these two classifications is incredibly difficult due to the nature of CMC; they appear and spread without leaving specific clues as to where they originated and how. Regarding this question, Virtanen et al. (2013) stated the following:

Such new functions result from the specific “faceting” structure (Herring, 2007) – the affordances and the communicative situation – of the medium, together with an enhanced metapragmatic awareness arising from the textual nature of most CMC (Herring, 1999) (p. 9).

In other words, they believe that these new expressive forms arise from the combined influence of the nature of the medium and the awareness, on the users’ part, of the effects and conditions of the language itself, thereby giving rise to forms fit to fill the needs of users of CMC on a particular platform. However, this diverges somewhat from the theory of technological determinism, as these effects are ultimately more variable than categorical, manifesting in different ways according to different languages and cultures (ibid.). The human element interacting with the

computer is essential to understanding the creation of these forms, and although patterns of usage do develop, the ways they are implemented vary according to the linguistic background and culture of the user.

2.2.5 Freedom of expression

Also related to the cultural background of the CMC user is the issue of freedom of expression, or free speech. Cultural perceptions vary, but most Western societies hold this concept in high regard and therefore it is inevitable that the question of free speech would appear in relation to these developments in CMC. Barlow (1996) has discussed how this concept has been relevant since the inception of the Internet: “[the creators of the Internet] believed that the Internet structurally encouraged free speech, by ‘routing around censorship’”. It is debatable to what extent this was accomplished in practice, but the nature of the Internet today as a platform for CMC has retained much of these characteristics which inhibit censorship on an international scale. Of course, national governments have the ability to censor and prevent access to certain Internet sites, but these limitations are narrow reaching and fairly easy to circumvent in most cases. On a smaller scale, individual Web sites can and do censor content, but as of yet there is no single governing authority that can censor the Internet as a whole, meaning that anyone with access to a computer can potentially create their own sites without worry of censorship. Indeed, the Internet has been touted as a platform of equality, as individuals and groups who might not otherwise have a chance to speak freely can utilize the Internet as a public forum in order to make themselves heard (Herring, 2002, p. 141). This is a rather optimistic standpoint, as access to the Internet can be limited by economic status and government interference, but nevertheless, such instances are relatively rare.

On the other hand, the unconstrained speech on the Internet has presented its own challenges, such as the low signal-to-noise ratio (i.e., more low quality communication than high quality communication) and the abundance of unregulated anti-social CMC behaviors (e.g., spamming,

flaming, hate speech, and sexual harassment). This reality brings to light the other side of the issue of free speech: should the individual's right to say anything be limited in the favor of the common good (Herring, 2002, p. 141)? Of course, some sites already utilize filters and employ forum moderators, but to what extent do these forces dictate what can and cannot be expressed? The area of Internet censorship remains minute in the grand scope of the system, and the very nature of its design largely prevents permanent censure of any single individual. That being said, this study will look at a new form of censorship that has developed as a cooperative effort among users to facilitate a more personalized and comfortable experience on microblogging platforms.

2.2.6 Community, social effects, and large scale discussions

Beyond enabling a wider range of freedom of expression, the Internet also makes possible, on an unprecedented scale, simultaneous conversation among extremely large numbers of people (Herring, 2002, p. 146). These instances act as efficient one-to-many and many-to-many communicative forms, but as Herring (2002) pointed out, they are not without their own challenges:

Coherence is difficult to maintain, in that conversation management (turn taking, exchange tracking, topic maintenance) tends to be fragmented in multiparticipant groups. In the absence of a strong moderator, computer-mediated groups tend toward disagreement and polarization, making consensus among large numbers of participants difficult to achieve (p. 146).

The "virtual community" fostered by CMC on the Internet takes on a variety of forms, the most prevalent being a sort of hybrid community in which social ties are "intermittent and varying in strength" (Herring, 2002, p. 143). These communities highlight CMC's interactive broadcasting functionality within the context of the Internet, and the implications for its participants are particularly visible, unlike some other forms of CMC on the Internet. It is these characteristics of large-scale communication on the Internet which give rise to a variety of adaptations attempting to minimize their effect. This, conflated with the behavioral tendencies provoked by the nature of CMC, leads to a number of unique new discourse phenomena. The combination of unregulated

polarization, anti-social behavior, anonymity, and the desire to maintain a status quo within an online CMC community has led to a particularly innovative form of censorship, which will be the main focus of this study.

2.3 Censorship

To begin with, the nature of censorship must be examined, for it has become ever more multifaceted with the diversification of consumable media. Censorship as it stands is “the suppression or prohibition of any parts of books, films, news, etc. that are considered obscene, politically unacceptable, or a threat to security,” as defined by the Oxford English dictionary. In modern society, the concept of censorship typically carries a somewhat negative connotation, as it is typically associated in the public mind as a tool implicit in systematic oppression and power politics. Arsan (2013) indeed describes censorship in the traditional sense as “a mechanism executed by the ruling class to consciously and willingly prevent the public from attaining certain information” (p. 447). Governments can utilize various regulations to prevent the free movement and spread of information, which then become internalized by media personnel, resulting in a policy of self-censorship without systematic enforcement (*ibid.*). It is possible to follow the various policies of censorship enforced by various governments throughout history, as they all fall into four distinct categories as defined by Phillips & Harslof (1997): ‘manifest-defensive’, ‘manifest-offensive’, ‘latent-defensive’, and ‘latent-offensive’. These terms are further defined by Arsan:

In ‘manifest-defensive’ censorship, media content is cut, banned, deleted, or censored for the lofty interests of the state. [...] In ‘manifest-offensive’ censorship, the attempts of the ruling class or government to reproduce and spread a false reality through public relations strategies and spin doctoring are very common. There exists a conscious and willing attempt to strictly control and reproduce what is presented to the public in this form of censorship. What is denoted by ‘latent-defensive’ censorship is the seizure and complete control of media channels that meet the informational needs of a society. The term ‘latent-offensive’ describes the silencing of the values and beliefs of people by the cultural hegemony of dominant groups. (p. 448)

These four systems of censorship often exist in tandem, working together to regulate the distribution of ideas within a social sphere. It is a multilayered phenomenon which has multiple

side effects, perhaps most noticeably the culturally ingrained sense of self-censorship resulting from socio-political pressure.

However, this is only one side of the coin; the nature of censorship is far more complex than such a singular reading might suggest. It is present in all forms of consumable media as well as in personal interaction and conduct, and has uses ranging from diplomacy and tact to simply functioning as a tool to avoid problematic situations in a domestic sphere. This present duality stems from a shift in the meaning of censorship. In the traditional sense, censorship implies that a certain power structure is acting within a social sphere. Yet, as the nature of journalism and media becomes increasingly more accessible both in terms of consumption and production, the nature of censorship also naturally shifts. The formal power structures have less influence and censorship becomes more self-directed, motivated by individual beliefs rather than widespread political agendas. However, this is not to say that these socio-political power structures have lost all influence; rather, their influence is increasingly seen through instances of self-censorship (Arsan, p. 458).

On the other hand, while censorship is moving towards a more personal sphere, it remains a highly politicized subject as its implementation poses great risks in terms of the freedom of expression. As Filak et al. (2009) state, “the ability to express one's self in an unfettered manner is at the core of the Bill of Rights to the U.S. Constitution and is often viewed as a ‘natural right,’ thus not endowed by government but rather inherent to the human condition” (p. 368). This being said, in most democratic countries, there are protections in place to prevent the misuse of censorship in a political sense. Though their effectiveness is debatable, they are nevertheless present, and act as politically acknowledged guardians of the freedoms of speech and the press, particularly in American society, where these rights are considered central to the country's sense of cultural identity.

However, it must be noted that self-censorship tends to fall outside these legal guidelines as it typically presents in the form of self-directed omission, which is particularly difficult to identify or

legislate. In that sense, legislation concerning censorship is limited, as it cannot protect individuals from forces acting outside the law. This is particularly noticeable as it is employed in social interaction.

2.4 Gatekeeping and content warnings in mainstream media

In addition to omission, there also exists a form of media regulation that diverges from the typical conception of censorship. Gatekeeping practices in mainstream media are a form of censorship which allows consumers to selectively partake in various media forms, as most commonly seen with films, television shows, and video games. These practices are put in place in order to restrict access to sensitive topics in entertainment. For example, in the United States, films are under the regulation of the Motion Picture Association of America's (MPAA) rating system as administered by the Classification & Ratings Administration (CARA); their film rating system classifies movies according to content, with a rating scale as follows: General Audiences (G), Parental Guidance Suggested (PG), Parents Strongly Cautioned (PG-13), Restricted (R), and Adults Only (NC-17) ("Film Ratings", 2015). Similar administrations are in place for other forms of consumable media in most countries. For example, the Entertainment Software Rating Board (ESRB) regulates computer and video games in North America, and the TV Parental Guidelines, established by the Federal Communications Commission (FCC) regulates the rating of television content in the US ("ESRB Ratings Guide", 2015).

These regulatory systems are an often overlooked form of censorship which has been deemed socially necessary for the benefit of the public majority. The rating systems in place protect individuals from being exposed to inappropriate or disturbing elements that they may want to avoid for any reason, and ensure that children are not so easily exposed to age-inappropriate content. Furthermore, this type of gatekeeping is an elective form of censorship in that it relies on the consumer's discretion whether or not it is actually utilized, thereby sidestepping the typical pitfalls of the more insidious forms of censorship, in which the consumer has no choice. In this

sense, censorship through content warnings participates in the creation of a safe and informed environment for the consumption of media without posing the risk of harm through misuse.

2.5 Censorship and psychology

In a social context, particularly in the domains which concern this study, the argument on behalf of censorship is not necessarily based on concepts such as political correctness or even the desire to stay out of trouble; rather, there are multiple defenses of censorship for the sake of psychological safety. There are multiple psychological disorders which can be triggered by external influences such as images or text which contain particular sensitive topics. These exteroceptive triggers can elicit a panic response from individuals suffering from anxiety and mood disorders, eating disorders, and perhaps most famously, post-traumatic stress disorder (PTSD) (Johnson et al., 2014). For these reasons, this study will look at and analyze the usage and occurrence of tags associated with three common phobias which typically elicit a panic response: emetophobia, agoraphobia, and arachnophobia.

The immediate response to a trigger in an individual with one or more of these disorders is typically a panic attack. These responses are classified by the DSM-V as recurrent expected panic attacks (ePAs), which occur in the presence of a clear external trigger that can be identified by the subject and are estimated to account for approximately 60% of panic attacks (Shulman et al., 1994). The symptoms are described by Johnson et al. (2014) as follows:

ePAs are comprised of an abrupt surge of symptoms that typically consist of cognitive symptoms like catastrophic fear of dying and/or losing control alongside autonomic-related symptoms such as palpitations, racing heart, hot flashes (heat sensations), and sweating. Other physical symptoms may include increased respiratory drive, paresthesias (numbing or tingling sensations), dizziness, and nausea. In the DSM-V, a PA consists of 4 or more of the 13 possible symptoms; two of these are cognitive/fear symptoms and eleven are physical symptoms. Typical PAs peak within ~10 min, have a duration of 25–45 min, and can occur during calm or anxious states (p. 430).

These panic attacks are highly distressing occurrences and individuals go to great lengths to avoid contact with specific exteroceptive triggers. This is where proponents of censorship base their argument: the psychological impact of sensitive topics in media is too dangerous to remain unregulated. This isn't to say that any and all sensitive material should be eliminated; rather, there is a necessity for different practices to sequester and properly label potentially triggering media (i.e. trigger and content warnings). This sentiment is particularly prevalent on social media sites where the user base controls the production and spread of content.

2.6 Self-censorship and social media

Self-censorship is defined by the Oxford English dictionary as “the exercising of control over what one says and does, especially to avoid criticism.” It is essentially an internalized form of censorship, directed at oneself for individual benefit. It can take multiple forms, but thus far self-censorship has been discussed most expansively in its presence in the general media, where it exists almost solely in the form of underreporting. As Germano & Meier (2013) state: “Self-censorship occurs when too few outlets are active in the market; in equilibrium they all endogenously choose the lowest possible level of reporting on the sensitive topic” (p. 121). This is censorship through omission, the most prevalent mode of self-censorship in professional forums, but by no means the only form.

As social media expands, a new type of self-censorship present in forms other than direct omission has become increasingly prevalent on microblogging sites such as Tumblr, which allow for the consumption of diverse media through a more complex platform. Here, the content is still being produced and posted, but through the use of metatags, the original poster allows themselves to be censored, should another user not wish to see or engage with content in that specific category. In this situation, self-censorship becomes a sort of dialogue between an individual and a social contract: they manually add classifications to their content to facilitate the comfort of other users

on the site, creating a safer, more personalizable atmosphere for that social media outlet. This form of self-censorship is a cooperative act, which could be perceived as either neo-common courtesy, or the relinquishing of certain inherent ideals of free speech, as the content creator thereby willingly limits their audience on the basis of topic and subject matter rather than the content or its message.

2.6.1. *Public and private self-censorship*

Again, the concept of self-censorship may at first appear to be contradictory to the conventional notion of censorship being a conflict between a censoring authority and a censored communicator. In self-censorship, the communicator also takes on the role of the authority, thereby operating independently of any external agents, at least at first glance. Mark Cohen (2001) considers this issue, and asserts that self-censorship is able to function in such a way since an individual is able to internalize public standards of censorship and apply them to themselves (p. 9). Cook and Heilmann (2013) further examine this issue on two fronts: public and private self-censorship. If public self-censorship is to be understood as a censee's response to an external censor, such as a legal code or social structure, then private self-censorship can be defined as a censee's response to an internal censoring agent, be it a personal code of conduct or set of individually formulated beliefs (p. 187). There may indeed be present an active dialogue between these two agents since external censors can directly influence internal beliefs, but their difference lies fundamentally in their motivation: the active agent in public self-censorship is the relationship between an individual and society, while the active agent in private self-censorship is the individual's relationship with themselves (ibid.).

These two forms of self-censorship can be seen in the popularization of trigger or content warnings on internet websites centered on social interaction. The first instances of content warnings in their most recognizable form appeared in online forums in the transitional period between the 1990s and the early 2000s, as seen in the forum archives of websites like

msmagazine.com. They subsequently gained traction on a site called LiveJournal, where users posted journal updates, fan fiction, and other samples of original creative writing. However, content warnings in this form did not become popularized as common practice until the advent of Tumblr. The spread to this particular microblogging site can most likely be attributed to shared demographic and format of both sites; yet, where LiveJournal was stagnant, Tumblr allowed for the rapid sharing and spread of content and ideas.

Both public and private self-censorship is reflected in the usage of content warnings on these two websites. Most evidently, perhaps, is the private element, as the addition of warnings onto self-produced content is reflective of the censee's internal moral landscape under the scrutiny of a censoring agent. The censorship – or aid to censorship – is self-directed and therefore private. However, there is also a public element at play; as the concept of trigger and content warnings spreads throughout the website the practice becomes standardized amongst its users. This results in a social pressure to adhere to the established norms, and many users who engage in such self-censorship are doing so in reaction to these new social constructs. In this way, self-censorship has become a part of the standard social interaction on these particular websites.

2.6.2. Microblogging as a mediated social practice

Microblogging has developed in the past years alongside social networking to become an increasingly popular mode of social interaction as a mediated social practice. However, before microblogging can be analyzed in the context of usage, the term 'mediated social practice' must be defined. Van Dijck (2011) explains the phenomenon as such:

Social practices are everyday activities that have a routine status, for example 'talking to friends' or 'updating oneself about the world'. Once routine practices are permeated by a specific ICM [information, communication, and media] tool we call them mediated social practices, for instance 'talking to friends on the phone' (p. 336)

According to this definition, microblogging platforms fall under the category of ICM tools, as they are used to perform a social action, and thereby microblogging can be classified as a mediated social practice. Furthermore, there is significant interplay between function and platform, as media tools are influenced by social practices, and social practices are influenced by media tools. This is exemplified by websites such as Google, where the brand name has become synonymous with the social practice: googling has come to be understood as an equivalent to internet search (ibid.).

It is the functional nature of microblogging platforms to offer a variety of ways to engage in social interaction: specifically two-way communication, one-to-many communication, and many-to-many communication. These functions are used to carry out a number of different activities, such as conversation, collaboration, self-expression, status updating, and information sharing (Van Dijck p. 337). Van Dijck (2011) further states that microblogging has undergone a shift away from two-way communication in favor of one-to-many and many-to-many communication, both which are more platform-specific methods of engaging in social action (p. 339). It is in this way that the ICM tool, the microblogging platform, has influenced the development of the social practice.

In the context of microblogging as a mediated social practice, self-censorship has developed into yet another aspect of the format in which users engage in one-to-many communication on microblogging sites, regardless of which activity the function is carrying out. Furthermore, the form which this self-censorship takes, in this case gatekeeping through content warnings, has been directly influenced by the format of the social practice. Here, language has taken on a tertiary function as an agent of censorship, in addition to conveying meaning and expressing tone. This has been accomplished partially through the nature of the platform, which allows for highly customized interaction, and partially through the nature of its users: individuals who create, consume, and share original content with the intention of casual communication and entertainment.

3 Data

The phenomenon of cooperative censorship is multi-faceted, and approaching it requires a practical knowledge of all factors involved. To explain briefly, individual users on the website Tumblr create blog entries or ‘posts’ which are comprised of two entities: the main body of content (text) and the metatags or ‘hashtags’ (tags), which serve an executive organizational function. All posts that are marked with a specific metatag are grouped together and searchable via that tag. In response to this functionality, a user-created browser extension was developed and widely downloaded, which allows users to blacklist their choice of metatags, meaning any post marked with that specific tag would be automatically hidden from view. This has led to the development of a system of trigger or content warnings which take the form of metatags, thereby enabling Tumblr users to participate in a form of cooperative censorship in which the original poster (OP) either allows or refuses censorship through the usage or exclusion of these metatags. This section will further describe the database from which the corpus was collected, explaining the nature of the data and its organization, as well as the exterior elements which contribute to its presentation and usage.

Section 3.1. will briefly explain the functionality of Tumblr and its role as a social media website, including the technical aspects of its usage. Section 3.2. will then further elaborate on the website’s system of metatags, and section 3.3. will introduce the browser extensions which interact with these metatags.

3.1 Tumblr as a microblogging platform

In practice, Tumblr is a social networking website comprised of microblogs or “Tumblr blogs”; there are currently over 223.6 million blogs which have produced over 104 billion posts, with over 80 million new posts being added daily (About: Tumblr, 2015). Tumblr was founded in 2007 as a

content-sharing site and allows users to create posts in multiple different media, including text, picture, video, audio, and link posts, which other users are then able to “like” or “reblog”. Reblogging is essentially re-broadcasting the original content on another blog while retaining credit to the original poster; this functionality is similar to Twitter’s retweet function or Facebook’s post-sharing function. This allows the content of Tumblr blogs a remarkable degree of mobility, with posts propagating through both reciprocal and non-reciprocal social networks and potentially reaching millions of users with the right exposure (Chang et al., 2014). Nevertheless, Tumblr is still considered a microblogging platform as the nature of the content is not as extensive as is typically required on blogging sites such as WordPad and LiveJournal.

Each Tumblr post is comprised of two elements: the content and the metatags. As stated, Tumblr utilizes a system of user-added metatags or “tags” to categorize content, and thereby individual posts can be catalogued and searched for using keywords, and the results organized according to popularity or age. The functionality of metatags will be further discussed in section 3.2.

The data collected from Tumblr will form the main corpus for this study. Tumblr was chosen as a primary source for data for a number of reasons, the first and most significant being the involved usage of post structure modifications and metatags, and the browser extensions that have been developed to enhance their functionality as both agents of organization as well as censorship. Furthermore, Tumblr blogs can be made either public or private, meaning all searchable posts are public content, which makes the data collection somewhat less involved, as there are fewer issues of privacy and research etiquette to be taken into account. The data can be accessed through the site’s own search function or by manually accessing the site’s archives, both of which are publicly available and contain all catalogued posts uploaded to the site. This data is persistent and can be re-accessed at any time via a designated url.

3.2 Metatags

As previously stated, Tumblr employs the usage of metatags, also known as hashtags or tags, in order to catalogue and organize posts, both on individual microblogs and on the site as a whole. Individual blogs are searchable with the simple addition of `"/tagged/keyword"` to their home URL; all the posts the user has tagged with that specific keyword will be listed. All Tumblr blogs can be searched in a similar fashion, by either using the search bar at the top right hand corner of the webpage, or by manually inputting the desired information as an addition to the URL (e.g. `"tumblr.com/tagged/keyword"`). On posts, metatags appear at the very bottom following a hashtag. These tags are added by the original poster and are by no means mandatory, but are frequently used for organizational purposes. On posts that a user reblogs, metatags may still be added, but these only exist within the users personal blog and will not make the post appear an additional time in site-wide searches. However, metatags on reblogged posts are frequently used to facilitate censorship through browser extensions, as will be discussed in the following section.

3.3 Browser extensions

Browser extensions are essentially downloadable third party computer programs which enhance the functionality of a web browser or website in some way. As Tumblr grew in popularity, so did the need for customizable enhancements that improved functionality issues which remained unaddressed by the original developers. Thus, free add-ons such as Missing-E and XKit became widely downloaded and utilized. With these programs came a blacklisting function: a computer script which filters out posts through the use of metatags. The individual user can input a certain word or phrase, and once this is 'blacklisted', any post with that specific word or phrase in its metatags will not show up in the user's feed. Because only tagged posts are filtered out, there grew a demand within the community to actively utilize metatags for the convenience of other users — in essence, to accommodate a form of selective, cooperative censorship.

4 Methodology

The methodological approach to this study will be twofold; it will be composed of a first stage involving diachronic frequency and content analysis, in conjunction with a second stage of further corpus linguistic analysis of the data with a theoretical grounding in computer mediated discourse. This research strategy provides a solid foundation for the study of the phenomenon in question, keeping in mind its relative newness and context-specific nature. In this section the research method of the study will be presented with a short explanation as to the decisions made regarding the data collection process and the selection of the areas of focus. Section 4.1. will explain the specifics of the first stage of this study, including the two-part analytic approach. Section 4.2. will then describe the second stage, its relevance to the first stage, and the justification for the choice of dataset.

4.1. Frequency and content analysis

In the first part of this study, the data was collected through a manual site-wide search of common metatags employed in censorship and self-censorship. Because of the nature of the data of interest, the selection of tags remains unavoidably subjective. The selection was made on the basis of over four years of passive observation as an active Tumblr user to determine some of the most frequently encountered trigger warnings, but as active engagement with the site naturally limits one's exposure to only a certain number of social spheres, it is unlikely that a fair representation of the site as a whole can be achieved. Thus, since a trigger is to be understood as a specific psychological phenomenon, the decision was made to limit the tags to the most common keywords associated with content that includes potentially harmful or disturbing elements. In this study, the parent tags #emetophobia, #agoraphobia, and #arachnophobia will be examined alongside the most common variations of each.

Once the parent tags and tag variants were decided, the data was manually collected. In order to gain access to the full data set and its metadata, it was necessary to employ the browser extension XKit previously discussed in section 3.3, which can be used to display information typically hidden by the standard interface. In this case, it was used to add a timestamp to each post accessed. This revealed the exact time and date each post was posted to the platform. Furthermore, since Tumblr's search function as it appears on the site is limited to popularity and real-time based archival methods via a scrolling page format, it was necessary to manually access the full chronological archives of each keyword as well as use XKit to change the scrolling interface to a page-by-page format. The chronological archives of any given keyword can be found by changing the url as follows:

<https://www.tumblr.com/search/keyword> → <https://www.tumblr.com/tagged/keyword>

These three modifications were necessary for the accurate and complete collection of the data set, which was then manually collected into a text file, preserving their chronological order, content, and relevant metadata. Post frequency was calculated from this data set by counting seven days' worth of posts from each month in order to determine the average number of posts per day over each year.

So as to glean a more comprehensive understanding of the nature of censorship on Tumblr, the data set will undergo analysis on a number of different levels in the scope of this study. The usage of these metatags will be analyzed first according to chronological frequency, determined by calculating yearly averages. Frequency within the data set as a whole will also be looked at, and possible trends across qualitative parameters will be further analyzed. Subsequently, a sample of individual posts from each topic will undergo a closer qualitative analysis in order to examine the possible evolution of the functionality of cooperative censorship in practice. It is important to note that due to the gargantuan size of the prospective data pool, it is impossible to take into account every instance of each specific tag in this manner; rather, a small sample taken at regular intervals covering the chronological lifespan of each tag will be presented.

4.2. Corpus linguistics

The second part of this study, the data was again collected manually from the four variants of a single parent tag. In this case, every single post in each tag was collected chronologically. The posts were then sorted by media type, since pictures, videos, and other multimedia forms are outside the scope of the study. The text posts, 568 in total, were compiled into a corpus of 41,106 words. The data was manually gathered over the course of several months in 2015 and 2016, and spans the years 2011-2015. Though this corpus isn't particularly large, it consists of every single instance of each tag variant and therefore gives a complete scope of the metatags in question. The tags in question are the four variants of #arachnophobia: #arachnophobia cw, #cw arachnophobia, #arachnophobia tw, and #tw arachnophobia. Arachnophobia was chosen out of the three tags analyzed in the first section due to a variety of factors. Out of the three, arachnophobia was the first to produce a censoring variant, it is used frequently enough to provide a sizeable dataset without being overwhelmingly large, and the content is more often text-based compared to the other metatags. The concordance of these three factors made the arachnophobia tags the most suitable for analysis considering the scope and focus of this study.

The four tag variants of arachnophobia will first undergo word frequency analysis in order to determine the specific vocabulary associated with the tags and the nature of the content being produced. This will be accomplished through a comparison to other standardized corpora as well as a sample of the much larger parent tag (#arachnophobia). Based on this analysis, the collocations of specific words will then be further analyzed in order to gain a more comprehensive understanding of the language in context.

5 Analysis of censorship through metatags

This section will present a two-fold analysis. The first stage involves both a quantitative and a qualitative analysis of the data set in terms of frequency and content. The development of a stringent metatagging system will be examined in the context of cooperative censorship, and the linguistic significance of using language as an agent of censorship rather than its object will be considered. Finally, in the second stage of the study, the usage of cooperative censorship will be further analyzed through a more thorough look at the usage of a specific set of metatag variants through the use of corpus linguistic analysis.

Section 5.1. looks at the development of the metatagging system and how the format has evolved over time. Section 5.2. presents the comparative chronological frequencies of each metatag in relation to the parent tag, and section 5.3. presents a more in-depth qualitative analysis of select posts from the most frequently used censoring metatags. Section 5.4. will then examine the comparative word frequencies of the arachnophobia tag variants, and further analyze a number of collocates and n-gram clusters as determined by the previous section.

5.1. Development of metatags in cooperative censorship practices

As mentioned before, the function of metatags on Tumblr first began as an organizational tool, as they were used to group similar posts into a space where they could be easily searched and catalogued. However, as is the nature of social media, the platform exists to be utilized and manipulated by the consumer, and as such, the systematic usage of metatags began to deviate from its original purpose, taking on secondary and tertiary functions which spread throughout the online communities and came into common practice. Their development and spread can be tracked by searching through each specific metatag, as the mechanism acts as an archival system.

The original or “parent” tag denotes a metatag without extraneous elements indicative of the intention to cater to blacklisting scripts. In this study, the parent tags are the following: #arachnophobia, #emetophobia, and #agoraphobia. Under each parent tag falls a variety of different iterations of the same concept, marked by either *tw* (trigger warning) or *cw* (content warning). Though associated, each of these variations presents as a unique metatag and can be searched or blacklisted independently of the others.

Table 5.1. First instances of metatag usage.

Metatag	Variations	First Occurrence
		D.M.YXXX
#arachnophobia	#spider	14 07 2008
	#spider tw	07 12 2012
	#spider cw	29 03 2013
	#tw spider	06 02 2012
	#cw spider	16 07 2014
	#arachnophobia	02 09 2009
	#arachnophobia tw	04 08 2011
	#arachnophobia cw	07 08 2012
	#tw arachnophobia	16 07 2012
	#cw arachnophobia	09 07 2013
#emetophobia	#emetophobia	26 05 2010
	#emetophobia tw	02 05 2012
	#emetophobia cw	28 08 2012
	#tw emetophobia	11 01 2012
	#cw emetophobia	12 01 2013
#agoraphobia	#agoraphobia	25 01 2010
	#agoraphobia tw	31 07 2013
	#agoraphobia cw	30 01 2013
	#tw agoraphobia	24 07 2012
	#cw agoraphobia	19 02 2014

Above is a chart of the dates of the first instances of each specific metatag. Notably, each parent tag was first used years before any of the variations associated with censorship. This is to be expected, as when Tumblr was first established in 2007, the tagging system was strictly organizational. However, as the site's popularity increased and its usership began expanding, there came a demand for additional functionality unable to be met by the site. Thus, users began writing their own scripts for browser extensions. The two most popular plug-ins offering censorship functionality in the form of a blacklisting function are Missing E and XKit. Quite notably, XKit was first released for download in 2011, and Missing E in 2012. These two years coincide almost exactly with the advent of all variations of each tag marked by either *tw* or *cw*. The earliest use of *tw* or trigger warning in any of the associated tags occurs in 2011, when the first post using the tag *#arachnophobia tw* appears. Indeed, tags marked with the shorthand for trigger warning (*tw*) predate almost all instances of the term content warning (*cw*), suggesting the term "trigger warning" was introduced to the social environment a measurable amount of time before the term "content warning". Perhaps this is due to the fact that the practice of censorship through language was born on a different platform, LiveJournal, which does not share Tumblr's gatekeeping capabilities; there the warning system in place was used to inform the reader of possibly triggering content, hence the name "trigger warning". Naturally, as Tumblr's functionality was supplemented, this became the initial term for censorship due to the shared demographics between the two sites.

Furthermore, due to the differences in denotation as well as date of presentation, it can be expected that any tag marked by either *cw* or *tw* contains content relating to the tag, while the parent tag can be expected to contain more topic-specific content. For example, based on observation, *#emetophobia* appears more likely to contain content concerning the actual phobia, while *#tw emetophobia* is more likely to contain mentions of vomit or other possibly triggering content for those who suffer from emetophobia. It is a slight distinction that may not be apparent or even applicable in other metatags, but the nature of emetophobia allows for such separation

and should be taken into consideration. While it remains outside of the scope of this study, it would benefit from further analysis. Such differences will be further considered in section 5.3.

Finally, the presentation of these variations indicative of self-censorship also point to one other peculiarity of self-censorship in social media. Contrary to Germano & Meier's (2013) assertion that self-censorship occurs when too few outlets are active in the market (p. 121), it is apparent that within the public sphere where the consumer and producer are one and the same, the increase of active Tumblr users has coincided with the increase in these self-censorship practices. This is perhaps due to the differing natures of commercial and casual media, wherein "official" news sources are subject to a certain power structure as well as competition, whereas "casual" media hold no obligations other than to its creator. The duality between public and private pressures on social media platforms has created a unique atmosphere, where the individual user is subject primarily to their own moral censor, but secondarily to established social pressures within the network itself (Cook & Heilmann, 2013, p. 183).

5.2. Comparative frequency analysis

It is possible to track the popularization of censorship-conducive metatags through chronological frequency analysis. In this section, the multiple variations of each parent tag will be analyzed through a comparative look at their relative frequency, starting with #emetophobia in Section 5.2.1., continuing on to #agoraphobia in Section 5.2.2., and finishing with #arachnophiba in Section 5.2.3.

5.2.1. #Emetophobia

The metatags for emetophobia, or the fear of vomit, present an apt opportunity for the comparative analysis of frequency, as the parent tag and its variations are both significantly active. The data set is not overwhelmingly large, nor is there a great size discrepancy between the parent tag (#emetophobia) and its most popular variation (#emetophobia tw). Thus, the results of the frequency analysis can be presented alongside each other on the same graph, aiding in the visualization of the data sets.

The graph on the following page illustrates the frequency of each tag in average number of posts per day over the course of five years, based on samples taken from each tag at regular intervals. The parent tag has a distinct positive slope, which is mirrored by the tag variations on a smaller scale. Most closely resembling the trajectory of the popularity of the parent tag is that of its variation #emetophobia tw. Comparatively, its popularity is 39% of the parent tag, a noticeable difference from the other variations, which range from 0.004% to 13%. This is an interesting divergence from the previous section: #tw emetophobia was born nearly four months before #emetophobia tw, and yet it is swiftly eclipsed in popularity. This suggests that the latter format was accepted over the former as the site standard within multiple social circles, thus contributing to its popularity, as the same pattern is reflected within the other sets of metatags, albeit not as strongly.

The patterns present also indicate a shift in the nature of communication, as noted by Van Dijck (2011), who asserts that microblogging has undergone a shift away from two-way communication in favor of one-to-many and many-to-many communication (p. 339). The popularization of #emetophobia tw alongside its parent tag is reflective of this. As the main function of this tagging system is to aid self-censorship for the sake of other users, it indicates that the content creator is socially conscious of the accessibility of their content through the microblogging platform, and are thus catering to the sense of public security rather acting than for the sake of a singular person, as is the case with two-way communication.

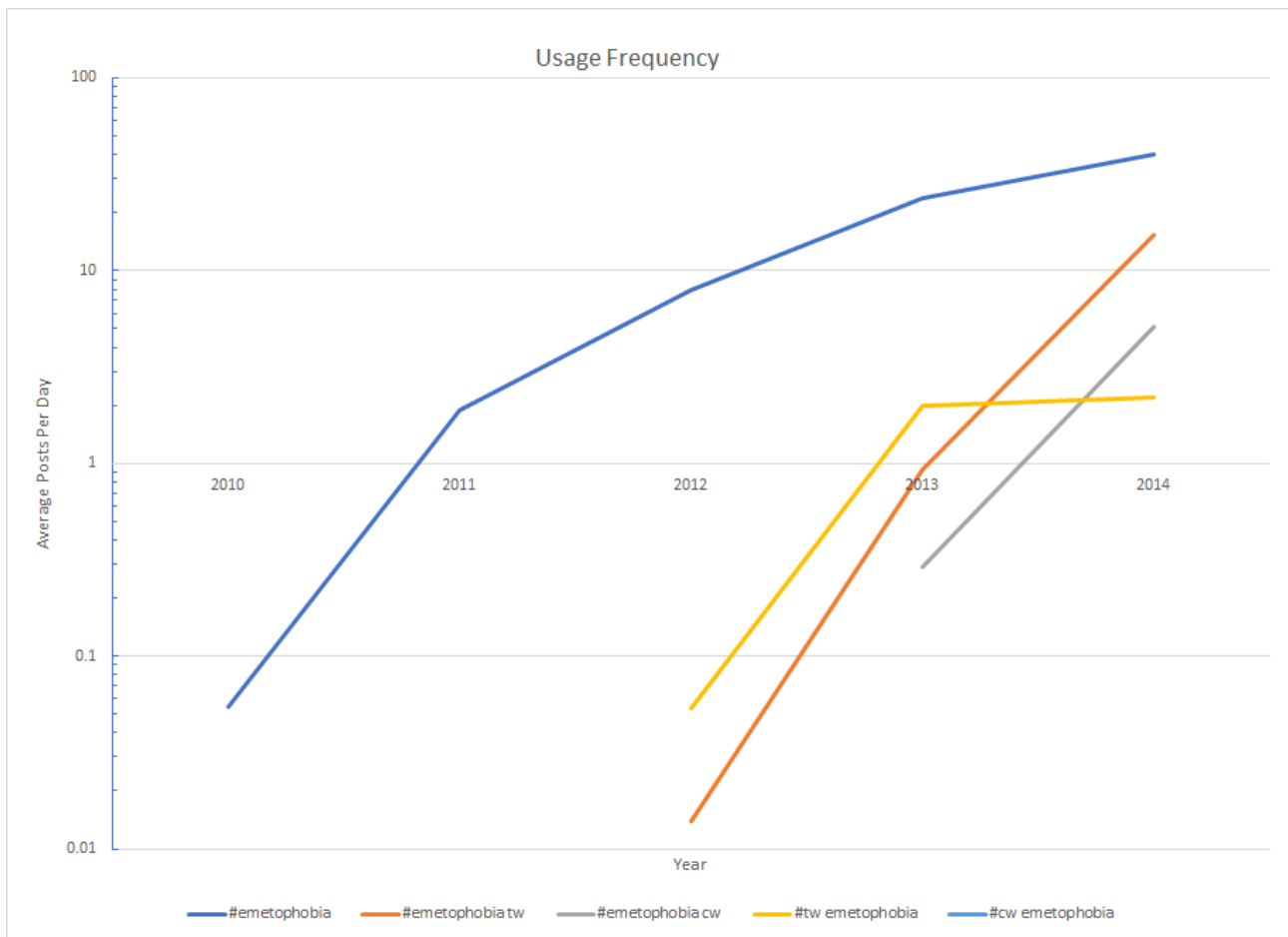


Figure 5.1. Frequency of usage of #emetophobia and its variations (2010–2014).

It is only within the last year or so that trigger and content warnings have become more popular, but the growth has been exponential. This hints at a larger shift in the collective social conscience of Tumblr’s main demographics. Whether it is indicative of some independent moral revelation or the result of social pressure and policing, it is impossible to say; nevertheless, it has cemented into an expected social practice within the framework of this microblogging website.

5.2.2. #Agoraphobia

Unlike #emetophobia, the parent tag for the agoraphobia cluster (#agoraphobia) is marginally more popular than any of its variations, most likely due to the fact that agoraphobia is more common than emetophobia, and therefore there is more topic-specific content to be found. Below, Figure 5.2 illustrates the frequency of each of the variations, but leaves out the trajectory of the parent tag, as the focus of this study is the variations, which are markedly less common than the parent tag. (In 2014, the average number of posts per day in the #agoraphobia tag was 31.6, in comparison to the most popular variation, #agoraphobia tw, which only has 0.12.) Nevertheless, the differences between each variation become quite apparent when compared side by side.

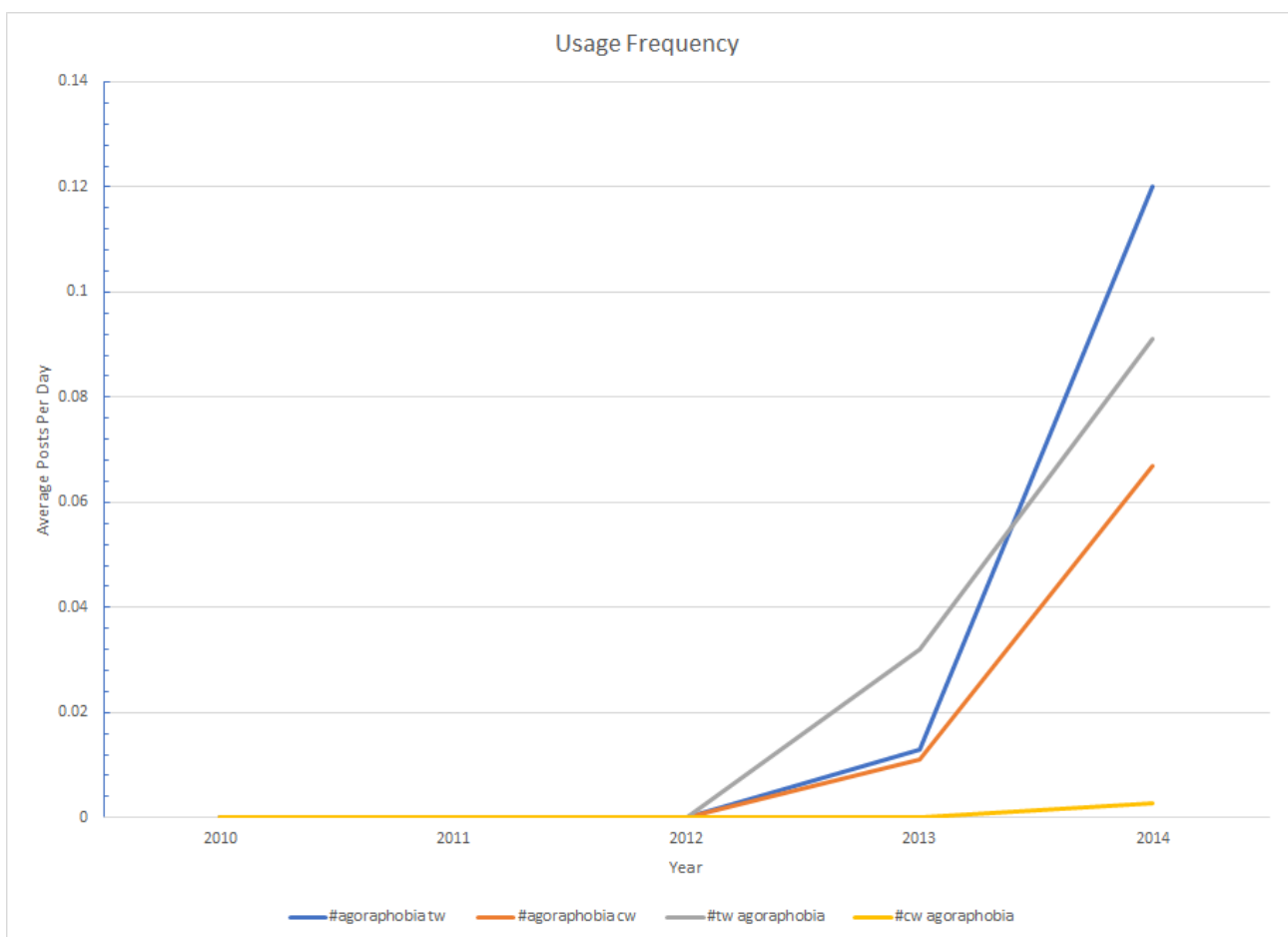


Figure 5.2. Frequency of usage of agoraphobia trigger and content warning tags (2010–2014).

Here becomes evident a great discrepancy between #cw agoraphobia and the other three variations. In fact, there exists only one post tagged #cw agoraphobia, made in 2014. This is due in part to the relatively low usage of any tags associated with agoraphobia as a trigger or content warning, and also perhaps due to the fact that the format #cw-TAG has shown itself as a markedly unpopular format across all data sets. The other censoring tags, however, all experienced a marked amount of growth considering the low popularity of the agoraphobia cluster in general. This also may be explained by the fact that due to the nature of agoraphobia, (the fear of being in crowds, public places, or open areas) it is markedly easier to produce content discussing the condition than produce content evoking the condition. Nevertheless, the trigger and content warning variations have still experienced growth along with the other clusters.

5.2.3. #Arachnophobia

A similar situation occurs with the arachnophobia cluster of metatags as with the agoraphobia cluster: the popularity of the parent tag widely eclipses that of any of its variations. In this case, it is most likely due to the ubiquitous nature of arachnophobia, as a topic, a condition, and a pop culture phenomenon (i.e. the 2014 film titled *Arachnophobia*; in this instance the tag would be topic specific rather than content specific). Basically, there is a wide range of cases where #arachnophobia is used in relation to content other than the condition itself. That being said, the development of the censoring varieties of the metatag can nevertheless be analyzed.

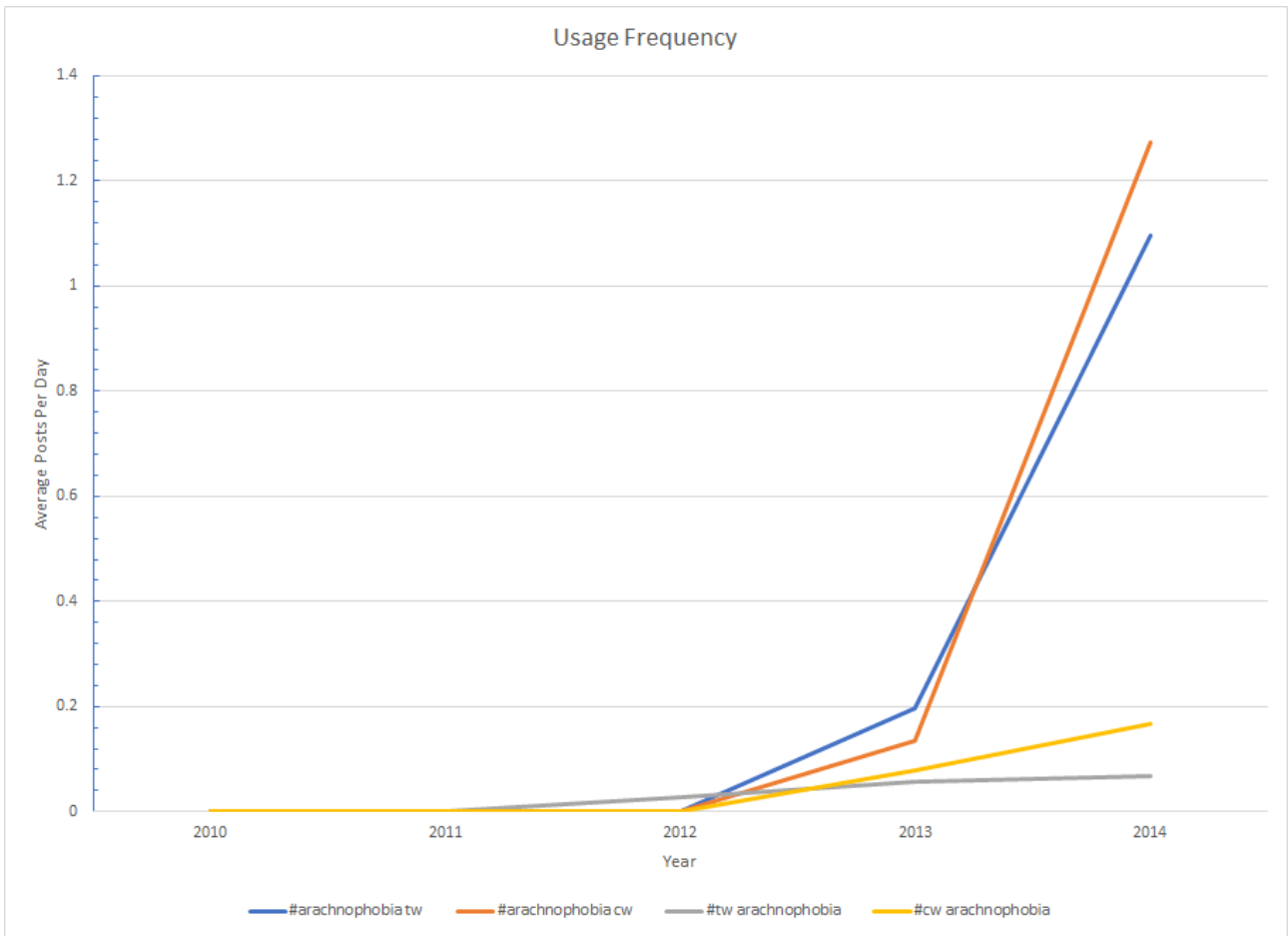


Figure 5.3. Frequency of usage of arachnophobia trigger and content warning tags (2010–2014).

In this case, the popularity of the two tags employing the format #TAG-t/cw greatly outnumbers that of the other two, as has been the case with most other metatag clusters (1.273 and 1.096 posts/day compared to 0.168 and 0.068 posts/day). The reason for this peculiarity might simply be attributed to the linguistic structure of the unit; “arachnophobia trigger warning,” where “arachnophobia” is acting as a descriptor, is somewhat more naturalistic than “trigger warning arachnophobia”.

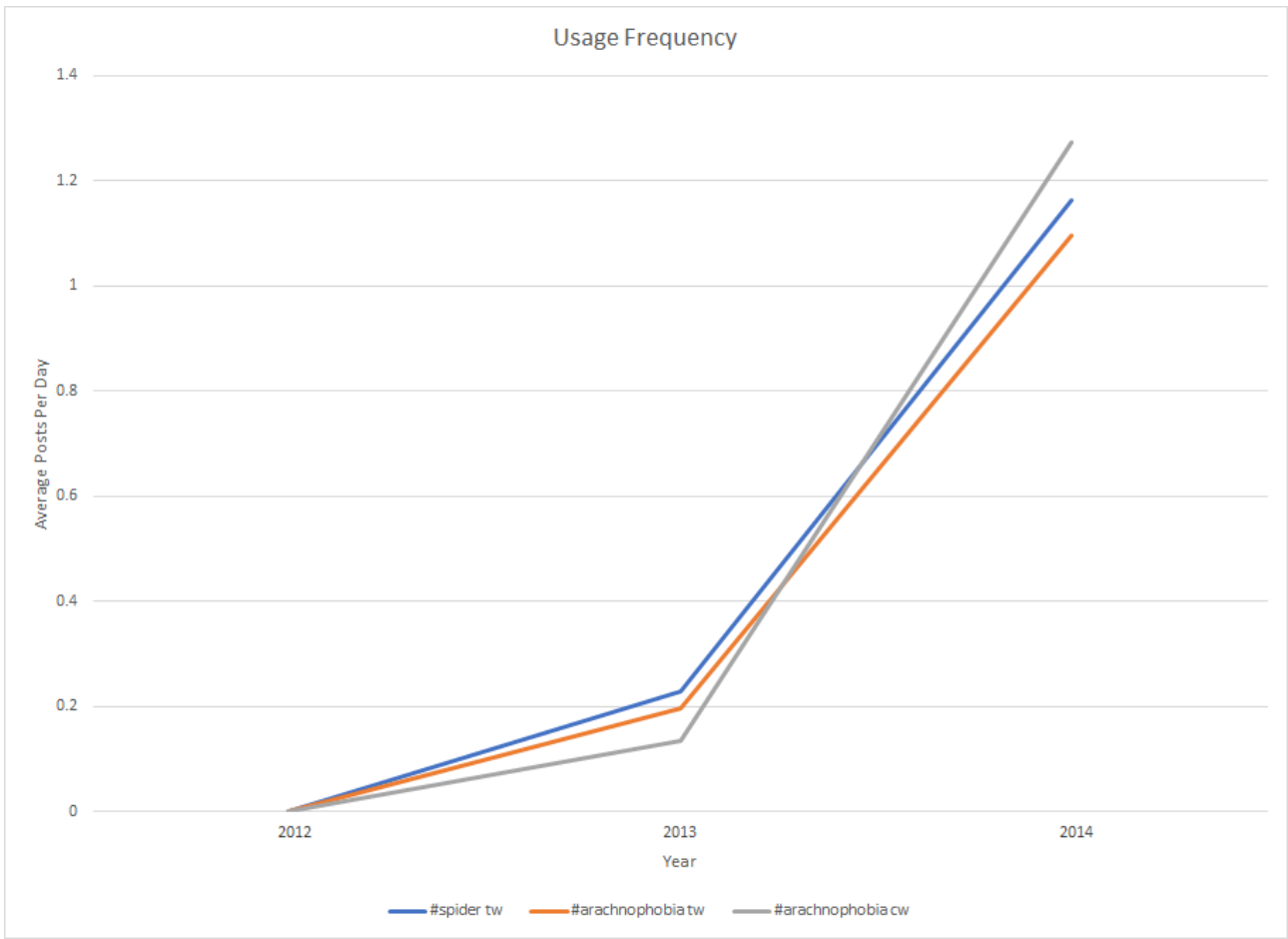


Figure 5.4. Comparative frequency of #arachnophobia tw, #arachnophobia cw and #spider tw (2012–2014).

Also notable is the nearly interchangeable popularity of #arachnophobia tw and #arachnophobia cw (as seen in Figure 5.3), and how closely both tags compare with another, almost identical metatag: #spider tw. As shown in Figure 5.4, the far simpler alternative has also gained in popularity in a nearly identical trajectory. This might be expected due to the same subject matter, but the preference of the #TAG-tw/cw format is also reflected in the spider metatag cluster.

5.3. Trigger warnings in use

This section presents a closer examination of a selection of individual posts which employ the most popular of the aforementioned censoring metatags. The content of each post will be considered in the context of the censoring function of the tag, as well as the overall role of censorship within each post. Again, Section 5.3.1. will present examples from the emetophobia clusters, Section 5.3.2. will examine a sample found in the agoraphobia tags, and Section 5.3.3. will deal with the arachnophobia clusters.

5.3.1. #Emetophobia tw

Individual posts found in the parent tag, #emetophobia, will first be examined so as to establish the general format of the content and to serve as a baseline for further comparison. The following example is from 2011, quite near the beginning of the tag's lifespan:

```
(1) February 19th 2011  
  
so anxious  
I hate being emetophobic. I hope today gets better.  
  
#emetophobia #anxiety
```

This post is a short personal expression relating to the topic of emetophobia. There is no risky content that might potentially trigger a panic response in another individual with emetophobia; rather, it is a short quip expressing the poster's personal relationship with the phobia. Therefore, it is understandable that it tagged #emetophobia without any warning labels attached. This topic-specific style of content is most common in the first few years of the tag's existence, as is reflected in the following example:

```
(2) June 16th 2012
```

I woke up and started to panic a little bit so I guess I'm on tumblr
at 4am

#emetophobia #panic disorder #the magic life of yuui

In this instance, there is no mention of emetophobia in the main body of the post. Instead, it is only added on as a tag, and thereby it can be inferred that the speaker's anxiety stems from their emetophobia, even though they have not stated it explicitly. This is an intensely casual style, possibly expecting that the readers of this particular microblog are already familiar with the original poster. In essence, this post is a quick journal-style status update, where emetophobia is relevant but does not pose a threat to any potential reader.

Generally, throughout the lifespan of the #emetophobia tag, most posts have been in this same vein, content-wise. However, within the past few years as the website and user base have expanded, a different style of post has become increasingly common. The following is one such example:

(3) February 1st 2015

...heterosexuals...*gagging sounds*

#emetophobia #text #ghost.txt

This particular post is short, yet contains a potentially triggering element for severe emetophobes: “*gagging sounds*”. With this in mind, it is clear that the tag #emetophobia is meant to serve as a warning to users sensitive to the subject, since the content of the post is not relevant to the discussion of the phobia itself. However, while #emetophobia is intended to function as a trigger tag, the execution is somewhat inappropriate, as this potentially triggering post is now also catalogued and searchable in a space where people talk about the phobia. This is an intrusion into a potentially safe space that has become increasingly common, as have posts complaining about this fact and urging other users to use a variation of the tag instead (e.g. #emetophobia tw, #tw emetophobia). The following two examples will examine the usage of #emetophobia tw in this context.

(4) May 16th 2013, 10:34:39 pm

i have this big math final tomorrow that i still havent studied for
lmao all ive done today is throw up 5 times and lay around in pain

#emetophobia tw #i hope im tagging that right

This post is a prime example of the cautionary metatag variant functioning as it is intended. The post mentions vomit or vomiting, which can trigger a severe response in a person suffering from emetophobia. The user, conscious of this, has added the tag **#emetophobia tw**, as well as a short expression of doubt, indicating their intent: they aim for their post to be censored by those who don't wish to see its content. If a user has the tag **#emetophobia tw** blacklisted, the post will be blocked and the content will not show up. In this way, the original poster is enabling themselves to be censored for the comfort of others, as is the case in the following example:

(5) April 30th 2014, 8:29:29 pm

cat food commercials where they try to show the food as like super
tasty and fancy and look at this awesome food, but you're just like
that is wet cat food and looks approximately like half digested
vomit.

#emetophobia tw #wet cat food is one of the most unappetizing looking
things to me #also it makes a shlorp sound

Again, the post in consideration contains a potentially disturbing element: the phrase "half digested vomit". However, it is marked for the presence of the word *vomit* rather than a reference or account of an actual event. Even though it is contextually innocuous, the poster has still added the tag so as not to upset anyone sensitive to such vocabulary. They are implicit in their own silencing, as the utilization of language has become both the object and the agent of self-directed censorship.

5.3.2. #Agoraphobia tw

As with the emetophobia metatags, similar patterns are present when considering the agoraphobia set. Again, the parent tag will first be examined before moving on to its variations. The following example is structurally similar to the posts found in the #emetophobia tag:

(6) January 20th 2013, 9:26:26 pm

Wanting to do something, anything. To maybe even go somewhere, anywhere. But not having the ability or the drive or the energy to even try.

#agoraphobia #agoraphobia things #social anxiety #social anxiety disorder #depression

Here, the original poster is again expressing their personal feelings as a sufferer of agoraphobia. The content is topic-specific and therefore relatively benign to other individuals also suffering from the same phobia. Therefore the content is not such that requires warning. Furthermore, another notable trend in the #agoraphobia tag is the high co-occurrence of tags related to anxiety and panic disorders (#social anxiety, #depression, etc., as seen in the example above). This is quite prevalent throughout the lifespan of the tagged posts, hinting both at the nature of the phobia as well as the comorbidity of these conditions.

(7) April 28th 2014, 1:48:11 am

Just thinking about going outside for a walk makes my anxiety act up. I cant even walk my dog new dog.

#agoraphobia #debilitating #anxiety #failure #hate myself

Here again are multiple familiar elements: the personal, conversational style of the post, the topic-specificity, and the related tags. There isn't mention of agoraphobia in the body of the post at all; rather, it is implied through the tags as the cause of the user's anxiety. In this way the metatags are functioning not as censoring agents or warnings, but as additional pieces of information relevant to the post's content.

The nature of agoraphobia is such that triggering content is somewhat difficult to come by, as a reaction in a person suffering from this phobia is typically elicited through an experience rather than through words. However, the usage of metatags depends entirely on the perception of the individual. Thus, there are instances like the following example where the content of the post is topic-specific as in the previous instances, but the original poster has nevertheless chosen to utilize a warning tag, #agoraphobia tw.

(8) August 24th 2014

I went to carnival for all of two hours and I was fine while I was there but as soon as I got back I passed out then started having panic attacks and that was at 3.40 and the fainting and attacks still haven't stopped over two and a half hours later.....

I'm proud of myself for confronting my phobia of crowds but pissed off at my body for still reacting like this

also I feel gross

#panic tw **#agoraphobia tw** #anxiety tw #like come on two hours and forty minutes and counting

The user is clearly conscious of the potential discomfort of others and is eager to employ censor-enabling metatags for the sake of public social wellbeing. Whether or not this may seem excessive is irrelevant, as the choice for self-censorship relies solely on the mediated interaction between the original poster and the individual who comes into contact with their post.

5.3.3 #Arachnophobia tw

The familiar distinction between topic-specific and triggering content is yet again seen in the comparison between posts tagged #arachnophobia and #arachnophobia tw, where the distinction is potentially the most definitive out of all three metatag groups. The tag #arachnophobia is typically added to post which involve some discourse on the elements of the phobia, as illustrated by the following example:

(9) December 30th 2011

OOC Fun Fact:

//I absolutely HATE spiders... They scare the living shit out of me.

~;

#Arachnophobia #Personal #Random #Fact #Arachnids #Spiders

This is, again, a personal post recounting the emotional experience of the original poster in a form of one-to-many communication, as explicitly denoted by the phrase “Fun Fact”. The poster is expressing their hatred for spiders, which they have chosen to classify as arachnophobia. Whether spiders actually elicit a panic response from them is irrelevant in this instance, as the content neither discusses the phobia in depth, nor presents any sensitive content. The same can be said for the following example:

(10) October 26th 2013

I just saw a post with a spider picture and nononononono I don't care that it docile and makes gold thread or used as mosquito nets over baby cribs it freaks me out to no end.

#arachnophobia #arachnophobia is a bitch #god i hate spiders

The user is simply sharing an experience and an emotional response without going into explicit detail, which could potentially cause others discomfort. Quite plainly, it is a post about arachnophobia, and presents a stark contrast to the following post:

(11) March 15th 2015

[IMAGE]

I made friends with a baby spider today (that's my pinky finger in the second picture)

#arachnophobia tw #spiders

This post differs both in content structure and in tone from the two prior, as it presents two images of a spider along with a very arachno-positive caption. Quite appropriately, this post is tagged **#arachnophobia tw** as the images as well as the mention of “spider” pose a very real threat to those suffering from severe arachnophobia. This reflects the established convention where oftentimes the format of the metatag is indicative of the nature of the content, be it topic-specific or content-specific. Moreover, the nature of the content is a clear risk, and the original poster

again has enabled themselves to be censored for the sake of the comfort of others. The trigger warning acts as such and allows the content to be blocked on an individual basis.

5.4. The arachnophobia tag variants

Following the analysis of these three parent tags and their marked censoring variants, it would be beneficial to more closely examine the language of each tag as a whole. However, this is simply not feasible for every tag due to the sheer amount of labor involved in collecting the data manually. Therefore, the scope of this section was restricted to the four most common variants of one parent tag. The parent tag chosen for this analysis was #arachnophobia, due in part to its popularity and its manageable size. Furthermore, content wise, the arachnophobia tags contain a large amount of textual data, which is necessary in order to compile a sizeable corpus. The four variants marked with censoring elements are the same four which have been analyzed in the previous sections: #arachnophobia cw, #cw arachnophobia, #arachnophobia tw, and #tw arachnophobia. By further scrutinizing the language associated with these tags through the use of a comprehensive corpus the relationship between content and metatag will be more precisely defined as it functions in this study, hopefully providing a more distinct understanding of the nature of cooperative censorship on Tumblr and the nuances of its usage.

First, section 5.4.1. will provide a brief overview of the composition of the dataset as a whole. Section 5.4.2. will present a comparative analysis of word frequencies based on a corpus of all four tags. Section 5.4.3. will then present a more in depth examination of clusters and concordances associated with select keywords chosen based on the findings of the previous section. Finally, Section 5.4.4 will provide a brief chronological analysis of the content-tag relationship on the basis of concordance plots.

5.4.1. #Arachnophobia cw, #cw arachnophobia, #arachnophobia tw, and #tw arachnophobia

As shown in section 5.2.3., the development of the popularity of the two tags in the format #TAG-t/cw greatly outnumbers that of the other two (in the format #t/cw-TAG), as determined by a

systematic diachronic sampling of each tag. However, a comprehensive count of all the posts in each tag shows that #arachnophobia tw is the actually most used tag over its lifespan. While #arachnophobia cw has gained recent popularity comparable to #arachnophobia tw, the cw variant hasn't been in sustained use as long as the tw variant which first appeared in 2011, whereas #arachnophobia cw came into usage in 2012. Therefore the amount of textual data in the #arachnophobia tw tag is markedly greater than the amount of textual data in the other tag variants.

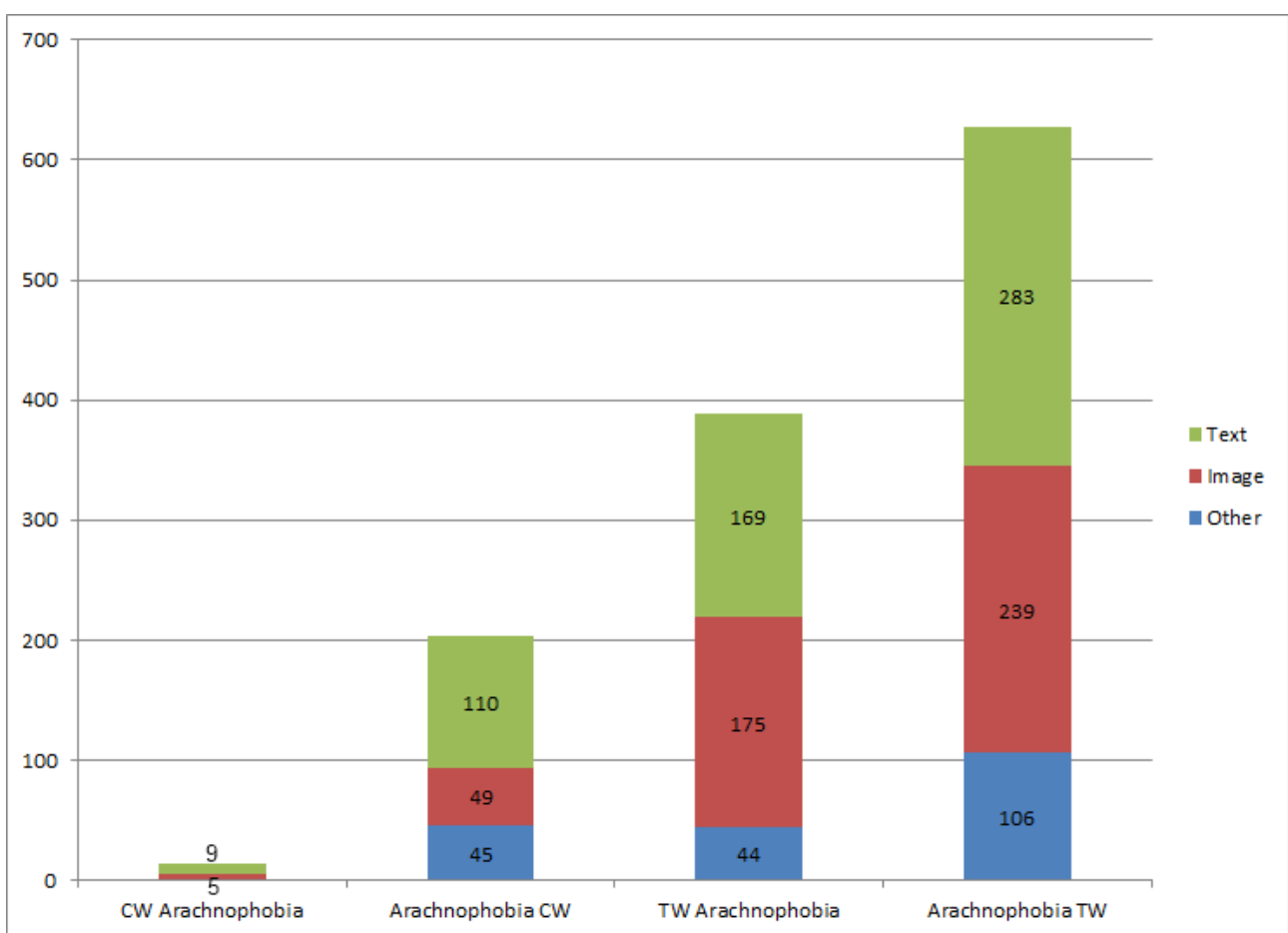


Figure 5.5. Frequency of post types in the four marked arachnophobia tags.

As shown in Figure 5.5, the metatags marked with tw contain more posts than the metatags marked with cw. This may be due in part to the popularity of the term “trigger warning” in the website’s social culture, whereas “content warning” conveys a more generic or clinical term. Since

the act of using these warnings is tied to the concept of psychological triggers in the site specific discourse, it is reasonable to expect that it has been in use for a longer period of time and therefore contains more posts.

In terms of composition, the posts in each tag can be categorized into three groups: text, image, and other. The text posts, which make up the focus of this analysis, contain solely text. The image posts are multimedia content, either photographs, .gif files, or videos, which will be excluded from analysis, as mass analysis of multimedia content is unfeasible and outside the scope of this study. The other category contains the rest: chat posts, messages between users, and hidden content which give the impression of a reasonable expectation of privacy. These will also be excluded because of their private or interaction-based nature, which ultimately lacks adequate context and verges on the ethically ambiguous. Posts in the 'other' category consistently make up the minority in each tag variant, and therefore their exclusion will not greatly affect the size of the dataset. These proportions are illustrated by Figure 5.6, which breaks down the composition of each tag by percentage. Text and image posts are relatively equal in number a majority of the tag variants, hinting at the nature of the content relevant to arachnophobia, which can generally be divided into textual accounts of encounters with spiders and pictures or videos of spiders.

Out of all the marked tag variants, only #tw arachnophobia contains a larger proportion of image posts as compared to text posts, if only by a slim margin (45% image posts, as opposed to 44% text posts). The rest of the tag variants contain more text posts than image by a distinct majority, as illustrated in Figure 5.6. #Arachnophobia tw is made up of 45% text posts and 38% image posts; #arachnophobia cw is composed of 54% text posts and 24% image posts; and #cw arachnophobia has the greatest proportion of text posts at 64%. However, as it has the smallest number of total posts (14) this distinction cannot be interpreted in a largely significant matter in the general analysis of the dataset as a whole, which is comprised of 1234 posts in total.

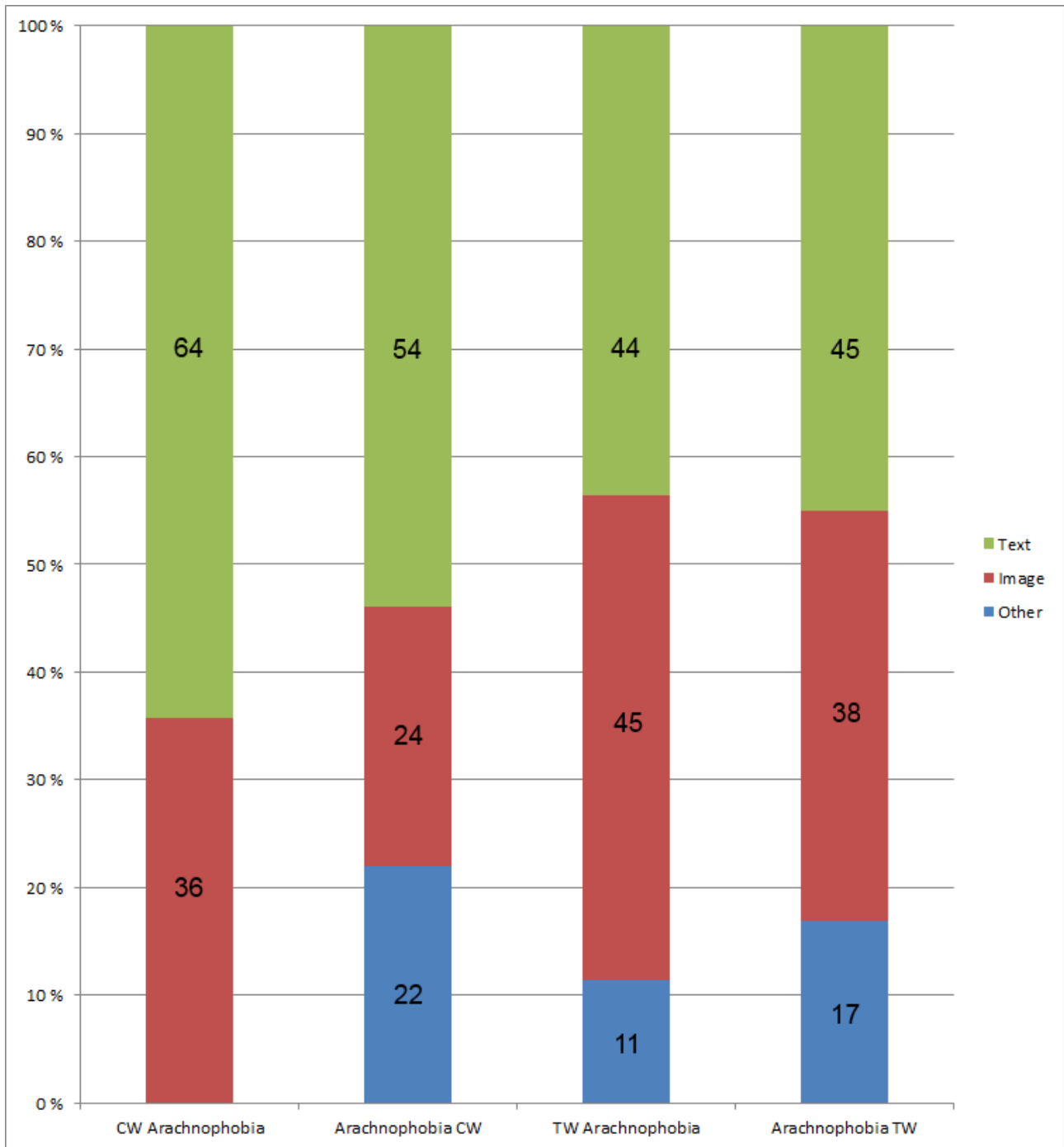


Figure 5.6. Post type make up of each tag variant by percentage.

Overall, despite their differences in size, the two metatags marked with tw contain similar proportions of text posts in comparison to other content. This is most likely due to their lexical similarities. Trigger warnings are used by a specific demographic within a specific social setting and therefore the content marked by these tags, no matter the order, will contain similarities. The

metatags marked with cw contain fewer posts and exhibit a greater variation of proportions. At this point it is not appropriate to make any further generalizations based on the dataset as it is divided. To further analyze the content of these tags it is necessary to turn to other quantitative means. In the following section, the content of all four tags will be compiled into a corpus which will undergo frequency analysis in order to determine the ways in which the censored content relates to the marked tags.

5.4.2. *Word frequency analysis*

In this section, general patterns in the lexicon of the censored text will be analyzed through the use of word frequency analysis. This will provide a number of hints as to the relationship between the agent of censure and the censored content which can then be analyzed within the context of its usage. By using software made for corpus analysis to count the word frequency of all arachnophobia tag variants, this data can then be compared to other standard corpora to differentiate distinct patterns. Again, the objective of this analysis is not to establish strict patterns of usage applicable to all trigger tags, but to describe the characteristics of this specific set of tags which may suggest something about the characteristics of trigger tag usage on Tumblr. This restriction is reflected by the relatively small size of the corpus at hand: the corpus is comprised only of the four arachnophobia trigger tag variants and contains 41,106 tokens collected from text-based posts. The corpus to be used as the basis for comparison in this study is the Corpus of Contemporary American English, which is comprised of 533,788,932 tokens spanning from 1990-2015, divided equally among spoken, fiction, popular magazines, newspapers, and academic texts (COCA, 2016). A list of the 100 most common words in the arachnophobia tags (arachnophobia corpus) can be found in Appendix A, and the normalized frequency list in Appendix B; a normalized frequency table of the 100 most common words in the English language from Corpus of Contemporary American English can be found in Appendix C.

For the purpose of this paper, the main focus of the frequency analysis has been limited to the top 100 most frequently used words in both the content and tags of each post. Because of the comparatively small size of the corpus, the focus must be restricted in order to avoid undue extrapolation considering the narrow scope of this study. Furthermore, both the main content of each post as well as its tags has been included, as the tags often include relevant information and commentary on the preceding post. However, the actual trigger tags themselves as well as the parent tag (#arachnophobia) have been excluded, as they exist as independent items, and in the case of frequency analysis provide no function within the text outside of organization. In order to faithfully trace the instances of the actual word [arachnophobia], the hashtags must be excluded. All numbers presented have been normalized per million so as to more readily form a comparison between the two corpora; however, the actual frequencies are also referenced when considering the metatag corpus by itself and can be found in Appendix A.

Before any comparisons can be made, the corpus collected in this study must first be discussed with particular attention paid to the content-specific vocabulary associated with arachnophobia and trigger tags. Table 5.2 displays the fifty words which appeared most frequently in the text posts catalogued by the marked tag variants of the parent tag #arachnophobia. These words make up the content offered for cooperative censorship, and therefore it is to be expected that some of the most common words relate to arachnophobia and spiders in some fashion. Indeed, the first content-specific term which appears is 'spider', which is the tenth most common word with a total of 649 occurrences. This is immediately followed by 'spiders'. It is interesting to note that both 'spider' and 'spiders' appear more frequently than 'arachnophobia', the actual keyword in the censoring tag which is the 85th most common term with only 66 occurrences. This suggests that spiders are discussed more than the phobia itself, hinting at a content-specific usage of the tags rather than a topic-specific usage. In other words, users are marking the presence of spiders for the benefit of those suffering from the phobia, rather than the discussion of the phobia itself.

Table 5.2. Fifty most frequent words in all marked arachnophobia tags (normalized per million).

Rank	Word	Frequency	Rank	Word	Frequency
1	i	47511	26	this	5717
2	and	31334	27	tw	5644
3	the	30677	28	with	5620
4	a	26128	29	out	5474
5	to	20946	30	for	5133
6	my	20265	31	not	4963
7	it	19827	32	they	4257
8	of	16859	33	at	4136
9	spider	15788	34	are	3941
10	spiders	15642	35	one	3868
11	in	13867	36	have	3844
12	was	9985	37	up	3844
13	that	9488	38	as	3819
14	s	8952	39	fucking	3649
15	so	8296	40	if	3503
16	on	8223	41	now	3454
17	is	8004	42	all	3430
18	t	7979	43	be	3308
19	me	7517	44	them	3260
20	but	7468	45	from	3187
21	like	7371	46	don	3163
22	just	7347	47	about	3017
23	you	6666	48	because	2992
24	m	6203	49	when	2968
25	there	255	50	can	2895

Also relevant is the term ‘tw’, which is the 27th most common term with 232 total occurrences in the corpus (5644 per million). In accordance with the previous sections’ analysis, it is to be expected that ‘tw’ appears more frequently than ‘cw’; however, in this dataset, all arachnophobia tw-tag variants have been removed. This suggests that the presence of these censoring tags does coincide with additional marked tags, such as #tw spider or #spider tw among other content relevant metatags. The same can be understood from the presence of ‘cw’ with 66 total

occurrences (1606 per million). The presence of these lexemes speaks to the usage of trigger tags, and suggests that multiple applicable triggers are commonly tagged concurrently.

Outside of the trigger tags, there are a number of words among the top 100 most frequent which stand out solely in the context of the corpus. However, analysis of these terms requires a point of comparison. In order to extricate the most significant patterns of usage, the frequency data from the collected corpus will be compared to a standardized corpus, in this case the Corpus of Contemporary American English. The Corpus of Contemporary American English (COCA) is the largest corpus of American English available, collected by Mark Davies at Brigham Young University (2010). The corpus is described as follows (2016):

[COCA is] the largest freely-available corpus of English, and the only large and balanced corpus of American English. COCA is probably the most widely-used corpus of English, and it is related to many other corpora of English that we have created, which offer unparalleled insight into variation in English [...] The corpus contains more than 520 million words of text (20 million words each year 1990–2015) and it is equally divided among spoken, fiction, popular magazines, newspapers, and academic texts.

In addition to the sheer scope of the corpus, since Tumblr is an American-based company with a majority base of American users influencing the usage of English, COCA was deemed the most appropriate choice for comparison. Table 5.3 shows the fifteen most common words in the COCA as well as their frequency.

Table 5.3. Fifteen most frequent words in the English language normalized per million (from Corpus of Contemporary American English).

Rank	Word	Frequency
1	the	41287
2	be	23503
3	and	20122
4	of	19378
5	a	19004
6	in	13107
7	to	11863
8	have	8063
9	to	7453
10	it	7255
11	I	7226
12	that	6428
13	for	6147
14	you	5772
15	he	5450

Immediately, one major difference between the two corpora becomes apparent: the rank of the first person singular pronoun *I*. *I* is ranked 11th most frequent in COCA, whereas *I* ranks as the number one most common word in the arachnophobia corpus. Given the personal nature of content published on microblogging platforms, it is not unexpected to find the first person singular pronoun near the top of the list, but the ten-place difference between the two corpora seems to suggest a more heightened focus on the speaker and their experiences in the arachnophobia corpus. It is reasonable to expect that the content of these trigger tags tends to deal with personal experiences in the context of some degree of trauma given the presence of a phobia, either in the speaker or their potential audience. Concerning the use of *I* in this context, Dunnack and Park (2009) suggest there is some degree of relation between pronoun usage and psychological

adjustment when dealing with trauma; they suggest this correlation is due to “the thought process that use of the pronoun I represents. This thought process is self-focus, thinking about oneself, and over time it likely aids in adjusting to trauma” (pp. 442–443). This aspect of pronoun usage in relating trauma is certainly plausible, but nevertheless only covers one singular aspect of the content of the trigger tags, and would be expected to appear in direct correlation with the word arachnophobia outside of the tags. Given the lesser frequency of the latter, it is unreasonable to account for pronoun usage solely as a result of the processing of micro-trauma related to the phobia. Possibly more prominent is the simple fact that microblogging platforms are used for personal expression, and a topic like arachnophobia tends to elicit personal responses like opinions and experiences (Thurlow et al., 2004, p. 14). Ultimately, the popularity of the first person singular pronoun is most likely the result of the convergence of these factors. Considering the usage of other pronouns, both corpora follow similar patterns, further highlighting the difference in the usage of the first person singular pronoun *I*. The third-person, singular neuter pronoun *it* is ranked seventh in the arachnophobia corpus and tenth in COCA. Regardless of content, this pronoun is relatively consistent in usage. In comparison, the first person singular possessive pronoun *my* has a greater difference in usage. *My* is ranked 44th in COCA and sixth in the arachnophobia corpus, mirroring the difference in *I* usage to a greater degree. This is also somewhat predictable considering the personal focus of the content.

Outside of pronouns, looking at the frequency of select nouns offers another perspective on the nature of the content of trigger tagged posts. These nouns were chosen based on the top 100 frequency list produced from the arachnophobia corpus. In comparison to the COCA frequency list, three nouns stood out as divergent from the norm: *spider(s)*, *room*, and *bed*. *Spider* and *spiders* ranked 10th and 11th, with 649 and 643 actual occurrences respectively (the lemma *spider* would be ranked second with 1292 total occurrences); whereas *spider* does not even fall within the top 5000 most frequent words in COCA. Naturally, this frequency can be explained by the focus of the content of the trigger tagged posts. Spider related content is tagged with arachnophobia related terms in consideration of sufferers of the phobia as well as for the purpose of topic-specific organization. However, in this case the presence of the other two words *room* and *bed* might not be as readily predictable as *spider*. *Room* is ranked 58th in the arachnophobia corpus and 228th in COCA, whereas *bed* is ranked 78th and 684th respectively. Nevertheless, considering

the settings where spiders dwell and are commonly encountered, it is reasonable to suspect that the presence of both *room* and *bed* hint at the popularity of reports of spider encounters within the trigger tagged posts. Analysis of each term’s collocates might prove more fruitful in cataloguing the actual popularity of these reports as a percentage of all posts, but the presence of these words suggests they are present in a substantial enough number to support the presence of a content specific tagging system.

Moreover, another notable word in the 100 most frequent list derived from the arachnophobia corpus is the verb *kill*. *Kill* stands out due in part to its connotation, but when compared to COCA, its prominence becomes more significant. *Kill* is at rank 91 in the arachnophobia corpus, while COCA has it ranked at 425. Again, the content offers valuable clues as to its prominence in the arachnophobia corpus; since there is a marked frequency of personal accounts of spider encounters, it is likely that a number of these are reporting either an attempt or a wish to kill a spider considering the implications of an arachnophobia trigger tag. This hints at hostile tone, but the instances of *kill* would also benefit from collocate analysis in order to better understand the context in which it occurs and the tonality to which it contributes. Similarly, the popularity of expletives in the top 100 list also hints at the emotive tone of the content of trigger tagged posts. The arachnophobia corpus has produced three expletives in the top 100: *fucking*, *fuck*, and *shit*.

Table 5.4. Comparative normalized frequencies of expletives in both corpora

Word	Corpus	
	Arachnophobia	COCA
fucking	3649	17
fuck	2360	18
shit	1411	27

Table 5.4 shows the respective ranks of the three most common expletives, and it is strikingly apparent that expletives are far more common in the arachnophobia corpus. As expletives tend to function as markers of tone as well as adjective intensifiers, their popularity suggests that the

majority of the content marked by trigger tags is highly emotive with a negative tone in the case of the arachnophobia corpus. Moreover, multiple studies have suggested that the function of expletives reaches beyond the typical limits of language. According to Bergen (2016), expletives are processed differently than other words, and have their own, “more emotionally connected” pathways in the brain which are preserved even when other parts of language processing are damaged. This emotional functionality is supported by Blakemore (2015), who argues that expletives have no true descriptive content and occur in conjunction with facial expressions, gestures, and tone of voice in order to “trigger procedures for the identification of emotional states”. This involvement in the processes for conveying tone and emotion support the emotional nature of the content marked by trigger tags.

5.4.3. Clusters and concordances

In order to better understand the context in which these terms occur, it is necessary to examine the most frequent clusters associated with each term. Moreover, certain terms would benefit from being analyzed through the use of concordances so as to achieve a more comprehensive overview of the conditions in which they are used; this is the main focus of this section. So as to better differentiate between patterns of topic-specific and content-specific trigger tagging, the clusters in which the word *arachnophobia* occurs will be looked at first in this section. Again, all these instances exclude trigger tags so as to clarify the difference between content and metatagging practices. Table 5.5 shows the most common 2-word arachnophobia clusters.

Table 5.5. Most common ‘arachnophobia’ clusters

Rank	Freq	Range	Cluster
1	6	1	have arachnophobia
2	2	1	an arachnophobia
3	2	1	my arachnophobia
4	2	1	severe arachnophobia

The most common cluster is ‘have arachnophobia’, which is to be expected as most personal posts related to the topic of arachnophobia are conveying personal experiences with the phobia. ‘My arachnophobia’ follows a similar pattern. Following these clusters which establish relationships to the phobia are adjectives which modify arachnophobia (e.g. severe, debilitating). These also relate to the personal experience of arachnophobia rather than spiders, thereby establishing a topic-specific usage within the trigger tags. Examples of usage within context are displayed in Figure 5.7.

wish hearts of stone had come with an [arachnophobia](#) warning because stumbling into a cav
i used to be. my mom had severe [arachnophobia](#) and i kind of learned to be afraid
ANY DAY #it's hell for people with [arachnophobia](#) 2015-04-04 Workin\x92 on that firs
over the research that has been done on [arachnophobia](#) in brief. I will hopefully finish it
my unfolded laundry and i dun really have [arachnophobia](#) but that was enough. and my fish rea
it\x92s a hat? Unless you have [arachnophobia](#), I highly recommend googling it. It\
come a long way since I developed debilitating [arachnophobia](#) as a child. I just squashed a spide
#spider cw 2015-02-13 i know i kind of have [Arachnophobia](#) But I have no idea it was that

Figure 5.7. ‘Arachnophobia’ concordances

These concordances all either relate experiences with arachnophobia or address individuals with arachnophobia, further supporting the topic-specific tagging system. However, these instances cannot be considered the majority, as they are far outnumbered by mentions of spiders, and, as addressed in Section 5.4.4., decline in popularity in favor for posts tagged on the basis of content (i.e. spiders). Table 5.6 displays the most common 2-word *spider* clusters, which, while directly related to arachnophobia, differ greatly in context and usage.

Table 5.6. Most common 2-word 'spider' clusters

Rank	Freq	Range	Cluster
1	175	1	a spider
2	67	1	the spider
3	23	1	huge spider
4	22	1	fucking spider
5	16	1	giant spider
6	14	1	ass spider
7	14	1	little spider
8	12	1	tw spider
9	9	1	tiny spider
10	7	1	big spider

Naturally the first two most frequent clusters contain articles, but all but one of the following clusters contain either an adjective or expletive functioning as an adjective (e.g. huge, fucking, giant, little). These serve to heighten the emotional impact of the word *spider*, as well as to convey a strong negative emotional tone. When the scope is expanded to 3-word clusters, these patterns remain, but are supplemented by verbs (e.g. be, kill, find, see).

Table 5.7. Most common 3-word 'spider' clusters

Rank	Freq	Range	Cluster
1	26	1	was a spider
2	17	1	a huge spider
3	14	1	s a spider
4	11	1	a giant spider
5	10	1	kill a spider
6	8	1	found a spider
7	8	1	with a spider
8	7	1	see a spider
9	6	1	a fucking spider
10	6	1	a massive spider
11	6	1	is a spider
12	5	1	and a spider
13	5	1	by a spider
14	5	1	huge ass spider
15	5	1	saw a spider

Table 5.7 displays the most frequently used 3-word *spider* clusters. In addition to negative emotive phrases (e.g. ‘kill a spider’, ‘a fucking spider’), there are experiences of spider encounters (e.g. ‘was a spider’, ‘found a spider’). These reports of spider encounters can be better understood when examining the concordances of the term *spider*, as seen in Figure 5.8.

australian feel when you find a huge dead spider in your house but then you realise it's
 skin and there's still a massive spider even bigger than the skin hiding somewhere
 rsonal 2015-12-21 i think there is a funnel web spider in the bathroom of my campsite. i also
 2015-11-21 [arachnophobia tw] there was just a spider on my laptop screen and it surprised me
 ers 2015-11-14 I just collided face first with a spider while I was trying to turn on a
 going to think about how long the huge spider I just found behind my shower curtain liner
 wtf 2015-11-06 OH GOD MONKEY WAS HUNTING A HUGE SPIDER BUT HE LOST IT AND I LOST IT

Figure 5.8. ‘Spider’ concordances

It is apparent that a majority of spider related posts are indeed reporting personal experiences with spiders. These are emotionally laden and tend to make use of adjectives which convey negative emotions such as fear, disgust, and stress. Content-wise, this is rather appropriate considering each of these posts has been tagged with an arachnophobia trigger warning. Out of all the verbs associated with these events, perhaps the most interesting to analyze within the context of tonal expressions is the verb *kill*. The most common clusters are presented in Table 5.9.

Table 5.9. Most common ‘kill’ clusters

Rank	Freq	Range	Cluster
1	22	1	kill it
2	10	1	kill a
3	10	1	kill a spider
4	6	1	kill it and
5	5	1	kill them
6	4	1	kill it with
7	4	1	kill me
8	4	1	kill the
9	2	1	kill and
10	2	1	kill spiders

As is to be expected, the verb *kill* appears most commonly in association with the words *it* and *spider*. However, there is a small but not insignificant instance of the cluster ‘kill me’. This is yet another marker of emotion, which functions as a way to convey extreme emotion in the absence of other resources like facial expression and gesture. Nevertheless, while clusters and concordances can offer a number of hints as to the overall characteristics of the content of trigger tagged posts, such as emotion and tone, there still remains an element which cannot be accessed through looking at clusters alone. In order to visualize the developments in the relationship between content and trigger tags, concordances must be examined chronologically.

5.4.4. Concordance plots

Analysis of frequency and collocates helps to highlight the focus of the trigger tagged posts, but a different approach is required to examine the actual relationship between the content and the usage of trigger tags. In order to see the development of usage patterns over time, it is beneficial to visualize the data through the use of a concordance plot. In this section, the data has been organized chronologically and then processed using AntConc software (Anthony, 2014) to produce the following concordance plots. The primary concern in using this approach was to map the occurrences of the word *arachnophobia* outside of the organizational tags so as to determine the relationship between content and tag. Figure 5.10 shows the concordance plot for the word *arachnophobia*, starting from 2012 and spanning to the end of 2015.



Figure 5.10. Concordance plot of ‘arachnophobia’ in trigger tagged posts (2012–2015)

Instances of the word *arachnophobia* are more closely concentrated in the beginning of the time line, and as the trigger tags become more popular, the usage of the word itself becomes sparser. This supports the claim that the usage of trigger tags has moved from topic-specific usage to a majority content-specific usage. In other words, the usage of trigger tags has moved from marking the presence of the discussion of arachnophobia itself to the presence of spiders. This pattern is somewhat mirrored in the usage of the word *phobia*, as seen in Figure 5.11.



Figure 5.11. Concordance plot of 'phobia' in trigger tagged posts (2012–2015)

The presence of the word *phobia*, while far less common than *arachnophobia*, nevertheless skews to the left, further supporting the movement from topic-specific tagging to content-specific tagging. The terms related to the phobia present a distinct pattern of development in trigger tag usage in the arachnophobia corpus, however, terms related to both topic and content remain relatively constant.



Figure 5.12. Concordance plot of 'spider' in trigger tagged posts (2012–2015)

As seen in Figure 5.12, the term *spider* is used regularly within the content of the tagged posts. This is to be expected, as spiders relate closely to both arachnophobia as a topic and arachnophobia-triggering content. A similar distribution can be found for the words *kill* and *fuck* (see Figures 5.13 and 5.14).



Figure 5.13. Concordance plot of ‘kill’ in trigger tagged posts (2012–2015)



Figure 5.14. Concordance plot of ‘fuck’ in trigger tagged posts (2012–2015)

Since *kill* in this case is more closely related to *spider* than to *arachnophobia*, it is to be expected that its usage mirrors the former rather than the latter. Its usage is informed by the arachnophobia trigger tags, in that both tend to carry a particular negative tone towards spiders. If one suffers from arachnophobia, it is reasonable to expect that one might express a wish to kill spiders; and conversely, if one kills a spider and writes about it on Tumblr, the content qualifies for an arachnophobia trigger tag according to established site-specific social protocol (Murray, 1988). This does not necessarily mean that all such reports will be tagged, but that a substantial number of trigger tagged posts include this content. Similarly, the negative emotive tones conveyed by the word *fuck* appear in a parallel distribution, suggesting the tone of the content remains relatively consistent.

Overall, the concordance plots produced from this data set appear to suggest a movement from a topic-specific tagging system to a content-specific tagging system over the lifespan of the arachnophobia trigger tags; however, the tone of the content has remained relatively consistent, as evidenced by a number of tonal markers. The use of metatagging conventions to enable censorship is assigned on a user-to-user basis, and executed on the basis of content rather than topic.

6 Conclusion

The objective of this paper was to present the development of a complex variety of censorship as a growing social discourse as popularized by social media, namely on microblogging sites such as Tumblr. The growing usage of trigger and content warnings as a classification system on personal posts reflects a new, more flexible type of cooperative censorship which is neither a response to an outside censor (Germano & Meier, 2013), nor purely self-directed (Cook & Heilmann, 2013). Rather, the censorship is situated at the intersection of public and private spheres, influenced by this very nature of social networking sites. In this form, language functions as an agent of both censorship and self-directed censorship. In other words, the act of censorship is born through a mediated interaction involving the consent and directive of two parties; the producer uses metatagging conventions to enable censorship, and the consumer initiates it.

This consensual censorship is a hallmark of the ever-evolving nature of one-to-many communication as seen on social media. As the masses begin to produce content, there arises a new need for regulation which is often times self-directed until it cements itself into the social conventions of the format. Indeed, the usage of trigger warnings mirrors the regulatory rating systems already in place for most consumable media, though it is markedly more flexible in nature. Furthermore, there is evidence that the nature of this censorship is actively evolving away from topic-specific censorship and towards content-specific censorship, which greatly broadens the flexibility of this system. Nevertheless, this willingness to allow oneself to be censored in the public sphere is indicative of a social shift, especially concerning topics such as mental health. Though no drastic conclusions can be drawn, it is safe to say that the awareness of the necessity of discretion is apparent.

Furthermore, the dynamic nature of these developments in censorship is far more complex than the narrow scope of this study can comprehensively illustrate. As there is virtually no prior research on this specific phenomenon, there is much left to be taken into consideration; this study can only hope to serve as an introductory exploration into the topic. A more expansive

investigation into the metatag-content relationship as well as the social conventions surrounding this phenomenon would be necessary to draw any further conclusions. As it stands, the development and popularization of trigger warnings as agents of censorship and site-specific social conventions highlights the dynamic social interplay that is censorship on social media.

REFERENCES

- About: Tumblr. (2015). Retrieved February 18, 2015, from <http://www.tumblr.com/about>
- agoraphobia. (2015). In OxfordDictionaries.com. Retrieved March 16, 2015, from <http://www.oxforddictionaries.com/definition/english/agoraphobia>
- Anthony, L. (2014). AntConc (Version 3.4.4) [Computer Software]. Tokyo, Japan: Waseda University. Available from <http://www.laurenceanthony.net/>
- Arsan, E. (2013). Killing me softly with his words: Censorship and self-censorship from the perspective of turkish journalists. *Turkish Studies*, 14(3), 447–462.
- Barlow, J. P. (1996). Declaration of the independence of cyberspace. Retrieved from <https://projects.eff.org/~barlow/Declaration-Final.html>
- Blakemore, D. (2015). Slurs and expletives: a case against a general account of expressive meaning. *Language Sciences*, 5222–35.
- Bergen, Benjamin K. What the F: What Swearing Reveals About Our Language, Our Brains, and Ourselves. Basic. Sept. 2016. 288p.
- Burkhalter, B. (1999). Reading race online. In M. Smith & P. Kollock (Eds.), *Communities in cyberspace* (pp. 60–75). London: Routledge.
- Censorship. (2015). In OxfordDictionaries.com. Retrieved March 12, 2015, from <http://www.oxforddictionaries.com/definition/english/censorship>
- Chang, Yi, Lei Tang, Yoshiyuki Inagaki, and Yan Liu. (2014). "What Is Tumblr: A Statistical Overview and Comparison." arXiv:1403.5206 [cs.SI].
- Cook, P., & Heilmann, C. (2013). Two types of self-censorship: Public and private. *Political Studies*, 61(1), 178–196.
- Davies, Mark. (2008-) *The Corpus of Contemporary American English (COCA): 520 million words, 1990-present*. Retrieved December 16, 2016 from <http://corpus.byu.edu/coca/>
- Davies, Mark (2010). "The Corpus of Contemporary American English as the First Reliable Monitor Corpus of English". *Literary and Linguistic Computing* 25 (4): 447–65.
- December, J. (1997). Notes on defining of computer-mediated communication. *Computer-Mediated Communication Magazine*, (3):1. Retrieved from <http://www.december.com/cmcmag/1997/jan/december.html>
- Dijck, José van. (2011). Tracing Twitter: The rise of a microblogging platform. *International Journal of Media & Cultural Politics*; p333–348.

- Dunnack, E. S., & Park, C. L. (2009). The Effect of an Expressive Writing Intervention on Pronouns: The Surprising Case of I. *Journal Of Loss & Trauma*, 14(6), 436–446.
- “ESRB Ratings Guide”. (2015). In ESRB.org. Retrieved March 15, 2015, from http://www.esrb.org/ratings/ratings_guide.jsp
- Filak, V. F., Reinardy, R. S., & Maksl, A. (2009). Expanding and validating applications of the willingness to self-censor scale: Self-censorship and media advisers' comfort level with controversial topics. *Journalism and Mass Communication Quarterly*, 86(2), 368–382.
- "Film Ratings". (2015). In MPAA.org. Retrieved March 15, 2015, from <http://www.mpa.org/film-ratings/>
- Germano, F., & Meier, M. (2013). Concentration and self-censorship in commercial media. *Journal of Public Economics*, 97(1), 117–130.
- Goggin, G. (2000). Pay per browse? The Web’s commercial futures. In D. Gauntlett (Ed.), *Web.studies: Rewiring media studies for the digital age* (pp. 103–112). London: Arnold.
- Herring, S. (Ed). (1996). *Computer-mediated communication: Linguistic, social, and cross-cultural perspectives*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Herring, S. (2002). “Computer-mediated communication on the internet”. *Annual Review of Information Science and Technology*. Volume 36, Issue 1, (pp. 109–168).
- Johnson, P. L., Federici, L. M., & Shekhar, A. (2014). Etiology, triggers and neurochemical circuits associated with unexpected, expected, and laboratory-induced panic attacks. *Neuroscience and Biobehavioral Reviews*, 46(P3), 429–454.
- Lee, Carmen K. M. (2011). “Micro-blogging and status updates on Facebook: Texts and practices”. In: Crispin Thurlow and Kristine Mroczek (eds.), *Digital Discourse: Language in the New Media*, 110–128. Oxford/New York: Oxford University Press.
- “Missing E”. (2015). In Missing-E.com. Retrieved March 16, 2015, from <http://missing-e.com/>
- Murray, D. E. (1988). The context of oral and written language: A framework for mode and medium switching. *Language in Society*, 17, 351–373.
- Myers, Greg. (2010). *The Discourse of Blogs and Wikis*. London: Continuum.
- “Overview”. Corpus of Contemporary American English. Retrieved May 30, 2016, from <http://corpus.byu.edu/coca/>
- Phillips, Peter, and Ivan Harslof. (1997). “Censorship within Modern Democratic Societies.” In: *Censored: The News That Didn’t Make the News*, edited by Peter Phillips and Project Censored. New York: Seven Stories Press.

- Self-censorship. (2015). In OxfordDictionaries.com. Retrieved March 12, 2015, from <http://www.oxforddictionaries.com/definition/english/self-censorship>
- Shulman, I.D., Cox, B.J., Swinson, R.P., Kuch, K., Reichman, J.T. (1994). Precipitating events, locations and reactions associated with initial unexpected panic attacks. *Behav. Res. Ther.* 32, 17–20.
- Thurlow, C., Lengel, L. & Tomic, A. (2004). *Computer Mediated Communication: Social Interaction and the Internet*. Thousand Oaks, CA: Sage.
- Virtanen, T., Stein, D., & Herring, S. C. (2013). *Pragmatics of Computer-mediated Communication*. Berlin: De Gruyter Mouton.
- Wakeford, N. (2000). New media, new methodologies: Studying the Web. In D. Gauntlett (Ed.), *Web.studies: Rewiring media studies for the digital age* (pp. 31–41). London: Arnold.
- “Word frequency data”. Corpus of Contemporary American English. Retrieved March 10, 2016, from <http://www.wordfrequency.info/top5000.asp>
- “XKit”. (2015). In XKit.info. Retrieved March 16, 2015, from <http://xkit.info/>

APPENDICES

Appendix A

Word frequency: 100 most common words in the marked arachnophobia tags.

Rank	Word	Frequency	Rank	Word	Frequency	Rank	Word	Frequency	Rank	Word	Frequency
1	i	1953	26	this	235	51	do	119	76	bed	73
2	and	1288	27	tw	232	52	am	113	77	by	73
3	the	1261	28	with	231	53	he	112	78	thing	73
4	a	1074	29	out	225	54	had	110	79	im	71
5	to	861	30	for	211	55	or	110	80	go	69
6	my	833	31	not	204	56	room	105	81	would	69
7	it	815	32	they	175	57	what	105	82	huge	68
8	of	693	33	at	170	58	your	98	83	arachnophobia	66
9	spider	649	34	are	162	59	fuck	97	84	cw	66
10	spiders	643	35	one	159	60	get	95	85	its	66
11	in	570	36	have	158	61	she	92	86	got	65
12	was	411	37	up	158	62	into	89	87	we	65
13	that	390	38	as	157	63	no	88	88	been	64
14	s	368	39	fucking	150	64	really	86	89	want	63
15	so	341	40	if	144	65	then	86	90	going	62
16	on	338	41	now	142	66	back	80	91	kill	62
17	is	329	42	all	141	67	think	80	92	more	61
18	t	328	43	be	136	68	little	79	93	some	61
19	me	309	44	them	134	69	off	76	94	has	60
20	but	307	45	from	131	70	over	76	95	him	60
21	like	303	46	don	130	71	an	75	96	personal	60
22	just	302	47	about	124	72	even	75	97	why	60
23	you	274	48	because	123	73	know	75	98	shit	58
24	m	255	49	when	122	74	re	75	99	ve	58
25	there	255	50	can	119	75	her	74	100	where	58

Appendix B

Normalized word frequency: 100 most common words in the marked arachnophobia tags.

Rank	Word	Frequency	Rank	Word	Frequency	Rank	Word	Frequency	Rank	Word	Frequency
1	i	47511	26	this	5717	51	do	2895	76	bed	1776
2	and	31334	27	tw	5644	52	am	2749	77	by	1776
3	the	30677	28	with	5620	53	he	2725	78	thing	1776
4	a	26128	29	out	5474	54	had	2676	79	im	1727
5	to	20946	30	for	5133	55	or	2676	80	go	1679
6	my	20265	31	not	4963	56	room	2554	81	would	1679
7	it	19827	32	they	4257	57	what	2554	82	huge	1654
8	of	16859	33	at	4136	58	your	2384	83	arachnophobia	1606
9	spider	15788	34	are	3941	59	fuck	2360	84	cw	1606
10	spiders	15642	35	one	3868	60	get	2311	85	its	1606
11	in	13867	36	have	3844	61	she	2238	86	got	1581
12	was	9985	37	up	3844	62	into	2165	87	we	1581
13	that	9488	38	as	3819	63	no	2140	88	been	1557
14	s	8952	39	fucking	3649	64	really	2092	89	want	1533
15	so	8296	40	if	3503	65	then	2092	90	going	1508
16	on	8223	41	now	3454	66	back	1946	91	kill	1508
17	is	8004	42	all	3430	67	think	1946	92	more	1484
18	t	7979	43	be	3308	68	little	1922	93	some	1484
19	me	7517	44	them	3260	69	off	1849	94	has	1460
20	but	7468	45	from	3187	70	over	1849	95	him	1460
21	like	7371	46	don	3163	71	an	1826	96	personal	1460
22	just	7347	47	about	3017	72	even	1826	97	why	1460
23	you	6666	48	because	2992	73	know	1826	98	shit	1411
24	m	6203	49	when	2968	74	re	1826	99	ve	1411
25	there	255	50	can	2895	75	her	1800	100	where	1411

Appendix C

Normalized word frequency: 100 most common words (from the Corpus of Contemporary American English).

Rank	Word	Frequency	Rank	Word	Frequency	Rank	Word	Frequency	Rank	Word	Frequency
1	the	41287	26	from	3065	51	one	1449	76	how	1010
2	be	23503	27	that	3208	52	time	1470	77	then	1019
3	and	20122	28	not	3070	53	there	1448	78	its	1011
4	of	19378	29	n't	3033	54	year	1441	79	our	984
5	a	19004	30	by	2792	55	so	1433	80	two	970
6	in	13107	31	she	2782	56	think	1417	81	more	965
7	to	11863	32	or	2584	57	when	1329	82	these	963
8	have	8063	33	as	2430	58	which	1295	83	want	957
9	to	7453	34	what	2213	59	them	1285	84	way	921
10	it	7255	35	go	2156	60	some	1271	85	look	881
11	I	7226	36	their	2029	61	me	1271	86	first	870
12	that	6428	37	can	1916	62	people	1270	87	also	868
13	for	6147	38	who	1908	63	take	1270	88	new	822
14	you	5772	39	get	1860	64	out	1270	89	because	817
15	he	5450	40	if	1816	65	into	1263	90	day	811
16	with	5026	41	would	1749	66	just	1252	91	more	788
17	on	4821	42	her	1734	67	see	1247	92	use	787
18	do	4656	43	all	1723	68	him	1243	93	no	772
19	say	3588	44	my	1672	69	your	1236	94	man	771
20	this	3532	45	make	1671	70	come	1177	95	find	768
21	they	3495	46	about	1638	71	could	1158	96	here	754
22	at	3411	47	know	1606	72	now	1135	97	thing	750
23	but	3375	48	will	1553	73	than	1086	98	give	740
24	we	3328	49	as	1545	74	like	1066	99	many	721
25	his	3311	50	up	1490	75	other	1026	100	well	720