#TWITTERDISCOURSEMARKERS:

A CORPORA BASED STUDY OF THE PRAGMATIC

FUNCTIONS OF HASHTAGS

by

DARCEY BROWNING

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON

August 2017

I will always be eternally grateful for my amazing husband Andy, who has made our seemingly unattainable goals possible. Who would have thought two nerdy kids from South Louisiana would make their dreams happen (despite the fact that it took almost 15 years to do it)? His persistence at the times I needed it the most and his ability to listen and give me advice on a subject that he had no formal training in was essential for the process, but his unconditional love and support was really what made this dissertation possible. I'm so grateful for the last 15 plus years together, and I look forward to further pursuing our goals while juggling all the joys of parenthood together-- we got this!

Finally, I am grateful to my son August, who inspired me to work hard during his naps, and play hard during our off time. I have to admit that if he would not have slept so well in the first few months of his life, many of these projects would be at a much different place. He provided a great deal of comedic relief to Andy and me while we were finishing our dissertations, and our son made us both more organized in our plans while teaching us to really relax and enjoy when we weren't doing work. Augy is such a sweet, clever, awesome kid, and I'll try my hardest not to geek out too much as he learns new linguistic innovations as he grows older or embarrass him with some intentional (and some unintentional) misuse of slang.

Abstract

#TWITTERDISCOURSEMARKERS:

A CORPORA BASED STUDY OF THE PRAGMATIC

FUNCTIONS OF HASHTAGS

Darcey Browning, PhD

The University of Texas at Arlington, 2017

Supervising Professor: Laurel Stvan

In this dissertation, I posit that hashtags can function as discourse markers, where space constraints of 140 characters on Twitter complicate their realization. Through the progression of research questions that shape each chapter, this dissertation analyzes how hashtags assist with felicitous communication to the intended audience of a tweet via four distinct corpora.

The first function of discourse markers investigated in this dissertation involves delays; in emotional narratives, we would expect to find discourse markers acting as delays. Starting with a corpus of survivor interviews, I investigate which traditional discourse markers appear frequently when survivors talk about violence and which content environments seem to prompt traditional, spoken discourse markers in these emotional narratives. Following up on the spoken discourse markers, I found similar patterns in survivor tweets from the #whyIstayed campaign. Using a corpus of 443 survivor tweets from across 73 days, I show how the initial placement of some hashtags before sensitive information can be a delay device. I also explore what linguistic content is more likely to cause this tweet-initial placement through statistical evidence, showing similarities between the hashtags and traditional discourse markers. Through answering the research questions for Chapter 2 and 3, I show how hashtags are act as delays in a similar fashion to some spoken DMs.

The second discourse marker function investigated in this dissertation is clarification. Starting with a general corpus of 1791 tweets from 2012, I explore the different functions of hashtags, finding that over 75% of hashtags in this corpus are involved with clarifying tweet content. Using the heuristics presented in Clark and Marshall (1981), I investigate how some hashtags can clarify the meaning of the tweet, through what I call tag reframing hashtags in Chapter 4. The third function is to reveal speaker's attitude, and in Chapter 5, I look at how two very different attitudes are expressed through the use of #NastyWoman. Using tag reframing hashtags in conjunction with a learning algorithm, I was able to analyze a second corpus of over 55,000 tweets from the 2016 presidential election to show the reclamation of a pejorative in Chapter 5. I show through the density of pejorative and reclaimed uses of #NastyWoman the

power struggle in the use of this term at the associated attitudes in the 18 days before the election.

Using a progression of corpus collection methods and data management, this dissertation shows that some hashtags are discourse markers, indicating that hashtags in computer mediated language pattern in form and function as discourse markers in other language genres do.

Table of Contents

List of Illustrations

List of Tables

CHAPTER 1

INTRODUCTION

1.1 Overview

Tweets, or postings on Twitter, are informal, conversational messages constrained to 140

characters or less. Hashtags, the combination of the # symbol and a series of words or

characters, are used for organization of tweets by searchable topics or events (Small 2011);

however, while they continue to be used as a way to search topics quickly, some hashtags

function to help create felicitous communication to the intended audience of the tweet in a

highly constrained space. This dissertation will show that hashtags in computer mediated

language pattern in form and function as other discourse markers in other languages. While

much has been noted about the nature of previously recognized discourse markers, discourse

particles, and pragmatic particles, there is a gap in the literature with respect to how hashtags

fit into this paradigm.

Both Chapters 2 and 3 use corpora of survivor discourse where DMs should appear because of

the emotional, narrative nature. In Chapter 2 I look at frequent spoken delaying discourse

markers that appear when people talk about violence. Following this, I present evidence of

hashtags acting as a delay device when sharing personal information on Twitter (in Chapter 3).

Furthermore, Chapter 3 is supported by the DM patterns in interview disclosures in Chapter 2,

since both chapters deal with survivor discourse where DMs should and do appear because of

the emotional, narrative nature.

The second discourse marker-like function of hashtags involves clarification of information: I show that hashtags clarify the message of a tweet by tag reframing (in Chapter 3). Chapter 5 provides an analysis on a specific hashtag using the tag reframing in the previous chapter. Using a progression of corpus collection methods and data management, this dissertation will show that some hashtags are discourse markers, and furthermore, hashtags in computer mediated language pattern in form and function as other discourse markers in other languages.

1.2 Literature on Computer Mediated Communication

Since computer mediated communication (CMC) shares qualities of both written and verbal communication (Crystal 2011), one could expect to see similar pragmatic devices being used for felicitous discourse. When an influx of people were introduced to computer mediated language, many relied heavily on communicating as one would in verbal and written correspondence- for instance, some extra information was given through parentheses or longer paragraphs of explanation; however, CMC developed in such a way that messages became shorter, and users relied more on abbreviations and acronyms. When parenthetical information was not used and parts of messages were omitted for brevity, interlocutors often engaged in infelicitous discourse. One can see that communication can sometimes be unsuccessful for users of computer-mediated discourse; so what devices developed within digital discourse?

1.2.1 Previous studies in pragmatic devices on CMC

Before hashtags, interlocutors of CMC developed other discourse markers. People relay emotion with extra letters in words (saying something like, "Yayyyy!" to express excitement") (Baldwin and Chai 2011). The repetition of letters are known to "add character and richness to the sentences" (Kalman and Gergle 2014). Others used emoticons to give extra information to the "hearers" or readers of their messages (Dresner and Herring 2012). Punctuation, like periods, can be used to assist with felicitous communication via digital devices; Gunraj et al. (2016) shows that periods alter the interpretation of a text message. Devices like punctuation, repetition of letters, and emoticons (all of which fall into a category sometimes called CMC cues) help facilitate less miscommunication; Vandergriff (2013) found that their participants use CMC cues "in combination/interaction with other emotive devices to convey socio-emotional information, including the sender's stance on the message or the sender's position vis-à-vis co-participants". Hashtags similarly assist with the interpretation of a tweet, but in a way much like that of devices already present in face-to-face communication.

Also, since turn-taking is often not as immediate as it is with verbal communication, people could interpret things incorrectly without having a clarification until the next time the speaker gets back online. Sanna-Kaisa Tanskanen and Johanna Karhukorpi (2008) studied CMC in emails, and they found that repair, a pragmatic way to clarify the meaning of a definite noun phrase, is not only used in verbal and written communication, but specifically, interlocutors in computer mediated discourse also use a form of repair known as Concise Repair during the same turn as the original message, but at the end of the message. Hashtags, which come at the end of a message on Twitter, are similar not only in the placement, but they can also act as a repair,

and thus can be categorized using Clark and Marshall's co-presence heuristics (1981), as shown in Chapter 4.

1.2.2 Corpus Collection on Twitter

Scholars, while looking at Twitter for other patterns, have written a great deal about how they gathered their data. Others have described the composition and creation of Twitter corpora (Petrovic et al. 2010; McCreadie et al. 2012; McMinn et al. 2013; Tatman 2015), and some of these corpus collection methods were more helpful than others for this dissertation. Chapter 4 uses the 2012 General corpus, downloaded from Illocution, Inc (2014), which was a website that collected tweets on Twitter and distributes corpora for other researchers. However, Chapter 3 and 5 both involved collecting the data for a specific purpose. To collect over 55,000 tweets over the span of 16 days before the 2016 presidential election, I used the R programming language script from Tatman (2015) regarding how to utilize the Twitter API in order to collect tweets. Using and adapting some of the corpus collection methods from previous literature, I was able to create four different corpora for this dissertation.

All of the corpora that analyze hashtags in this dissertation are from Twitter since this is the site from which hashtags originated. Furthermore, there are a number of reasons for restricting the study of pragmatic use of hashtags to the way hashtags are used on Twitter. Unlike hashtags that are used on Facebook, the usage of hashtags on Twitter is constantly evolving because of the frequency of use as a pragmatic device. Even though hashtags are used often on Instagram, which is a social networking application which uses images as the means of microblogging), people use hashtags somewhat differently because of the interaction of hashtags,

pictures, and captions. The other ways in which researchers look at hashtags are discussed in section 1.2.3.

1.2.3 Literature on Hashtags

Scholarly work on hashtags is more common than one might expect. There are a number of articles that look at hashtags for studies in other fields. For instance, Horeck (2014) looks at the use of #AskThicke and similar hashtags in discussions of feminism and rape culture. Epidemiologists, such as Chunara et al. (2012), find correlations with amount of discussions of outbreaks in tweets, including those that involved #cholera, and the amount of officially diagnosed people. Romero et al. (2011) evaluates the spread of ideas by considering the categories of content discussed using hashtags. This dissertation will be taking a much different approach to studying hashtags.

Many studies of hashtags actually look at Twitter data through the schema of political discourse. Small (2011) discusses the use of #cdnpoli in political discourse. She discusses how this particular hashtag on Twitter does "not provide a forum for political discussion" like one might find in political blogs, but instead, this hashtag is used for "political expression". Some use hashtags to get others involved in a social movement; for instance, Bruns et al. (2013) discuss the role of social media, including the use of #egypt and #libya, in the movement surrounding the 2011 Arab Spring while analyzing the patterns that emerged in the millions of tweets on this topic. Davis (2013) highlights the changes to elections due to the popularity of political discussions on Twitter, especially through hashtags. While my work is not directly related to politics (though chapter 3 analyzes tweets covering election debates), the functions

proposed will hopefully accommodate the pragmatic uses of these political hashtags since such hashtags are still acting as discourse markers.

Some looked at hashtag activism in very different ways. Freelon et al. (2016) talks about hashtags associated with the Black Lives Matter (BLM) movement, and found "unlike with live-tweeted TV shows or natural disasters where large audiences all participate at the same time, it takes longer for BLM activists to share information, gather evidence, and build support for action." However, Yang 2016 uses the hashtag #Blacklivesmatter to discuss the narrative agency in hashtag activism, with respect to structure. He posits that the syntactic structure of the hashtag is not motivating the narrative, but rather the participation of others. Furthermore, hashtag activism often features a narrative structure with a beginning, climax, and conclusion (Yang 2016; Clark 2016). Clark (2016) looks at the hashtag #WhyIStayed- which is discussed in Chapter 4 in this dissertation- through the schema of hashtag feminism. Weathers et al. (2016) gives descriptions of themes found in the #WhyIStayed campaign, and I will use these as a method to study the reasons for initial hashtag placement in a survivor tweet corpus. This dissertation, especially chapter 3, will be using these elements of the hashtag form to show how people use an initial hashtag placement as a delay device in survivor stories.

Not all research on hashtags takes a descriptive approach: Potts et al. (2011) calls for people to use a more unified hashtag during events to make it easier to search for specific things, especially during natural disasters, when many look to Twitter for news information. While many conferences and events promote the use of an official hashtag, standardizing all hashtags is rather unlikely since hashtags are generated by the users (Bruns and Burgess 2011). Because

people use hashtags in creative ways, however, we can study their pragmatic contributions to tweets.

Some research takes a broader look at the overall linguistic use of hashtags, while other research looks at particular hashtags for linguistic study. Zappavigna (2012) takes a qualitative look at the grammatical functions of hashtags in a larger work on Twitter discourse. Through discourse analysis, Page (2012) analyzes the use of hashtags as a means of gaining a larger audience. Sharma (2013) shows how users employ blacktags, which are "racialized hashtags - for example, #ifsantawasblack or #onlyintheghetto"- to signal race and race related issues. The use of these blacktags creates a sense of unity in a community, and Chapter 2 will be addressing this issue since a subset of certain hashtags can signal community membership. Caleffi (2015) similarly notes that hashtags refer to the topic and create "communities of people interested in that topic" while calling for greater linguistic research into the uses of hashtags.

Some sources even look at the pragmatic use of hashtags. For instance, Scott (2015) analyzes hashtags through the schema of relevance theory, and found that "they have been appropriated by users to perform other roles in the communicative process." This dissertation situates the analysis of hashtags into the framework of discourse markers, which fills a gap in the literature since there is no overarching analysis for the pragmatic functions of DM-like hashtags.

1.3 Discourse Marker Literature

Much has been noted about the nature of discourse markers, and these theories about discourse markers, discourse particles, and pragmatic particles can be helpful when determining how hashtags fit into the framework of discourse markers. I posit that hashtags show many of the same characteristics of discourse markers, including features like optionality, orality, and multicategorality. Hashtags are similar to some DMs, but differ in some notable ways. Many DMs have more than one function depending on the context, and hashtags as DMs also have many functions, some of which parallel some functions of some DMs.

Definitions or lists of properties and characteristics of DMs often vary by scholar, and rigorous definitions are difficult because of the variety of functions and form of different DMs. Stvan (2006) describes discourse markers as, "often defined as a set of words that has uses that are neither clearly meaning-bearing, nor purely grammatical in function." Jucker and Ziv (1998) call this a "fuzzy concept" because of the variation in what gets labeled discourse markers. Discourse markers (DMs) include a variety of words and phrases, like *like*, *oh*, *well*, *um*, and *y'know*, which are often utilized in spoken language; however, many native speakers fail to see the contribution of DMs to an utterance because these words and phrases are optional (Schourup 1999) and stigmatized (Brinton 1996). In fact, DMs are common in conversation, but often overlooked, ignored, or edited out. In section 1.3.1, I will define DMs, with a set of criteria that will structure our look at DMs and hashtags.

1.3.1 Definition of Discourse Marker

Discourse markers are types of extra sentential words that are used to organize within a discourse, while revealing context connections and speaker attitudes. For this dissertation, I will use the definition of DMs provided by Schiffrin (1987) where these discourse devices are "a functional class of verbal (and non-verbal) devices which provide contextual coordinates for ongoing talk." Furthermore, my current definition of DMs have the following criteria: DMs must (i) be positioned outside the clause structure; AND one of the following: (ii) work to signal the coherence relation or position of the utterance to some other part of the discourse AND/OR (iii) show the speaker's attitude toward an utterance, AND/OR (iv) signal a lower register (i.e. informality).

For criterion (i), DMs are positioned outside the clause structure, and an example of this is shown with *well* in (1) and *then* in (2).

       (1) **Well** [I   grew up uh out   in the suburbs] [...] (Schiffrin 1987: 106)
         DM   S   VP   PP

       (2) [other people   must have picked them   up]   then (Haselow 2011)
         S   VP   DM

In both (1) and (2) the DMs *well* and *then* appear outside of the main clauses. Criterion (i) is similar to syntactic optionality, where the DM is detachable and independent of the syntax of the sentence (Schiffrin 1987). In contrast, an example of *well* not acting in this way is in (3) where *well* is not a DM:

(3) [He   did   **well**   on the exam]

          S   VP   PP

The adverb *well* is not outside of the clause, and is instead an integral part of the verb phrase of this utterance. Instead, *well* is inside the VP, as shown in the gloss.

According to criterion (ii), DMs work to signal the coherence relation or position of the utterance to some other part of the discourse. An example of a traditional discourse marker position of an utterance is in (4) where *and* is a connector of events showing a relationship of temporal sequencing:

> (4) I uh go on trips with 'em [...] **and** I bring 'em here, [...] **and** we have supper, or dinner here, [...] I don't see any problem (Schiffrin 1987)

In (5), *so* is showing sequence (Müller 2005), and signaling position of the utterance to some other part of the discourse.

> (5) [...] restaurant he's an artist. **so** he comes over and starts talking to them. (Müller 2005)

In contrast, an example of *and* not acting as a connector of discourse or positioning any utterance part in relation to a larger context is in (6), where *and* is just joining noun phrases within a clause.

> (6)   [He   ate peas   **and** carrots]
>   [S   VP[V   NP   conj NP]]

For criterion (ii), this looks at the connective ability of DMs, where they situate an utterance to the larger context.

Criterion (iii) illustrates the speaker's attitude toward an utterance, and we see this flavoring of a statement of speaker attitude in (7).

(7a) Oh, Karen is moving back home.

(7b) Well, Karen is moving back home.

In (7a), the speaker is showing more excitement about the event than does the speaker in (7b) who may be more reluctant about sharing this news. The semantic meaning in both (7a) and (7b) both consist of Karen moving back home. The discourse markers flavor the meaning of these utterances, where (7a) shows that the speaker is happy about this information, while (7b) is uttered by someone who is not happy with the news. By flavoring the meaning and revealing the speaker's attitude, these two discourse markers change the pragmatic meaning since *oh* is a marker of information management (Schiffrin 1987) while *well* is a marker of cause and effect (Schiffrin 1987) and can precede information that one might be reluctant to share (Jucker 1993).

Criterion (iv) signals informality, and an example of a DM being used in an informal way appears in (8) where the contraction version of *you know* appears twice in the utterance.

(8) I believe ... that...**y'know** it's fate

So eh **y'know** it just seems that that's how things work (Schiffrin 1987)

The informal tone signaled by this DM is further supported by the contracted version of *you know* and the other contractions used in this utterance. Some DMs do not fit this criterion (iv), like *consequently* and *indeed* which are more formal, as shown in (9) and (10).

(9) Sue won't eat. **Consequently**, she will lose weight. (Fraser 1999: 943)

(10) The United States was at peace with that nation and, at the solicitation of Japan, was still in conversation with its Government and its Emperor looking toward the maintenance of peace in the Pacific. **Indeed**, one hour after Japanese air squadrons had commenced bombing in Oahu, the Japanese Ambassador to the United States and his colleague delivered to the Secretary of State a formal reply to a recent American message. (Han 2011)

In fact, *indeed* in (10) was used in a speech declaring war on Japan after Pearl Harbor, a time where formal speech would be needed by the President. Not all DMs signal informality, but DMs like *y'know* and *innit* do signal a lower register.

In the next two sections, I will use this definition provided above to show how hashtags fit the criteria of DMs, while also relating it to other studies in DMs. In section 1.3.2, relevant literature on characteristics of discourse markers is discussed. In section 1.3.3, I discuss the functions of DMs that are relevant to hashtags.

1.3.2 Features of Discourse Markers

Schiffrin (1987) addresses some of the properties of discourse markers, by calling them multifunctional, not-obligatory, and syntactically diverse. In Schourup (1999), he lists the characteristics of DMs as: connectivity, optionality, non-truth-conditionality, weak clause association, initiality, orality, and multi-categorality. Müller (2005) also includes properties like

word class, phonological features, and lack of semantic content, in conjunction with the characteristics discussed in Schiffrin (1987) and Schourup (1999). Combining all these frameworks, I address some of the more pertinent characteristics of DMs and how hashtags fit into the schema of DMs in sections 1.3.2.1-1.3.2.5.

1.3.2.1 Hashtags and Outside Clause Structure

The first criterion for DMs outlined in this dissertation is that the DM must be positioned outside the clause structure. Hashtags appear to be quite similar in distribution to other discourse markers. According to Fraser (1988), DMs usually appear utterance initial, sometimes utterance medial, and rarely in the utterance final position. An example of an clause initial discourse marker is *well*, as shown below:

(11) **Well** I grew up uh out in the suburbs [...] (Schiffrin 1987: 106)

An example of an utterance medial discourse marker is *like,* as shown in (12). One utterance final discourse marker is *then* in (13).

(12) and they go back and forth **like** two or three times (Müller 2005: 210)

(13) other people must have picked them up **then** (Haselow 2011)

Although *like* in (12) appears in the clause, it can be moved around without changing the meaning of the utterance, seen in (12a) and (12b) below:

(12a) and **like** they go back and forth two or three times

(12b) and they **like** go back and forth two or three times

Because of this, *like* is still considered extra sentential.

In contrast, if we look again at example (1) and (3), the DM can be removed from it, as shown in (1b) whereas *well* cannot be removed without losing the semantic meaning of sentence.

(1a) Well [I   grew up uh out in the suburbs] [...] (Schiffrin 1987: 106)
     DM   S   VP   PP

(1b) ----- [I   grew up uh out   in the suburbs] [...] (Schiffrin 1987: 106)
     ---   S   VP   PP

(3a) [He   did   well   on the exam]
       S   VP   adv   PP

(3b) * [He   did   -----   on the exam]
        S   VP   ----   P

Hashtags can similarly appear in an initial and final position, outside of the clause of the utterance, as shown in (14) and (15); the difference with hashtags and DMs is that only 25.8% of the tweets in a general corpus featured an initial hashtag (see section 3.4.1 for more information). Instead, many hashtags (like those discussed in Chapter 4) appear in the sentence final position, as in (15).

(14) Initial: **#whyIleft** I was tired as simple as that just tired

(15) Final: I left because I loved me **#WhyILeft**

The most popular placement of many hashtags is the final position, like that in (15), which is similar to the positioning of utterance final *then*. Other discourse markers like *oh*, *so*, *and*, and *because* (Schiffrin 1988) are more likely to appear at the beginning of an utterance, like the hashtags in (14).

Hashtags like in (16) appear in a medial position, but do not fit the criterion for DMs because they are essential to the structure of the tweet, so this does not fit criterion (i).

(16a) RT @green_ranger23: [I swear if I hear **#GangnamStyle** one more my ima go craz

(16b) RT @green_ranger23: [I swear if I hear ---- one more my ima go craz

Removing *#GangnamStyle* removes the object and semantic theme of *hear*. This use of *#GangnamStyle* is not discourse marker-like because *#GangnamStyle in* (16) does not fit criterion (i). In (17), *#GangnamStyle* does comply with criterion (ii) since it appears after the clause in the tweet.

(17a) [Somebody please tell me why i cant stop listening to this song]

#**gangnamstyle** #problems

(17b) [Somebody please tell me why i cant stop listening to this song]

----- #problems

Example (17) does fit criterion (i) because the hashtag shows up outside the main clause in (17), so it is extra-sentential. This non-optional hashtag functions differently from the two functions investigated in Chapter 2 and Chapter 3 of this dissertation; thus, *#GangnamStyle* is not

discourse marker-like. Instead, Chapter 2 and 3 deal with two of the functions of hashtags that are syntactically and semantically optional.

Criterion (i), where tweets show up outside of the clause, is similar to what other scholars call syntactic optionality, a characteristic of DMs that applies to DM-like hashtags. Schiffrin (1987) discusses how discourse markers are usually detachable and independent of the syntax of the sentence. This is true for many hashtags; if one removes the hashtags in (18), the tweet would appear as (19), and the syntactic analysis of this utterance would not be affected since the hashtag is just attached to the end of the tweet.

(18) [With the roof closing, fans in attendance tonight also receive three innings of a free sauna.] #Brewers

(19) [With the roof closing, fans in attendance tonight also receive three innings of a free sauna.]-----

Removing *#Brewers* does not change the sentence. Not all hashtags are syntactically optional; however, in (20), the hashtag is part of the sentence itself:

(20a) RT @green_ranger23: I swear if I hear #GangnamStyle one more my ima go crazy

(20b) RT @green_ranger23: I swear if I hear --- one more my ima go crazy

Removing *#GangnamStyle* changes the structure of the tweet, as shown in (20b). There is another "optionality", as discussed in Schourup (1999), that is relevant to the analysis of hashtags as DMs.

Another example of a hashtag not fitting criterion (i) appears in (21) where removing the hashtag removes the subject of the sentence:

(21a) You know that song that goes "breaking up is hard to do"? Well, we think **#workingout** is even harder.

(21b) You know that song that goes "breaking up is hard to do"? Well, we think -------- is even harder.

Because of this, the hashtags in (20) and (21) do not fit criterion (i) and are not considered DM-like according to the definition provided.

1.3.2.2 Hashtags and signaling a lower register

Criterion (iv) is where the DM signals a lower register. Schourup (1999) discusses how many DMs are associated with informal speech, calling this characteristic orality. While hashtags are not spoken, hashtags are used in the informal medium of Twitter. Computer mediated communication (CMC) shares qualities of both written and verbal communication (Crystal 2011), and because of this, one could expect to see similar pragmatic devices being used for Twitter, but space constraints and lack of truly synchronous conversations alter the realization of these pragmatic devices. Because of this, a hashtag may not be an oral device, but the informality and nature of online communication on Twitter lends itself to still being considered a DM despite being written.

17

Examples of DMs signalling an informal register appear in (22) and (23).

(22) I believe ... that...y'know it's fate

So eh y'know it just seems that that's how things work (Schiffrin 1987)

(23)   i think **like** the only big parties around here are **like**... the ski and snowboard club ones... (Tree 2014)

Tree (2014) explains that one would "expect *like*s and *you know*s to be less common in writing than speaking" because writing is usually precise whereas these two DMs indicate imprecision. According to Louwerse and Mitchell (2003), "Spoken discourse had more than ten times more discourse particles [DMs] than written discourse." The lower frequency in formal environments emphasizes the less formal nature of these two DMs.

Hashtags are also used in informal writing, like *#fiyah* in (24).

(24) This is proper mega. Little feller had got too much sauce **#fiyah**

Slang terms like *mega* and *sauce* combine with the hashtag in (24) to create an informal tone to the tweet, in a similar way to *y'know* and *like* in (22) and (23).

DMs are often stigmatized because of the informal nature of DMs. Brinton (1996) states, "Because of their frequency and oral nature, pragmatic markers are stylistically stigmatized and negatively evaluated, especially in written how formal discourse.[...]" when listing a number of the characteristics of DMs. This is also similar to hashtags, as seen in the Gizmodo article, when Biddle (2011) states, "Unfortunately, the hashtag is ruining talking. [...] But at their most

annoying, the colloquial hashtag has burst out of its use as a sorting tool and become a linguistic tumor—a tic more irritating than any banal link or lazy image meme," and this annoyance the author describes is very similar to how many react to discourse markers.

1.3.2.3 Hashtags and Attitude

DMs can reveal speaker's attitude, according to criterion (iii). According to Jucker (1993), *well* "reflects the speaker's attitude towards the question."

> (25) A: but who has to buy it
>
> B: **well** the - the state has to buy it but ... (Jucker 1993)

Similarly, #Dafuq expresses confusion and the similar feelings of @staissyfrancis towards the tweet content in Figure 1-1.
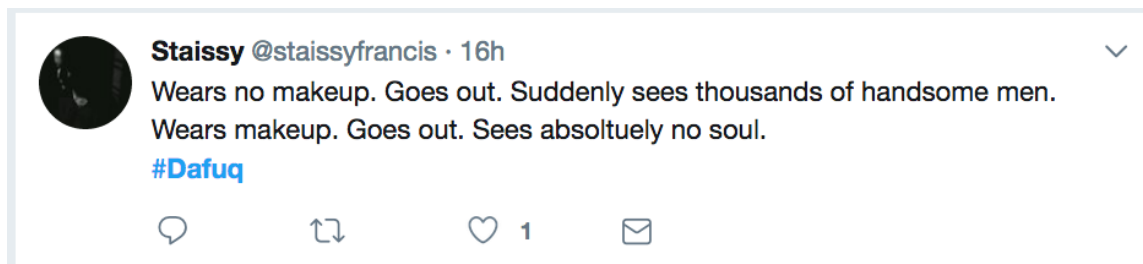
Figure 1- 1: Attitude hashtag

DMs and DM-like hashtags can both flavor an utterance to reveal something about the speaker's attitude, and this feature fits criterion (iii).

19

1.3.2.4 Hashtags and connection

Another characteristic of DMs discussed in the literature is in criterion (ii) which consists of the connective ability of DMs. This is sometimes called connectivity (Schourup 1999). The most common connection that a hashtag provides is connecting a tweet to the larger conversation. De Moor (2010) highlights this when he states, "By preceding a term with a hashtag '#', users can contribute to a conversation about a topic by searching on this term"; similarly, Yus (2011) talks about the connection on Twitter via hashtag by noting the "ability to sustain dense interactions [...] with an explicit wide audience." Using the hashtag can allow a tweeter to connect to a conversation on a specific topic, and by doing so, they can reach an audience larger than their own Twitter followers.

Not all DMs connect on a sentence level. Haselow (2011) states, "final *then* may link the utterance with which it occurs to a larger discourse segment in which a specific state of affairs was discussed or negotiated," and this connective ability is true of many hashtags since hashtags can connect an individual tweet to a larger conversation. In Figure 1-2, 1- 3, and 1-4, tweeters use the hashtags #WhyIStayed and #WhyILeft to situate their tweet content into larger conversations about abuse and violence. In Figure 1-2, three people share survivor stories, as shown below:
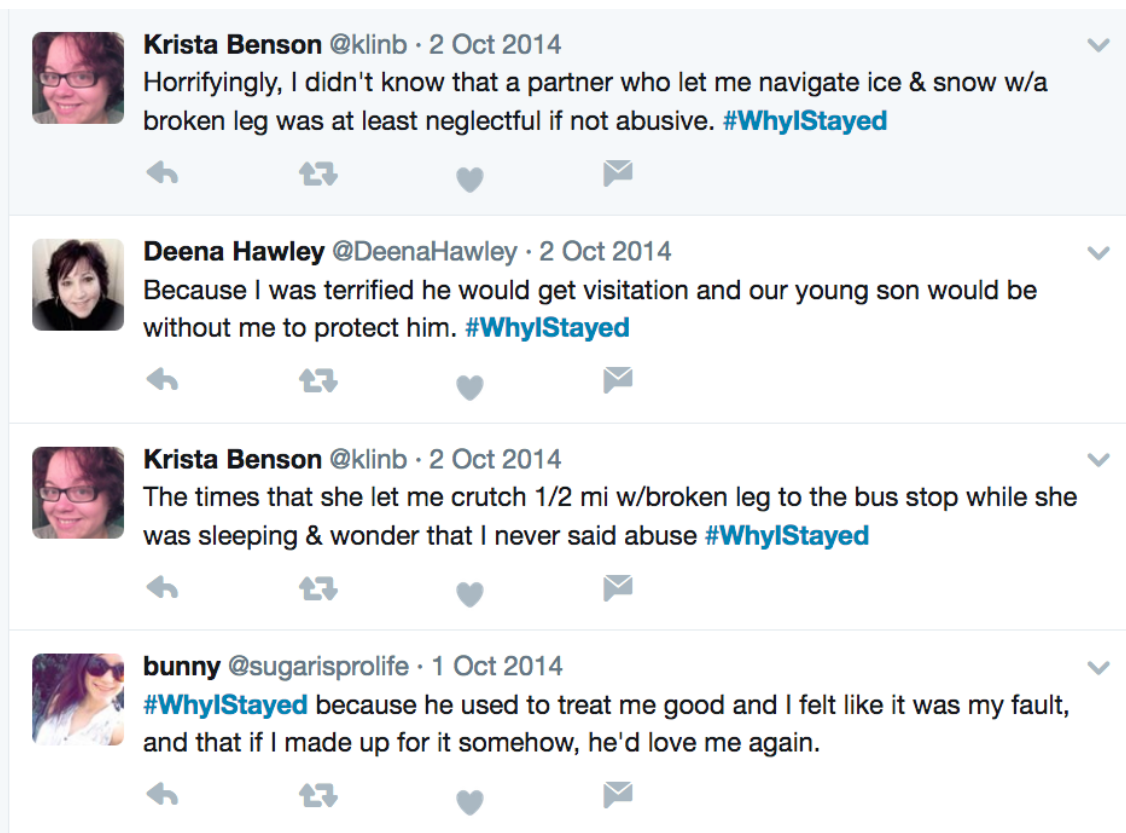
Figure 1- 2: Survivors connecting

The people participating in this conversation on domestic abuse are only limited by access to this social media; therefore, there are quite a number of people who are participating. Because of this, the hashtag could be also working in some capacity as a topic marker, akin to *you know* in Erman (2001) where speakers "use *you know* more frequently to ensure listener involvement, by highlighting a new referent in the discourse, which was apparent in connection with turn-taking" as this topic marking device (more about this function is suggested in the Future Studies section in 6.3.1).

Similarly, in Figure 1-3, @OmniaTweets is using *#WhyIStayed* as a connective to the larger conversation, but in a different way. This user is using the hashtag to talk to those sharing their

stories by telling them the verbal abuse is something they dealt with long before the physical

abuse. Also, by using the reply feature to address @OmniaTweets, @rythemdivine also uses the

hashtag *#WhyIStayed* to not only directly address the first tweet in conversation, but also join

the larger conversation on abuse, as shown below:

Figure 1- 3: Connecting to survivors

While the reply feature and tagging the username also acts as connectives, the hashtag here is

adding to the connection.

Furthermore, we see the same hashtag also being used as a connective in Figure 1-4, but it is

being used by @mcurtis12news to hear more stories from those that might be following along.
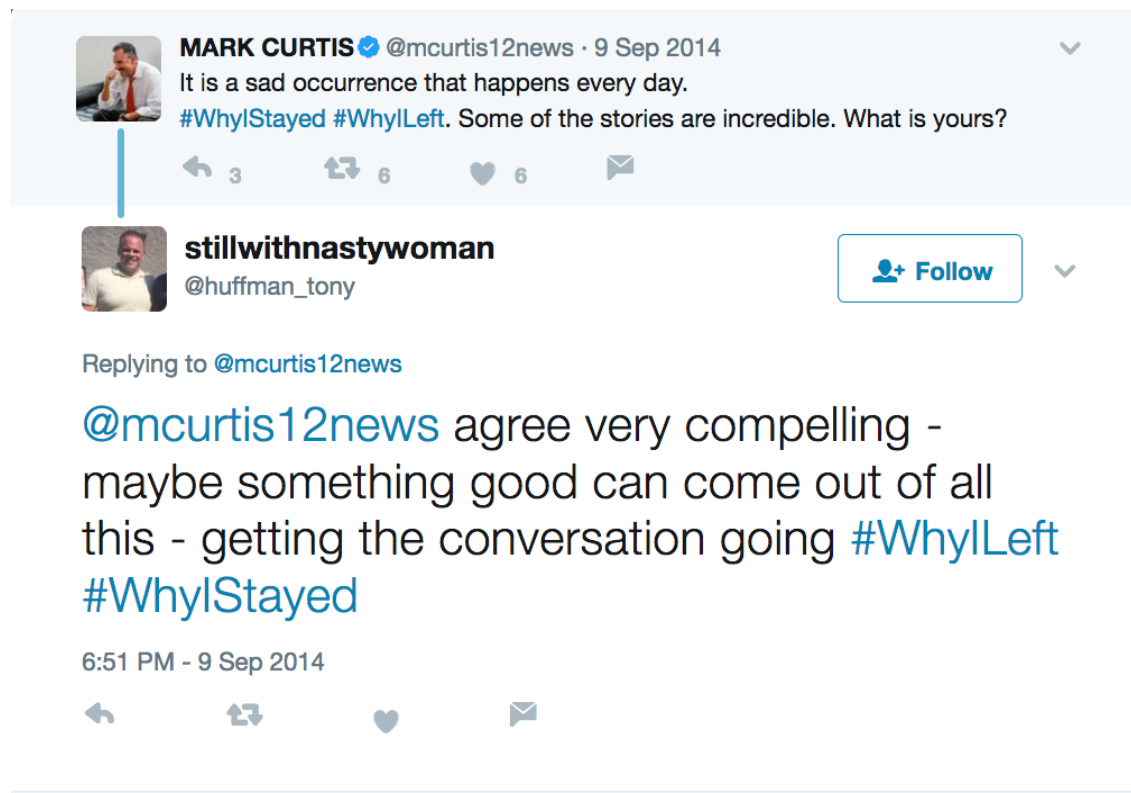
Figure 1- 4: News connective

In these three examples of tweet conversations, we see many tweeters joining the larger conversation about abuse using some trending hashtags. While their contributions to the larger narrative were all very different, they all use the hashtags as connectives. How do hashtags compare to the larger body of connecting DMs, though?

While hashtags do connect to a larger conversation, hashtags may not function in exactly the same way as other well-studied connecting DMs. Schiffrin (1987) discusses the idea that DMs display a relationship between two sentences, while Fraser (1999) asserts that discourse markers impose the relationship. Since scholars who study DMs treat this connection differently, this section will follow from Schiffrin's framework.

Schiffrin (1987) discusses the connective properties of *and*, *but*, and *so*, and gives this example of *and* acting as a connector of events:

> (26) I uh go on trips with 'em [...] and I bring 'em here, [...] and we have supper, or
>
> dinner here, [...] I don't see any problem (Schiffrin 1987)

Müller (2005) shows the connective ability of sequential *so* in the following example:

> (27) [...] restaurant he's an artist. *so* he comes over and starts talking to them. (Müller
>
> 2005)

Similarly, hashtags can act as a connector of events or sequences. In Figure 1-5, *#but* is connecting the two clauses in the tweet, as shown in the following:
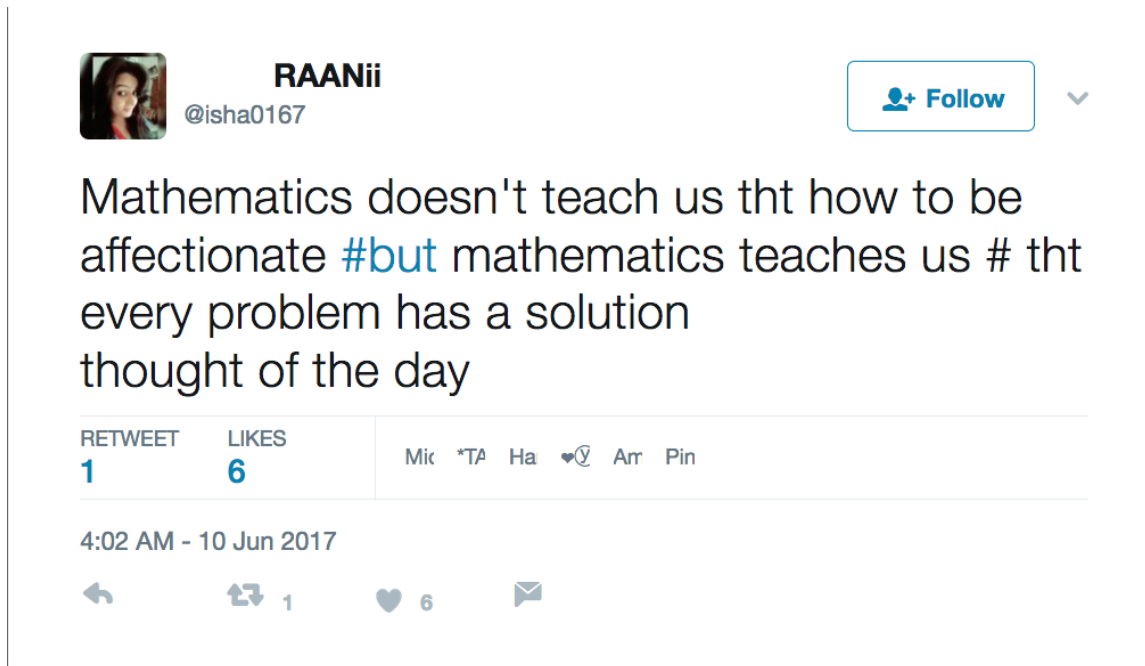
Figure 1- 5: Connecting clauses

Similarly, in Figure 1-6, the hashtag #and is connecting two clauses in two separate tweets from

the same speaker.

Figure 1- 6: Connecting clauses with #and

However, the power of the connector function here is mostly in the conjunctions *and* and *but*.

Also acting as a connector in these tweets is the reply function. So, do we see this connection

with hashtags that are not conjunctions?

The hashtag *#TIL* is an acronym for *today I learned*, where users would usually make a

statement about some fact that they learned during that day. The hashtag *#TIL* is not

normally a connector of events; however, in figure 1-7, the hashtag connects a series of tweets by

the user @look_sharpe, who made a thread of daily tweets on things he learned that day.

Although he had other tweets during this time period that were not about the event of learning

a new thing each day, he connected this series of events through the use of #TIL, as shown

below:

Figure 1- 7: Connecting events from same speaker

Here we have the reply feature and the thread of tweets on different days using the hashtag as a connector. These tweets with *#TIL* are connecting events where the speaker learned something new that day, in a similar way to the connection with *and* in example (26), shown again here as (28).

> (28) I uh go on trips with 'em [...] and I bring 'em here, [...] and we have supper, or
>
> dinner here, [...] I don't see any problem (Schiffrin 1987)

However, this connection is different in that the tweets were written over a series of days (instead of a succession of sentences uttered by a speaker in a narrative). While hashtags can act as connectors to the larger conversation, it may be less clear how they connect one tweet to another on a lower level.

An example of a hashtag not acting as a connection is in Figure 1-8 where *#ConcentrationFaceIsNotCool* only appears once in the search function on twitter.

Figure 1- 8: Not connecting hashtag

Because of this, this hashtag does not fit the criterion (ii).

According to Schourup (1999), one of the ways DMs are optional involves the relationship signaled by the DM. Particularly, this relationship is "still available to the hearer, though no longer explicitly cued" when the DM is removed. In Figure 1-9, the user @BevJohnst uses *#WhyIStayed* to talk to others about abuse, but this meaning is still available since this person tweeted @DivANarcBlog, which shows up as a thread on Twitter.

Figure 1- 9: News connective

This is an example of optionality since the pragmatic meaning (joining a conversation) is still available (tweeting using the reply function) but not as readily accessible.

Furthermore, Scott (2015) states, "Neither does the hashtag contribute any speech act or attitudinal information. Rather, it provides background contextual information which guides the overall interpretation of the utterance. Similarly, one might argue that the use of #WhyIStayed and #WhyILeft in both tweets in Figure 1-1 are optional since both could be found talking to others about similar topics using different search terms (like *sad occurrence*) using a time constraint on the search. However, the hashtags make this connector to the larger conversation more readily available.

1.3.2.6 Semantic contributions of DMs

Semantic contribution of DMs is not part of the criteria of DMs because there is a variation in the semantic contribution of some DMs.

In particular, if we look at how DMs change the meaning of the whole utterance, DMs do not contribute to the truth conditions of the overall utterance, as shown in examples (29) and (30), where (29) is uttered by Karen's high school friends, while (30) is said by her college friends.

    (29) Oh, Karen is moving back home.
    (30) Well, Karen is moving back home.

This characteristic, called semantic optionality in Schoroup (1999), is concerned with changing the truth condition of *Karen is moving back home* with the DM *oh* or *well*. The semantic

meaning in both (29) and (30) both consist of Karen moving back home. The discourse markers flavor the meaning of these utterances, where (29) shows that the speaker is happy about this information, while (30) is uttered by someone who is not happy with the news. By flavoring the meaning, these two discourse markers change the pragmatic meaning since *oh* is a marker of information management (Schiffrin 1987) while *well* is a marker of cause and effect (Schiffrin 1987) and can precede information that one might be reluctant to share (Jucker 1993).

This change in pragmatic meaning without changing the semantics is true of many hashtags, as well, where the tweet (without the hashtag) is not changed by removing the hashtags. Example (31a) is an original tweet, while example (31b) is the same tweet after removing both hashtags *#somuchtodo* and *#oohlookfacebook.*

> (31a) i need to stop procrastinating #somuchtodo #oohlookfacebook
> (31b) i need to stop procrastinating

Both the base tweet in (31a) and (31b) have the same truth value: the tweeter needs to stop procrastinating. Changing the hashtags or deleting the hashtag does not change whether *i need to stop procrastinating* is true or false. However, while the content remains the same if the hashtags are removed, the pragmatic meaning changes, since adding the hashtags *#somuchtodo* and *#oohlookfacebook* changes the pragmatic interpretation of the tweet. Specifically, *#somuchtodo* reinforces the idea that the person needs to stop procrastinating, while *#oohlookfacebook* gives an example of how this person is procrastinating. DM-like hashtags do not change the truth-conditions of the base tweet.

Removing #*Facebook* in (32), however, does change the truth conditions and semantic meaning of the tweet, where the agent of the verb *launched* is missing if the hashtag is removed.

> (32) **#Facebook** has just launched a job search feature http://read.bi/2tNQaZN
> v/@ainecain #FutureOfWork

The hashtag in (32) fails to meet the criteria for being considered a DM because it does not meet criterion (i) where a DM must be extra sentential.

Hashtags and some DMs may carry their own meaning, however. Since words, phrases, or symbols appear after the #, there is often some meaning in the hashtag itself. If we look again at (31), the hashtags become infelicitous when we change the content of the hashtag to something contradictory, as shown in (31c) and (31d).

> (31a) i need to stop procrastinating #somuchtodo #oohlookfacebook
>
> (31c) i need to stop procrastinating #somuchtodo #oohlookfacebook
>
> (31d) i need to stop procrastinating #ineverprocrastinate

On the surface, (31d) is a clash, but it can be felicitous through sarcasm where the speaker would be flouting the Gricean Maxim of quality. Looking at the semantic contributions of example (31a) - (31d), DM-like hashtags carry some meaning, but this is similar to some DMs that carry their own meaning, like *say* or *you know what I mean* which are more contentful that DMs that are sometimes considered fillers like *uh* and *um.*

1.3.2.4 Hashtags and Multi-categorality

Hashtags often feature words or phrases after the #, where either the hashtag can sometimes fit into other parts of speech or other parts of speech are imbedded in the hashtag. Schourup (1999) discusses how discourse markers, over time have fit into various syntactic categories. For instance, *now* and *actually* are two discourse markers that are adverbial (Schourup 1999), like the following hashtag that acts like an adverb:

Figure 1- 10: Hashtag adverb

Schiffrin (1987) talks about conjunctions and their functions as discourse markers, and we see *#and* here acting like a connector of two noun phrases:

Figure 1- 11: Hashtag conjunction

We also see some hashtags acting like interjections, like #ouch:
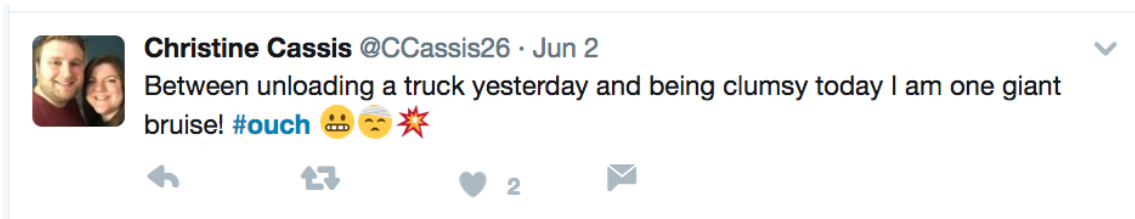
Figure 1- 12: Hashtag interjection

We even see this hashtag on the verb here in *#lied*.

Figure 1- 13: Hashtag verb

In Figure 1-13, it is not being used as an extrasentential DM, but *#lied* is being used as a DM in Figure 1-14, as follows:

Figure 1- 14: Hashtag verb as DM

While the hashtag *#lied* in Figure 1-13 fails to meet criterion (i) since it is part of the main clause, this same hashtag gets used outside the clause in Figure 1-14, in the same way that *say* can be used as both a DM and a verb (but not at the same time).

1.3.2.7 Multitasking DMs and Hashtags

Some discourse markers function in a variety of ways in the same utterance, and this is discussed in Müller (2005) when she states that DMs "under scrutiny fulfill more than one function or at least have sub-functions" which means DMs can have different functions within the same utterance. This is also true of hashtags. DM-like hashtags connect to larger conversations, as discussed in 1.3.2.5, while also functioning in other DM-like ways. For instance, two of the functions of DM-like hashtags studied in this dissertation include clarifying the message and acting as a delay device, which are more observable functions of hashtags.

Schiffrin (1987) and Brinton (2010) both discuss a number of discourse markers that vary in function, and this dissertation will build on it because it will be looking at the functions of hashtags as discourse markers. Not only do hashtags have a similar form to other discourse markers, hashtags promote successful communication on Twitter and other forms of CMC as discourse markers that either repair part of the tweet or signal hesitation. Discourse markers can signal some clarifying information or repair of information. According to Tree and Schrock (1999), "A repair entails a change of state from the original version to the corrected version," and an example of clarification of information appears in (33):

(33) And then he cited four presidents, um Jefferson Eisenhower Kennedy, and uh

someone else for um. . . for being adulterers, but- but being very great presidents, **oh** and

Roosevelt.

Schiffrin (1987) describes this clarification function of *oh* as being involved with "information

management." Similarly, *I mean* also manages information when it clarifies the original message,

as shown in (34)

(34) she appears to be perfectly happy - . **I mean** she can't be a hundred per cent happy,

nobody is, but she appears to be happy (Tree and Schrock 2002)

Tree and Schrock (2002) describes *I mean* as a DM that can "forewarn adjustments", while

Schiffrin (1987) states that "*I mean* focuses on the speaker's modification of his/her own talk."

Chapter 4 will discuss the type of hashtags that clarify the message of the tweet in a similar way

to *oh* and *I mean*.

Discourse markers can act as delay devices. Some English discourse markers that are known to

act as a hesitation or delay are *well* and *uh*, as shown in (35) and (36).

(35) Uh, **well**, I'm not sure about how much you know about China. Well, it's a
beautiful city. (Baiat et. at 2013)

(36) th- there is a a **uh** a potential problem (Clark 2002)

According to Jucker (1993), *well* can function as a delay device; however, the author also notes how the function of delay device also overlaps with the function of face-threat mitigator. Clark (2002) contrasts the difference between the "best known delay signals [...] *uh* and *um*" noting that these two DMs signal different length delays. Buysse (2012) notes how the discourse marker *so* can be used with vowel lengthening to mark "hesitation or reflection" (2012:1770). Similarly, Wang (2011) discusses how the Japanese DM *ano* and the Mandarin Chinese DM *nage* expresses hesitation in sharing personal information. Hashtags also act as delay devices, as discussed in Chapter 3.

Jucker & Ziv (1998) list some of the many functions of discourse markers as intimacy signals, hesitation markers, repair markers, and attitude markers. The repair markers are relevant to the functions described in Chapter 2 (and applied in 3), while the hesitation markers are relevant to Chapter 2 and 3. Functions of hashtags that are not discussed at length but are present include topic orientation markers (Fraser 2009) and attitude markers (Jucker & Ziv 1998).

1.3.3 Approaches to studies of DMs

Discourse marker research can have different scopes. Some research concentrates on specific discourse markers, like *well* in Jucker (1993), and *dui bu dui* in Chen & He (2001). Brinton (2010) gives an overview of several discourse markers. Other studies limit their scope to just two or three discourse markers, like with *say* and *why* in Stvan (2006) and *say* and *zeg* in Van Olmen (2013), where the different markers are functioning in some similar yet distinct ways. Some chapters of this dissertation limit the scope to compiling a corpus with just one or two hashtags. This is very different from the chapter that analyzes a general corpus, where there are

no limitations on the hashtags. Brinton (1996) and Brinton (2010) both analyze markers

diachronically, and two of the content chapters in this dissertation will similarly do studies of

how hashtags were used over time.

1.4 Research questions

Pulling from these topics, I have distilled them into the following main research questions:

(I) What are the discourse functions of hashtags

(II) How do these compare to functions of other discourse markers?

To investigate these two main research questions, each chapter addresses a different research

question, as follows:

(III) Which traditional discourse markers act as delay devices in survivor story interviews? What environments are motivating the delay discourse markers?

(IV) For those hashtags involved in hesitation, what are the motivating factors behind the initial placement of the hashtag? Are some environments in tweets disclosing violence more likely to feature an initial hashtag?

(V) When hashtags clarify a message or signal pragmatic repair, what categories based on copresence heuristics exist in the data? Furthermore, are some categories more used than others?

(VI) How can tag reframing be used to classify data and assist with analysis based on community membership?

In order to effectively answer these research questions, four distinct Twitter corpora were used

to form the analysis. To answer research question (III), the corpus for Chapter 2 consisted of the

transcripts from spoken interviews of survivors, gathered and transcribed from videos from web

sources. In Chapter 3, research question (IV) was answered using a corpus that consisted of at most 20 tweets per day, but for a long time period: I collected 443 older tweets using Google Webscraper over 73 days. In Chapter 4, research question (V) was answered using an already collected general corpus of tweets. In Chapter 5, research question (VI) was answered with a big-data corpus of almost 56,000 tweets over 18 days using RStudio and the API provided by Twitter. Chapter 3 and 5 needed very specialized corpora narrowed down by the specific hashtag to investigate the research questions for each. In particular, it would have been impossible to analyze hesitation placement of hashtags from Chapter 3 in a general corpus since most tweets have hashtags that appear in a tweet final position. While the methodologies may be very different in each chapter, this dissertation makes use of the four different corpora of each to answer the four research questions.

In this dissertation, I posit that hashtags are discourse markers where space constraints of 140 characters on Twitter complicate their realizations. In Chapter 2, I look at spoken discourse markers that appear when the speaker discusses emotional narratives. In Chapter 3, I present hesitation hashtags and show through statistical evidence how some tweet-initial hashtags act as a hesitation at the beginning of the tweet. In Chapter 4, I investigate the tag reframing hashtags and how they can change the meaning of the tweet using the heuristics presented in Clark and Marshall (1981). In Chapter 5, I show how tag reframing can actually be used in conjunction with a learning algorithm to analyze the reclamation of a pejorative adjective via hashtag. This dissertation, through the progression of research questions answered in each chapter, analyzes how hashtags assist with felicitous communication to the intended audience of the tweet.

CHAPTER 2

DELAYING DISCOURSE MARKERS IN SURVIVOR INTERVIEWS

My research is predominantly in CMC, as shown through the introductory chapter of this dissertation; however, for a point of contrast, this chapter instead looks at a spoken corpus where people also use DMs and other delay devices to talk about surviving violence.

This project for Chapter 2 began after already investigating the preliminary patterns I discuss in Chapter 3 with Twitter data from the 2014 Domestic Abuse Awareness campaign, where I found hashtags appearing with an initial placement more frequently when people talk about surviving violence. Because I was interested in seeing what people linguistically do to signal delay in spoken disclosures of survival, I move beyond Twitter data in this chapter, to look at the discourse markers used in spoken survivor interviews to see if they share the same delaying function of initial hashtag placement seen in Chapter 3.

In section 2.1, I introduce the research question in this chapter, while I explore relevant literature in section 2.2. The methodology, including how I created the corpus and what the corpus consists of, is discussed in 2.3. In 2.4 and 2.5, an analysis for this delaying *and* is given, and in 2.6, I present conclusions from this chapter.

2.1 Background

2.1.1 Overview

This chapter looks at patterns in discourse markers and other delay devices in interviews and other narratives containing public disclosures of violence. DMs like *so* (Buysse 2012) and *well* (Jucker 1993) are known to precede information that one might be reluctant to share. Because the data for this corpus consists of both emotional narratives that are of a somewhat public nature, speakers may want to avoid disclosing such personal information about their traumatic experience too directly. Because of this, we would expect to find a number of delay devices. However, *well* only appeared 11 times, while *so* appears 105 times. Because some of these tokens of *well* and *so* were not marking a delay, I was curious to see which discourse markers speakers do use to show hesitation. In exploring this, I found 682 tokens of *and*, and I postulate that at least 70 of these tokens of *and* act as more than just a connector. I show through the analysis of data from a corpus of survivor speeches that *and*, like *ano* and *nage* (Wang 2011), acts as a hesitation before the speaker divulges personal information.

Using a corpus of survivor stories, I show that *and* is a frequent discourse marker in this type of corpus, but while many of these tokens of *and* function as a connector of ideas and assist with the organization of the discourse, some go beyond this function, and act as a delay device. Delaying *and* can signal to the audience that the speaker is hesitant to share the information, and it allows the speaker to distance himself or herself from the revelation.

2.1.2 Discourse Marker AND

As defined in Chapter 1, DMs are extra sentential words that are used to organize within a discourse, while revealing context connections and speaker attitudes. Along with that definition, DMs must be (i) positioned outside the clause structure; AND must also do at least one of the following: (ii) work to signal the coherence relation or position of the utterance to some other part of the discourse AND/OR (iii) show the speaker's attitude toward an utterance, AND/OR (iv) signal a lower register (i.e. informality).

Some instances of *and* are DMs, as shown in (1) where *and* is (i) positioned outside the clauses and (ii) signals a relationship between the two clauses, in this case temporal sequencing:

> (37) [I uh go on trips with 'em [...]] **and** [I bring 'em here, [...]] **and** [we have supper, or dinner here, [...] I don't see any problem] (Schiffrin 1987)

In contrast, other instances of *and* act as just a connector of phrases, like example (38):

> (38) [He   ate   peas   **and** carrots]
>   S   VP [V   NP   conj NP]

This instance of *and* is not a DM because it is not (i) positioned outside the clause, as shown with the gloss under (38). Also, this instance of *and* does not fit any of the other criterion since, e.g., it does not (ii) show an ordering relationship between *peas* and *carrots.*

Since instances of *and* where it connects separate clauses already act as a DM by meeting criterion (i) and (ii), where the main function is connection, I next investigate whether this DM might also be multifunctional, like some other DMs. In particular, in this chapter I show how *and* can combine with other DMs to not just connect clauses, but also go beyond this function to act as a delay device when used in emotionally revealing utterances.

2.2 Literature

In section 2.2.1, I discuss the literature involved with interviews and public discourse as data. In section 2.2.2, I will present previous studies on delay devices that are relevant to the study of the interviews in this chapter.

2.2.1 Literature on Interviews and public speeches

Since this chapter uses a corpus of interviews and public disclosures of violence, I first look at previous linguistic studies that used similar data sources. Schiffrin (2001; 2001; 2002; 2003; 2003), for example, has studied the characteristics of oral histories in a number of works, particularly with respect to Holocaust survivor stories. This chapter will build on such work by also studying two Holocaust survivor stories in this corpus. One of the survivor stories in my corpus is a more formal televised interview. Prepared talks, like those given in interviews by the speakers for this survivor corpus, may be void of many discourse markers because of the prepared nature. However, Han (2011) finds that not only are discourse markers used in public speeches, but *and* is the most frequent marker in a corpus of famous public speeches. Because Han found so many discourse markers in famous public speeches, we would expect to find something similar in this study, as well. And indeed that was the case for my corpus.

2.2.2 Literature on DMs

Previous DM scholars have studied *and*, but their research fails to account for this hesitation function found in the current study. Dorgeloh (2004) notes the frequency of sentence-initial *and* in different types of text. Her work found that *and* is used more frequently in denarrativized texts; we would expect to mostly find this sentence-initial *and* in genres that are still somewhat narrative in nature, like in the narratives in this survivor discourse corpus.

Schiffrin (1986) claims that *and* has two functions: "a structural device for building a text, and a marker of speaker continuation in interaction." According to Schiffrin, the structural function is "only because of its relation to its containing text" whereas the continuation function is because of its context-bound use. In other words, *and* is multifunctional, but these two functions are derived from separate discourse motivations.

Schiffrin (1987), in a book about different discourse markers, looks at how *and* is used in discourse, but analyzes two basic functions of *and*: (i) *and* can coordinate ideas, and (ii) it can signal a continuation of the discourse. However, even though this work shows how this connector acts in relation to other connectors and how it serves to embed ideas, this work fails to look at other possible functions of *and*. A previous work of Schiffrin (1986) looked at *and* specifically, but noted how tense does not change with *and* like it does other DMs like *but* or *so*. Schiffrin (2006) looked at more functions of *and*, and one of the functions proposed in her work is similar to that presented in this dissertation. One function of *and* called "in a list of something about yourself now" is found in a survivor story with a Holocaust survivor. Finding other funtions of *and* is similar to the work in Chapter 2, but another similarity is the corpus: I

44

compiled a corpus of survivor stories, two of which are survivors of the Holocaust. Schiffrin found *and* when talking about personal information in the present, where this dissertation finds delaying *and* being used when revealing traumatic information about the past.

2.3 Data & Methodology

In section 2.3.1, I discuss how I created the corpus, and in section 2.3.2, I give a description of the data in this corpus.

2.3.1 Corpus Creation

To study the delaying function of *and*, a corpus of survivor stories was gathered from available sources on the Internet. Twelve stories make up this survivor corpus: One is from an interview on televised news where a rape survivor talks at length to a journalist (WTNH News8 2012). Two other speakers were part of the Holocaust Oral History Archive (Horwitz 2007; Vine 1983), and the remaining nine survivor stories were collected by the Office for Victims of Crime as a training material (OVC TTAC 2015). For the news story, no text file was available, but the entire half hour interview was available on video through Youtube so I was able to replay it and transcribe the material. For the two Holocaust survivors, their stories were documented in text on the website and the sound file was available to assist with the detailed transcription. Finally, the stories used for training were collected both in video and text file, which assisted with the detailed transcription, as well.

After finding these twelve survivor narratives, the stories were carefully transcribed into a .txt file, where it was then processed using the concordance software AntConc (Anthony 2014).

Using the word list function in AntConc, there were many tokens of *and* that appeared in the corpus. After looking at *and* in the KWIC window, I was able to identify the environments for each and where it went beyond just connecting ideas.

2.3.2 Data Description

The corpus consisted of 14,473 tokens, with 1,852 word types, according to AntConc. The corpus consisted of 12 speakers, many of which experienced different types of violence. Two directly involved some type of sexual assault, but the story from the domestic abuse survivor relayed information about how sex was used against her in this relationship. The story of Rob contained different types of abuse he suffered as a child, some of which was at the hands of a pedophile.

Two men recall their experiences getting assaulted by strangers. Two stories were by Holocaust survivors recalling their time in concentration camps. Another story involved the mother's perspective after an accident with a drunk driver, and she tells about her experience caring for her severely injured child and her subsequent death from injuries related to that accident. Similarly, not all of the stories directly involved the speaker: Peggy's story involved recalling the events that surrounded the death of her son by an arsonist while Teri's story involved the events around her son's death after getting involved with a gang. Jee Young's story involved recalling the events when her brother was attacked during a hate crime.

All of these stories come from very different types of people: they vary in age, sex, location, and background. However, as I will show in the subsequent section, they all use *and*, specifically *and um* or *and uh* or *and* in repetition before revealing something traumatic.

As a point of comparison, there were 682 tokens of *and*, and this is key, or markedly high in relative frequency, compared to the Brown corpus (Francis & Kucera 1961): using AntConc to compare the corpora, *and* has a keyness of 150.065. In the next section, I will show the functions and types of this quite frequent word in the survivor corpus.

2.3.3 *Coding delaying and*

In this survivor corpus, many people are divulging personal information about some traumatic event. In the following section, I show that *and* appears in a number of environments, but these all function as a delay device before an utterance recalling particularly painful information. The power of the delay device in these survivor stories is in the combination of *and* with another entity, like appearing before *uh* and *um* (Section 2.3.3.1), and appearing in repetition (Section 2.3.3.2). Other instances of *and* may also have that delay function (for instance, if it features vowel lengthening, it may also be a delay). All of these delaying *and* instances may be on a scale of delay, where some instances may have more power of a delay than others.

Some of these scalar delays appear in Section 2.3.3.3, when the delay appears before an utterance marked with agency.

2.3.3.1 *Appearing before uh and um*

When many of the speakers talk about the perpetrator, they use the phrase *and um* or *and uh* before the statement, as seen in example (39-42):

(39) I remember seeing the blood spraying on the pavement and uh he picked me up by my hair, opened the passenger back door behind the driver side and threw me in the car. (WTNH News8 2012)

(40) I was saying no and please and um he was just swearing at me telling me to shut up and um at that point he uhh pulled his pants down and forced me to give him oral sex (WTNH News8 2012)

(41) And uh when he took me out like that, it changed my life. (OVC TTAC 2015)

(42) All you heard is, is, is, is uh, shootings and hittings and uh, they were running around with these sticks and hitting you over the head. (Vine 1983)

In example (39) and (40), Amy is prefacing the details of her rape with and *uh* and *and um,*

delaying the telling of this news. In (41), Alan, who was crying by this point in the video, uses

and *uh* in a similar way when talking about his attackers. Finally, in (42), George, while

discussing the violence surrounding his arrival at Auschwitz, uses *and uh*, as well. In this corpus,

there were actually 67 instances of *and* that collocate with *uh/um*, as seen below in Figure 2-1:
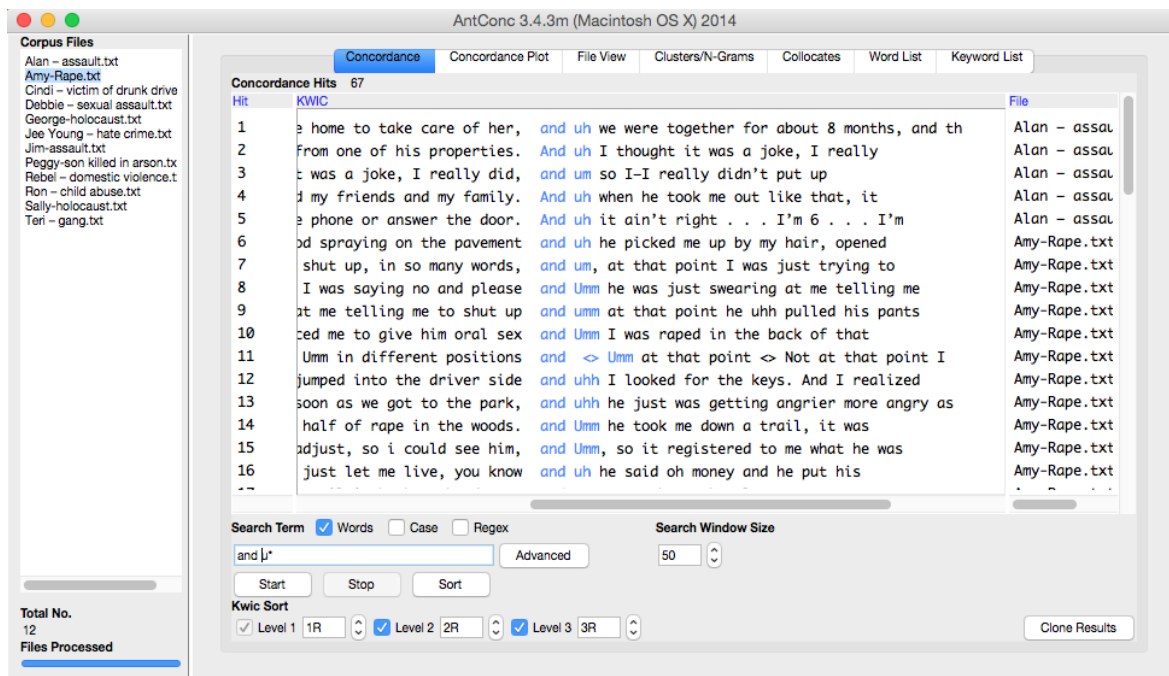
Figure 2- 1: AntConc Concordance of and uh/um

The speakers are using *and* here, while in an emotional state of retelling their story, as a way to preface sensitive information. While all of these speakers directly experienced the violence, this same pattern is also present when the sister of an assault victim discusses the actions against her brother when a group committed a hate crime against him, where *and uh* comes before *they*, the pronoun that refers to his perpetrators:

> (43) Let's jump on him now when he's alone. Um My brother turned around, **and um** they circled him, **and um** they were like "Chink, go back to your country" **and um** we're not even from China. (OVC TTAC 2015)

These instances of *and* which collocate with *uh* and *um* are acting as delaying markers. Not all

of these tokens of delaying *and*, which collocate with *uh* and *um*, appear preceding information

of the offender- sometimes they used this when recalling other traumatic details, as seen in (44):

> (44) they would get sores and uh, and infection and, and, and uh, get poisoned and die
> (Vine 1983)

In (45), when the mother reveals to the camera that her son passed away due to the actions of

another, she also uses this delaying marker:

> (45)And uh you know, Jerry didn't make it, you know (OVC TTAC 2015)

Similarly, Alan prefaces the information about his current state of mind after the assault

with *and um*, as seen in (3), which appears below as (46), and (47):

> (46) And uh when he took me out like that, it changed my life. (OVC TTAC 2015)

> (47) I'm scared to death to answer the phone or answer the door. And uh it ain't right
> (OVC TTAC 2015)

In both (46) and (47), Alan uses the same discourse marker *and uh* to preface information he

may be reluctant to tell the person interviewing him. One example seemed potentially

problematic for this pattern:

> (48) **and umm** his wife came to the door (WTNH News8 2012)

At this point in the video, Amy is talking about the events after the attacker left her for dead. It seems, by this line, she has finally found a safe haven- so why the *and um*? However, by taking into consideration the larger context, we can see that the bad news comes a little later:

> (49) **and umm** his wife came to the door and they shut the lights off and locked the door and they sorta just called 911 but left me outside **and umm** at that point i just said oh geez (WTNH News8 2012)

Amy was not really safe: the couple left her on the porch bleeding, which is troublesome to the speaker. These examples show how *and*, when it appears with *uh* or *um*, is a discourse marker that functions as a delaying device; however, does *and* necessarily have to appear before *uh* or *um* to have this function? When *and* appears after *uh* or *um*, it also acts as a delaying device, and there are 12 tokens of *and* in this corpus that fit the bill. In (50), the speaker discusses "the most painful" part of the abuse she suffered:

> (50) He was sexually abusive. um... and I think of all of it, that was probably the most painful and still probably the hardest to get past um(OVC TTAC 2015)

During this, she also used pauses and she wiped her tears during this line, as well. This speaker is using both *um* and *and* as a way to delay this revelation. This pattern can also be seen in (51) below:

> (51) I had blood was coming out of my mouth. um and I think I scared the dickens out of him (WTNH News8 2012)

> (52) mean time i was just rolling with any impact i was receiving, uh and at some point i just told myself let go let go let go. I just didn't want to feel it anymore. and thats when i just stopped responding. (WTNH News8 2012)

In (51), Amy is recalling the events of showing up on the porch of someone who lived nearby the location of her attack, and in (52), she is telling about the attack itself. In both, she uses *and* after *uh* and *um* when recalling specific details. Does *and* only act as a delaying marker when it collocates with *uh* or *um*?

2.3.3.2 Appearing in repetition

There were some instances of *and* that clearly signal the speaker's hesitation of sharing information. In (53), George discusses the horrid living conditions at Auschwitz using the repetition of the word *and* to hesitate relaying this information to the interviewer:

> (53) The new camp was filth and dirt and, and, and, and these cell blocks were not done yet and it was uh, uh, uh, conditions were horrible. (Vine 1983)

George uses the same device when discussing all of the death that surrounded him there:

> (54) A lot of people committed suicide. Like, went over to the fence, got electrocuted, because they couldn't wait, that could take sometimes three days. And is, and, and, and, you know, and you knew in your own mind, and we were experts already at that time, we knew that once they take you, you go (Vine 1983)

There were 27 tokens of delaying *and* that seem to fit into this environment specifically. This shows that *and* does not necessarily have to collocate in some way with *uh* or *um*, but these are the most apparent using concordance software.

2.3.4 Issues with Coding

2.3.4.1 Connecting ands (not delays)

Not all tokens of *and* are delaying discourse markers; some merely organize the discourse by

acting as a connector. We see the connector *and* in Peggy's story, where she uses this term to

connect two noun phrases:

(55) Chrissy **and** I were watching TV the other day (OVC TTAC 2015)

This fails to meet criterion (i). This type of *and* only connects and nothing in the utterance or

this part of the video seems to indicate that the speaker is hesitating in any way. Instead, this

term is being used to organize the internal structure of the sentence. Similarly, in (56), Alan

uses the first instance of *and* as a discourse marker to connect this utterance with the prior

discourse.

(56) **And** I always had confidence in myself, to take care of myself and my friends and
my family. (OVC TTAC 2015)

This is different from example (3), which is a previous utterance from Alan, mentioned here

(57):

(57) **And uh** when he took me out like that, it changed my life. (OVC TTAC 2015:
Alan, 2015)

In this, he is clearly using *and* as a delay device before revealing information about his assailant, while in (21), he uses sentence initial *and* in a continuation of the discourse.

Sections 2.3.3.1 and section 2.3.3.2 show examples of and that show up with other delays where the and is. This is different from connector and which Amy uses to signal the continuation of the speaker in (18):

(58) **And** I locked the door. (WTNH News8 2012)

Although this meets criterion (i), it also functions for (ii) where it connects, but it does not meet the other criteria.

2.3.4.2 Difficult to determine type of and

While *uh* and *um* might be indicators of a hesitation, the content of the upcoming utterance prefaced by delaying *and* may be motivating the delay device, and in turn, it may be a greater predictor on the function of *and* before an utterance.

There were many examples where taking prosody into consideration could have helped the coding of delaying *and.* Because of this, more instances of *and* may be acting as a delay, but without the other delays, there would need to be another indicator of delay, like length of vowels (like the vowel lengthening found with hesitation *so* in Buysse (2012)).

In many of the examples, it is difficult from the text to actually determine which type of discourse marker is being used. For instance, in (59), it seems like Amy also uses *and* as a way

to connect the events when she took power over her situation, but she does use many other

discourse markers, like *so* in the same utterance, so the type of *and* is difficult to determine.

> (59) So, I snuck the phone and I hid it underneath umm a coat in the back seat of my
> car and I dialed 911 without him knowing.(WTNH News8 2012)

Because these two tokens of *and* also appear with other discourse markers, one might jump to

conclusions and say both are acting as delay devices, but it seems like the other discourse

markers are used in (59) not as delay devices. *So* seems to be connecting this in the discourse

and situating it as a continuation to further the narrative, while *umm* seems like a disfluency. If

that is the case, it is less likely that *and* is acting as a delay device, but more as a connector.

2.4 Analysis

2.4.1 Agency and delaying and

One might expect, in a corpus of survivor stories, the speakers may use language to convey the

fact that they feel victimized since "victims are, by definition, passive objects who have been

acted upon by other forces, not active agents" (Gilligan 2003:30). Similarly, Schiffrin (2002)

found "a mix of agency and passivity in the descriptions of forced labor assignments in camps,"

but she notes that her analysis supports previous work in oral histories that show these types of

narratives fail to uphold "artificially imposed dichotomy between survivor as agentive "hero" or

passive "ghost." Delaying *and* appears in the following example where the passive construction is

used to refer to the sickness and maladies affecting the people at the concentration camp.

55

(60) Uh, they would get sores and uh, and infection and, and, and uh, get poisoned and die, you know, from just wearing these things (Vine 1983)

The delay devices used in (60) come before passivation of poison, and verbs like *die* obscure the agent, or perpetrator responsible for the loss of lives at this time. The same pattern emerges in (61), when Jee Young recalls what brother's attackers said to her after:

(61) **And um** the response I got from him was, "I'm sorry I caused you such inconvenience, and um but if you think I'm a racist, I'm not 'cause I don't treat people by their color." (OVC TTAC 2015: Jee Young, 2015)

In (62), instead of phrasing it as making the attacker the agent of the sentence, the verb *respond* was nominalized and the agent of "respond" is moved into the prepositional phrase, which deemphasizes and obscures the agent and the speaker in a way that is very similar to passive constructions. Also, given the information on how victims might use language to portray their victimization, we would also expect the speaker to not have agency or power in much of the discourse, and this can be seen in (62), where Alan starts to recall the events at the beginning of his assault.

(62) **And uh** I thought it was a joke, I really did, **and um** so I...I really didn't put up any resistance, **and** then he slid out a metal baseball bat, **and** he started to attack me (OVC TTAC 2015: Alan, 2015)}

In (63), he uses *and uh* and *and um* before an utterance where the subject of those utterances fails have an affect on some outside entity, which is one of the indicators of agency. However, he also uses delaying *and* when the third person pronoun enacted the violence when he slid out a metal baseball bat and started to attack. Agency and the content of the disclosure may be motivating the delaying *and* appearing before utterance since we see *and* appearing when the perpetrator has agency and the survivor does not. This is quite different from how Alan later speaks about his future, after the attack:

> (63) I look to the future **and** I see things, but it's... I'm not impatient, but I really want to get back to where I was (OVC TTAC 2015: Alan, 2015)

When he takes the active role of the verb *see*, he uses *and* less, and does not use delaying *and*, but the token of *and* in (63) is acting as just a connector.

After using AntConc to see what contexts *and uh* or *and um* appear in, I found that from the 67 tokens, 5 of these are passivized in some way that obscure the survivor/victim and the assailant. However, many more seem to show some pattern with agency: 25 of the 67 tokens appear where I lacks agency or control over another outside entity or power in the situation, like in example (64) and (65).

> (64) **and um** we were hit head on by a drunk (OVC TTAC 2015)

> (65) **and umm** i was such a mess, my eyes were swollen shut.(WTNH News8 2012)

Also, 21 of the 67 tokens come before a third person pronoun (singular or plural), referring to the assailant, that appears to have agency of some kind.

(66) **and uh**, he decided to burn the synagogue. (Horwitz 2007)

(67) **and uh** he picked me up by my hair, opened the passenger back door behind the driver side and threw me in the car. (WTNH News8 2012)

This means only 16 of the 67 tokens of *and uh* and *and um* are unaccounted for with this.

2.5 Conclusions

The discourse marker *and* does appear many times in the corpus of survivor stories, as shown by the keyness value in section 2.3, and part of the reason for the large frequency may be due to the fact that speakers are using *and* in a way that goes beyond just connecting ideas. Delaying *and* can be found with *uh* and *um*, but it also can be easily found in repetition, as well. Delaying *and* seems to most often appear with some type of agency for the assailant or lack of agency for the victim, when we just narrowed it to the 67 tokens of and narrowed by the query *and u**, which indicates that agency may be motivating the usage of this delaying *and*. If speakers know they are about to reveal information about how someone else had power over their traumatic experience or if speakers know they are about to reveal not having any power in the situation, he or she seems more likely to hesitate, and employ the discourse marker *and* as delay device, akin to how early hashtag placement was used in Chapter 3. In the next chapter, similar patterns with hashtags are presented.

CHAPTER 3

DELAYING DISCLOSURE VIA HASHTAG PLACEMENT

In my previous observations on Twitter data, I noticed that most tweets have hashtags at the

end of the tweet. However, this notion of hashtag placement was challenged in October 2014

when I started collecting some tweets that were talking about survivor stories. I noticed that

many more of these tweets featured an initial hashtag placement. Using a corpus of tweets from

the 2014 Domestic Abuse Awareness campaign on Twitter that used *#whyistayed* or *#whyIleft*,

I noted an unusual pattern in the hashtag placement that I hypothesized might correlate with

the content of the tweet. In particular, I noted that a large number of these tweets had the

hashtag at the beginning, rather than the end of the tweet, as shown in Figure 3-1, where only

25% of tweets in general are at beginning of the tweet, vs. 39% in initial position for the

survivor story tweets.

Table 3- 1: Comparison of initial hashtags in survivor corpus to general corpus

	Yes Initial	No Initial	Total
survivor tweet corpus	173 (39.05%)	270 (60.95%)	443 (100%)
2012 general corpus	463 (25.82%)	1330 (74.18%)	1793 (100%)

Since fronting the hashtag delayed the main tweet message, I examined whether there was a similarity between this hashtag placement and other uses of delay devices. Especially considering that other discourse markers are used as a delay device. This chapter will show that hashtags, when placed in this tweet initial position, can act as a delay before private information, in a similar way to spoken DMs discussed in chapter 2.

3.1 Introduction

In the campaign, people used the hashtags *#whyIstayed* and *#whyIleft* not only to situate their personal story into some larger campaign, but also to divulge to this global community potentially intimate information in the same tweet as the hashtag. However, in doing so, members of this community are also revealing more than their suffering: the hashtag placement is motivated by the content of the tweet, and there is some statistical significance associated with these patterns. Some hashtags that appear at the beginning of the tweet are doing more than just marking the topic- some are acting as a delay device, akin to some discourse markers. While this is not the only function of hashtags or discourse markers, evidence shows that this is one of the functions that is quite prevalent in this content specific corpus.

This study shows how early hashtag placement can act as a delay device in tweets that convey survivor stories. While speakers naturally hesitate when talking (Erard 2008), more hesitation is used when discussing sensitive information; delaying devices, including discourse markers,

frequently appear in emotional narratives (Romano 2014). Given the colloquial nature of many tweets, delay devices should appear in emotional narrative tweets; however, the character limitations on Twitter (users are confined to 140 characters or less) complicate how hesitations and discourse markers are realized.

Hashtags, as discourse markers in computer mediated language, can act as delay devices in computer mediated communication when they appear in a clause initial position in a tweet. In particular, the placement of the hashtags *#whyIstayed* and #whyIleft have changed over the course of the broader discussion on domestic abuse. Since the first instance of #whyIstayed used with a self-disclosure actually features a late placement hashtag (hashtags in a clause final position in a tweet), one might expect this to affect how others use this hashtag to join the larger discussion. However, the prevalence of early hashtags (hashtags in a clause-initial position in a tweet) in the campaign as a whole is indicative of pragmatically motivated delay that the early placement creates.

3.1.1 Background

During the 2014 Domestic Abuse Awareness campaign, two hashtags *#whyIstayed* and *#whyIleft* started trending, in part due to the controversy surrounding the news stories about how Ray Rice, an NFL football player, was arrested for assaulting his fiancée Janay (Erzikova et al. 2016). In particular, these hashtags emerged at a time when the media was victim blaming Janay for staying with her abusive partner. As a reaction to that, survivors of intimate partner violence started tweeting their stories to make others aware of the fact that it is often difficult to leave abusive situations.

The hashtag #*WhyIStayed* was begun by Beverly Gooden, a domestic abuse advocate, when she

tweeted stories of her abuse and the abuse of others (Matson 2016). The first tweet of the

campaign from September 8, 2014 appears below in Figure 3-1.
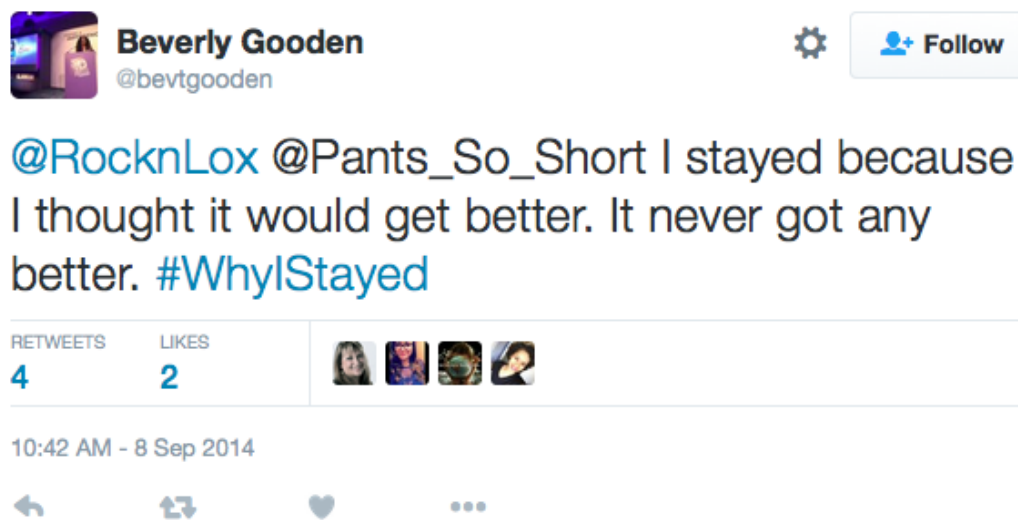
Figure 3- 1: First Tweet of Campaign

The hashtag #whyistayed was not inherently linked to survivor stories, however. Before the

campaign, #whyIstayed was used to talk about vacation and community outreach-- it was not

used to talk about stories of abuse or survival, as shown in Figure 3-2.

Figure 3- 2: Before Campaign

Before the campaign, it was only used 6 times in in 2014. However, after the campaign started, it was used 1879 times on the first day September 8, 2014.

Weathers et al. (2016) discuss four themes of topics found in the first day of the campaign, which include "four reasons for staying in abusive relationships: (lack of) resources, responsibility for abuse, fear, and gender-linked power" while investigating "how women tell their stories within the larger societal structure in which they live through Twitter." I use this categorization system to code the tweets in my dataset, as detailed in section 3.4.3.

Cravens et al. (2015) looks at the tweets using *#whyistayed* and *#whyileft* from a clinician's perspective, and their analysis shows that there are a number of factors that affect both staying and leaving abusive relationships. Their study ultimately highlights how studying this campaign can be fruitful for therapists because of the pervasiveness of intimate partner violence. Clark

(2016) states, "In the case of #WhyIStayed, the hashtag acted as an easily personalized storytelling prompt, which provided a particular narrative focus for survivors to frame their diverse experiences in a compelling manner in 140 characters or less." This chapter builds on these works by using the short narrative form of these tweets and investigating which of the the reasons for staying and leaving abusive relationships motivates hesitation, which is realized as an initial hashtag placement in these tweets.

Previous studies of this campaign similarly look at this campaign because of the self-disclosures of violence, while this current study will look at how participants in this campaign situate their stories using linguistic devices, like hashtag placement.

3.1.2 Delaying Hashtags

This survivor corpus features more initial hashtags when compared to other a corpus of tweets that are not specific in topic. Many tweets in this corpus involve divulging sensitive, personal information that one may be reluctant to share. Examples of sharing personal information using initial *#whyistayed* and *#whyileft,* appear in (68)-(70):

(68) **#whyileft** he hit our 4 year old and shook her till she vomited.

(69) **#WhyIStayed** i thought the mindgames would end. i thought the mental abuse would stop. i thought if i loved her enough she'd forget the pain.

(70) **#WhyIStayed** He convinced me that he needed my help, and #WhenILeft he called the police on me because I asked him to sit

Tweets were more likely to feature an initial placement hashtag than in a general corpus (X-squared = 29.899, df = 1, p-value <.001), as mentioned in section 3.1. Some hashtags act as a

64

delay device before sensitive information, and the initial placement can be used in such a way to express hesitation in revealing that information. This function is most recently seen with the hashtags that became more prominent in 2014 with the domestic abuse awareness campaign *#whyIstayed* and *#whyIleft,* where people use these hashtags to not only situate their personal story into a larger campaign, but also as a means of divulging personal information after the hashtag.

3.2 Literature on delay devices

Some discourse markers are known for functioning as delay devices. For example, Buysse (2012) notes how the discourse marker *so* can be used with vowel lengthening to mark "hesitation or reflection" (2012:1770). According to Jucker (1993), *well* can function as a delay device; however, the author also notes how the function of delay device also overlaps with the function of face-threat mitigator. Since divulging personal information too quickly when it is not reciprocated can be face threatening (Merrigan 2000), the use of these two types of DMs in the survivor tweets would be expected because of divulging personal information. Furthermore, Wang (2011) explains how Japanese *ano* and Chinese *nage* appear before private or potentially embarrassing information when it functions to express hesitation in sharing personal information. Similarly, I have found that some hashtag placement acted as a delay device, especially in tweets that also served as survivor stories since each speaker is revealing their own traumatic experiences. DMs acting as delay devices, like *well* and *so* have been discussed in previous research; however, previous research fails to note how the placement of some hashtags function in a very similar way to these other markers.

3.3 Data and Methodology

For this study, disclosures via tweet were used to find delaying hashtags. Although Twitter data

can be found in already compiled corpora, compiling this survivor corpus allowed me to look at

particular tweets that were involved in the revelation of personal information.

3.3.1 Hesitations in Disclosure Corpus

One of the aims of this study in delaying hashtags was to look at the patterns that emerged

with beginning hashtags in a corpus of personal disclosures. While others have looked at

personal disclosures for other reasons, this particular study attempts to investigate how self-

disclosure tweets differ from other types of tweets. Han (2011) finds that discourse markers are

used frequently in public speeches, and because of this, we would expect discourse markers to

have a presence in prepared, public tweets, a similarly edited genre.

For this study in delaying hashtags, 1445 tweets were collected to see how this might differ from

the findings in Chapter 2 with spoken DMs. However, the tweets were limited to those that just

disclosed stories of surviving violence, using previous studies that looked at disclosure as my

guide. Harvey (2012) looked at the discourse of adolescent disclosure of depression by using a

corpus of anonymous emails from a medical website geared towards teens. Shaw (2007) instead

focused on just four participants. While my study is built on Shaw's findings, it approached the

participant pool in a similar manner to that of Harvey (2012) by taking an anonymous collection

of online text that is specified by the context, and instead have tweets from a random collection

of participants in this campaign. However, only 443 tweets out of the 1445 tweets actually involved self disclosure; many of the other tweets discussed the campaign as a whole.

The genre of survivor discourse was chosen to see what linguistic devices are employed when someone is divulging potentially intimate information to a larger, public audience. In (71), the speaker uses *#whyIleft* before talking about the horrors of what happened to her child:

(71) #whyileft he hit our 4 year old and shook her till she vomited.

One might imagine that if a friend were divulging this information, it would sound more like example (73), which begins with delaying discourse markers, than like (72), which is missing discourse markers:

(72) he hit our 4 year old and shook her till she vomited.
(73) so, um, well, he hit our 4 year old and shook her till she vomited.

This native intuition is supported by Romano (2014), who notes how people use a number of delaying devices (including discourse markers) in emotional narratives. Since the present study uses a corpus that features many people telling their emotional stories, we would expect to find many delay devices.

Because the data for this subset corpus consisted of emotional narratives that are of a somewhat public nature, speakers may want to avoid disclosing such personal information about their traumatic experience too directly. Because of this, we would expect to find a number of delay devices, and I found that these delay devices are realized as hashtags.

3.3.2 Corpus collection methods

Many popular activist campaigns use branded hashtags as a means of collectively talking about the same subject, and this is great for creating LSP corpora since many of the tweets will inherently be narrowed down by subject type. Twitter data is already written, and in an electronic form, but by compiling them into a Unicode version of a text file, one can easily use this data in concordance software. Also, this data is in English, which is a direct result of the fact that the initial query terms, these branded hashtags, are English clauses.

Since the 2014 domestic abuse awareness campaign focuses on the awareness, people used these hashtags to collectively talk about abuse, disclosing survivor stories on Twitter. Using the search function on Twitter, these survivor stories were collected by searching for campaign tweets using the hashtags *#whyIstayed* and *#whyIleft*. Also, this data is in English, which is a direct result of the fact that the initial query terms, these branded hashtags, are English clauses.

One method that many researchers use to collect Twitter data involves using the Twitter API. However, that would not work for this project since the API only retains tweets from the last 14 days. The search function on Twitter, though, allows searches from specific dates, so this method was used to collect this corpus.

Another problem that complicated data collection is many of the tweets using *#whyistayed* and *#whyileft* were not talking about survival or experiences with abuse. Instead, they were talking about the trendy hashtags, the campaign itself, or merchandise. Because of this, search terms

were collected that were associated with non-disclosures, and those were eliminated from the final data set. In the first part of data collection, the following terms were filtered using the search URL: *amazon, youtube, powerful, voices, campaign, trending, stories, magazine, editorial, conversation, yahoo, twitter, search, pizza, digiorno, ray, message, scan, applause, movement, awareness*, and *information*. In the second pass, using the conditional highlighting function of Excel, tweets with the following terms were removed from the final data set: *access hollywood, DrPhil, buffy, facebook, hashtags, forum, http, interview, Ray, Rice, Janay*, and *celebs*.

Using the extension for Google Chrome browser called Web Scraper, at most 20 tweets a day that used #whyIstayed and #whyIleft were collected. The limit of 20 was because the twitter search function automatically pulls up a maximum of 20 tweets.

After downloading Web Scraper, the extension was opened up and a new sitemap was created using the Twitter URL for the search of tweets for the first day of the campaign. From there, a new selector "each.tweet" was made in the "root", which gathered the entire tweet from the website on google chrome. In "each.tweet", two selectors, "tweet.text" and "Tweet.all", were created. The selector "tweet.text" only gathered the tweet content, while "tweet.text" collected the entire tweet. From there,"Browse" was selected under the "Sitemaps" tab in Web scraper to verify that the tweets were populating the columns correctly, then in the tab "Sitemap (twitter)," "Export data as CSV" was selected to make this file information into an analyzable file, using a unique name to save it.

The URL for twitter search for *#whyistayed* and *#whyileft* was inputted in Webscraper, and it took those tweets from the page and saved it into csv format. The problem with Webscraper is

that when it interacted with Twitter, it would only scrape up to 20 tweets at a time because 20

tweets is the automatic number of tweets that comes in the search, without further scrolling.

Even with more scrolling, Webscraper would only take the first 20 tweets because of Twitter's

automatic output.

Because it would only take 20 tweets at a time in Webscraper, other parameters, like date, had

to be changed in the input URL, else only 20 tweets would have been collected in this process.

Because of that, the URL had to be specific with dates, so at most 20 tweets per day could be

collected. So the first URL input on Webscraper found tweets using *#whyIstayed* OR *#whyileft*

specified only tweets on September 8, 2014. After saving this CSV file, "Edit the metadata" was

selected to change search parameters in the URL to the next day. After doing this, "Browse" was

selected under the "Sitemaps" tab in Web scraper to verify that the tweets were populating the

columns correctly. Again, in the tab "Sitemap (twitter)," "Export data as CSV" was selected, and

this file was saved using a unique name. These steps were repeated several times, until 73 days

worth of tweets were gathered. After day 73, there was a drop in the number of tweets that

used these hashtags.

Not all the tweets for the campaign disclosed information about surviving abuse. In fact, many

tweets were comprised of people talking about the campaign itself or using this trending hashtag

without context for the larger campaign. In order to filter these out during the initial search and

collection via Web scraper, the following terms were filtered using the search URL: amazon,

youtube, powerful, voices, campaign, trending, stories, magazine, editorial, conversation, yahoo,

twitter, search, pizza, digiorno, ray, message, scan, applause, movement, awareness, and

information.

There were 1445 tweets collected over 73 days, beginning on the first day of the campaign

(September 8, 2014), and stopping the collection of tweets when the campaign started to

dwindle in number (November 19, 2014). Many disclosure tweets were gathered each day, up to

20. There was an average of 18.15 tweets collected a day, with a minimum of 6 and a maximum

of 20.

These 1445 tweets were not all disclosures of abuse, despite the search parameters, so in Excel,

conditional highlighting was used to create rules to highlight those tweets that had the following

terms: access hollywood, DrPhil, buffy, facebook, hashtags, forum, http, interview, Ray, Rice,

Janay, and celebs. Using the sort function, I moved those highlighted from the previous set to

the bottom of the CSV file. From there, as I organized and analyzed the data, I manually

highlighted the tweets that were not disclosures.

After narrowing the data to just disclosures, using the methodology described above, there were

443 tweets over 73 days.

3.3.3 Description of the data

This corpus consists of tweets collected from September 8, 2014 until November 19, 2014, which

is 73 days. The average number of disclosure tweets per day was 6.08, with as many as 16 (three

days) and minimum of zero (one day). These campaign tweets can be split into two categories:

phase 1 and phase 2. The density of tweets in this corpus deviate significantly from the

normality according to the Shapiro-Wilk test ($p < .001$), and the histogram displaying this can be

viewed in Figure 3-3 below:

Histogram of tweet$day
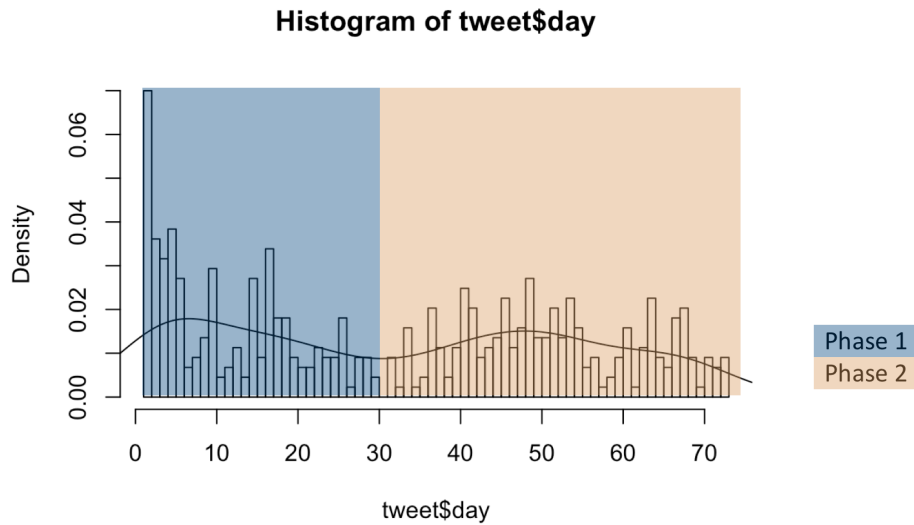
Figure 3- 3: Density of Tweets Per Day

As noted in Figure 3-3 , there were two surges in campaign participation. On day 31 (October 8, 2014), there were zero disclosures collected. While there were 20 tweets initially collected, all tweets for this day were non-disclosures that made it through the query despite the initial parameters. This coincided with an article published online on the abuse relationship between Buffy and Riley, characters from the television show *Buffy the Vampire Slayer.*The article compares the abusive nature of Buffy's relationship with Riley in the episodes that originally aired in 1999-2002 to the events with Ray and Janay Rice that sparked discussion for this 2014 campaign. This resurgence of the topic in online discussions seems to have motivated the surge in phase 2.

3.3.4 Hashtag Placement

Hashtags in this corpus are categorized as being in one of the three following placement areas: initial, medial, and final. Initial hashtags appear in a clause-initial position in relation to the associated tweet content. An example of this placement is *#WhyILeft* is in example (74).

> (74) **#WhyILeft** I knew, that no matter how perfect I was. It would never be enough to stop him. I had to take care of me.

Medial hashtags appear in the clause, and the content of the hashtag is essential for the understanding of the clause as a whole. An example of this is *#adopted* in (75).

> (75) #WhyILeft Because being treated inferior because you were **#adopted** is NOT okay. #respect

Final hashtags appear in a clause-final position, as shown with the hashtags *#depressed*, *#crying*, and *#WhyIStayed* in (76).

> (76) Sometimes I don't understand why...why they treat me like this..I did sth wrong? **#depressed #crying #WhyIStayed**

Some hashtags appear in the middle of a tweet, but they are characterized as either clause-initial or -final hashtags, as in (77) and (78), respectively:

(77) **#WhyIStayed**: bc I was just a kid & "he's your daddy no matter what so you have to love him". **#WhyILeft:** Love doesn't hurt. #breakthesilence

(78) The promises felt good, even if I knew it was bad. #WhyIStayed My life didn't feel worth living anymore. #whyileft

In (7), *#WhyILeft* is part of the second clause of the tweet, and in relation to that content, it is in the initial placement. In (8), *#WhyIStayed* is part of the first clause of the tweet, and in relation to that content, it is in the final placement.

Using a corpus of tweets from the 2014 Domestic Abuse Awareness campaign on Twitter that *#whyistayed* or *#whyIleft*, I was able to note similarities between the hashtag placement and how other discourse markers are used as a delay device.

Since initial hashtag placement is the hesitation placement, Table 4-1 shows the frequency of initial hashtags in both the survivor tweet corpus and a 2012 tweet corpus. The 2012 corpus consisted of 1,793 tweets where 463 tweets (25.82%) begin with an initial hashtag. This is significantly different from the pattern shown in the survivor tweet corpus: 173 tweets out of 443 (39.05%) began with an initial hashtag, as shown in Table 3-2:

Table 3- 2: Comparison of initial hashtags in survivor corpus to general corpus

	Yes Initial	No Initial	Total
survivor tweet corpus	173 (39.1%)	270 (60.9%)	443 (100%)
2012 general corpus	463 (25.8%)	1330 (74.2%)	1793 (100%)

It is statistically unlikely that there would be 173 tweets that begin with initial hashtags in this corpus using the general tweet corpus for comparison ($p < .001$). I postulate that this is a significant difference in probabilities of hashtag placement is due to the function of the hashtags in the survivor tweet corpus: some hashtags in the initial position act as delay devices when appearing before disclosing sensitive information.

Because there are statistically more tweets that feature an initial hashtag in a corpus of disclosures, initial placement of hashtags is doing more than just connecting these tweets to the larger conversation on domestic abuse: they are acting as a delay device before revealing details of abuse akin to other discourse markers like *well* (Jucker 1993), Japanese *ano* (Wang 2011), and Chinese *nage* (Wang 2011).

3.3.5 Phases of the Campaign

Studying the frequency of tweets that use these hashtags showed that there are two peaks to the campaign, which was divided into two phases of the campaign. Phase 1, as shown earlier in Figure 3-3, consists of the first 30 days of the campaign. Phase 2 consists of days 31 through 73.
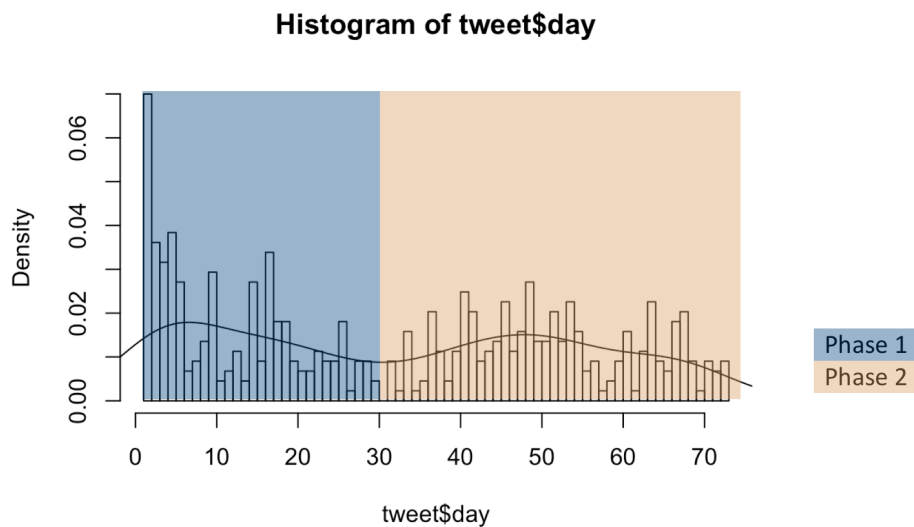
Figure 3- 3: Density of Tweets Per Day

While the two peaks in usage motivated this campaign to be split into two phases, the two phases were not significantly different in size. There were more tweets in phase 2 (227) of the campaign than phase 1 (216), but this difference is not significant (X-squared = 0.27314, df = 1, p-value = 0.60), so this should not affect the results when looking at the interaction between phases and the variables explored in section 3.4.

As previously mentioned, an article about *Buffy the Vampire Slayer* published on day 31 (October 8, 2014), coincided with the rise in disclosure tweets after that day, so this served as the break between phases 1 and 2. I will use phases of the campaign to show content differences that arise in the campaign over time in relation to initial hashtag placement.

There were 97 tweets (44.9%) that featured an initial hashtag before the disclosure in phase 1, whereas there were 75 tweets (33.0%) that had an initial hashtag in phase 2 before the disclosure.

Tweets that feature initial hashtag depending on phase

Phase 1 — YES 97 / NO 119

Phase 2 — YES 75 / NO 152
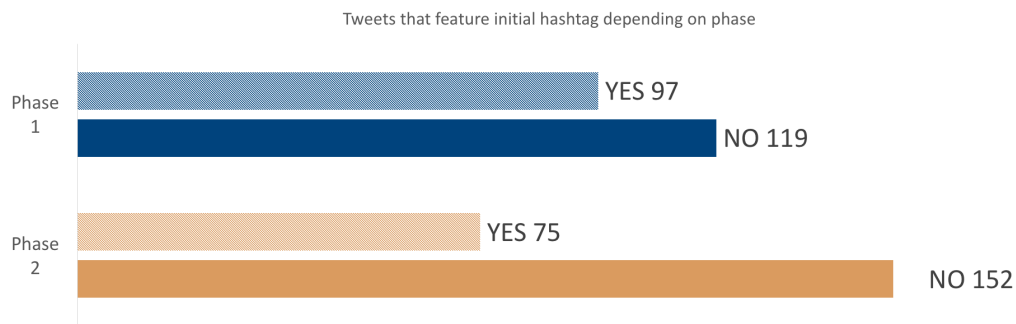
Figure 3- 4: Chart of tweets that feature initial hashtag depending on phase

Disclosure tweets were more likely to begin with Initial hashtags are dependent on the phase of the campaign (X-squared = 6.0732, df = 1, p-value = 0.01373).

This means people were more likely to use the delay device of initial placement in the first phase of the campaign. When the online conversation on domestic abuse was still rather new, people felt the need to hesitate more. Once the second phase of the campaign began, people did not use the initial hashtag as often--- they did not feel the need to hesitate as much after so many people talked about their abuse during the first phase.

3.4 Analysis

In 3.4.1, I show that this corpus has an unusually large amount of tweets with a tweet initial hashtag when compared to a general corpus, and in sections 3.4.2 and 3.4.3, I show what factors like the phases of campaign and the actual hashtag being used affects the hashtag placement. In 3.4.4 and 3.4.5 analyze the content of the tweet and related factors and initial hashtag placement, particularly with respect to mention of abuser and themes of the campaign.

In the next section, we will see how this delaying function changes over time in the campaign.

3.4.1 Left or Stayed

One variable that may affect the initial hashtag placement that is involved with content of the tweet is the content of the hashtag itself. There were 254 tweets (57.3%) in this corpus that used only the hashtag #whyistayed. There were 132 tweets (29.8%) that used only #whyileft. There were 57 (12.9%) of the tweets that used both #whyistayed and #whyileft.

The following table shows the number of initial placement hashtags according to the hashtag used in the tweet. Of the 311 tweets that used #whyistayed (254 only using one hashtag and 57

using both), 139 (44.7%) tweets began with an initial placement hashtag, whereas 55% did not

have an initial hashtag. Only 67 (35.4%) of 189 tweets that used *#whyileft* (132 tweets with

only one hashtag and 57 with both) began with an initial placement hashtag, but 64.6% did not

have initial hashtag placement.

Table 3- 3: Initial hashtag placement depending on #whyistayed or #whyileft

	Yes Initial	No Initial	Total
#whyistayed	139 (44.69%)	172 (55.31%)	311 (100%)
#whyileft	67 (35.45%)	122 (64.55%)	189 (100%)

According to the chi-squared test for independence, the use of the initial placement hashtag (the

dependent variable) is dependent on presence of *#whyistayed* (the independent variable) (X-

squared = 13.178, df = 1, p-value = 0.0002833). However, the use of the initial placement

hashtag is not dependent on presence of *#whyileft* (X-squared = 1.5428, df = 1, p-value =

0.2142).

These results indicate that this initial placement of the hashtag as a means of delay was used

more often when people talk about why they stayed in the unhealthy relationship. However,

when the person talked about why they left, they did not use this delay device as much.

However, this is complicated by the fact that the use of *#whyistayed* and *#whyileft* varied over the course of the campaign, as shown in Tables 3-4, 3-5, and 3-6. In the 216 tweets from phase 1, there were 139 (64%) tweets with *#whyistayed*, whereas there are 51 (23.61%) with *#whyileft* and 26 (12.04%) with both. In phase 2 of the campaign, there were 227 tweets total, with 115 (50.66%) of them having *#whyistayed,* 81 (35.68%) with *#whyileft*, and 31 (13.66%) with both hashtags.

The percentage of initial hashtag placement in the general corpus is 25.8%, so anything more than that is still unusual. In table 3-4, there is a larger percentage of initial hashtag placement for tweets with just #whyistayed in the first phase of the campaign, compared to the second, but this percentage only differs by 3.2%. Phase 1 with 43.2% and Phase 2 with 40.0% are both higher than the general corpus percentage for initial hashtag placement and higher than the percentage found for all the survivor tweets, which was 39.1% with initial hashtag placement.

Table 3- 4: Tweets with only #whyistayed, depending on the campaign phase

#whyistayed	Yes Initial	No Initial	Total
Phase 1	60 (43.2%)	79 (56.8%)	139 (100%)
Phase 2	46 (40.0%)	69 (60.0%)	115 (100%)

There is a different pattern in initial hashtag placement in tweets that use #whyileft. In table 3-5, there is a larger percentage of initial hashtag placement for tweets with just #whyileft in the

first phase of the campaign, compared to the second phase. Phase 1 with 39.2% has a greater percentage than that found in the general corpus (25.8% of initial hashtag placement in tweets) and the survivor corpus (39.1% of initial hashtag placement in tweets). Phase 2 with 17.3% is much less than the general corpus and the survivor corpus.

Table 3- 5: Tweets with only #whyileft, depending on the campaign phase

#whyileft	Yes Initial	No Initial	Total
Phase 1	20 (39.2%)	31 (60.8%)	51 (100%)
Phase 2	14 (17.3%)	67 (82.7%)	81 (100%)

When the tweet uses both hashtags, there was 73.1% of tweets in the first phase of the campaign with initial hashtag placement, as shown in Table 3-6. In Phase 2, 45.2% of tweets had initial hashtag placement Both of these percentages are greater than the general and survivor corpus.

Table 3- 6: Tweets with both hashtags, depending on the phase of the campaign

Both hashtags	Yes Initial	No Initial	Total
Phase 1	19 (73.1%)	7 (26.9%)	26 (100%)
Phase 2	14 (45.2%)	17 (54.8%)	31 (100%)

The use of initial hashtags is dependent on the use of *#whyistayed* (X-squared = 13.178, df = 1, p-value < 0.001). The use of initial hashtags is not found to be dependent on the use of *#whyileft* (X-squared = 1.5428, df = 1, p-value = 0.2142).

In short, people were not only more likely to use initial hashtags in phase 1 of the campaign, but they were also more likely to use initial hashtags when using the hashtag *#whyistayed*.

Since *#whyistayed* and *#whyileft* have different content involved, in the next section I will investigate what content might motivate the initial hashtag placement.

3.4.2 Mention of the Abuser

Because disclosure of private information seems to be relevant, another aspect to examine is what happens if the abuser is mentioned. There were 245 of the 443 tweets that mention the abuser, like in (79)-(82), using different referential expressions:

(79) #WhyIStayed **he** burned all my other bridges, I was afraid I would be alone

(80) #WhyIStayed because you get ground down living with **someone** who threatens you, screams at you & gaslights you.

(81) Dear **X** #TruthIs -I was so focused on the PROMISE I saw in **you** I refused to see the PERSON you are. #WhyILeft #MyTime #GodsPromise #ImFree

(82) #WhyIStayed married for 7.5 yrs, 6 abusive. It went from **him** choking me, holding me down to a gun in my face everyday for 2 yrs.

People were more likely to tweet about their abuser in phase 1 of the campaign than phase 2, as shown in Table 3-7.

Table 3- 7: Abuser mention according to phase of campaign

	Phase 1	Phase 2	Total
Abuser Mentioned	140 (64.8%)	105 (46.3%)	245 (100%)
No Abuser Mentioned	76 (38.4%)	122 (61.6%)	198 (100%)

There was an interaction of variables between mention of abuser and phase of campaign (X-squared = 14.682, df = 1, p-value = 0.0001273).

While mentioning the abuser is dependent on the phase of the campaign, the initial hashtag placement is not dependent on the mention of the abuser (X-squared = 1.7861, df = 1, p-value = 0.1814), as shown in Table 3-8:

Table 3- 8: Abuser mention and initial hashtag placement

	Initial placement	No initial placement	Total
Abuser Mentioned	103 (42.0%)	142 (58.0 %)	245 (100%)
No Abuser Mentioned	70 (35.4%)	128 (64.6%)	198 (100%)

If we look at just the tweets with initial hashtag placement, however, there are a significantly larger number of these that mention abuser than those that do not (X-squared = 6.2948, df = 1, p-value < 0.05).

Also, the hashtag used for the tweet did not depend on the mention of abuser, for both #whyistayed (X-squared = 0.011013, df = 1, p-value = 0.9164) and #whyileft (X-squared = 0.035419, df = 1, p-value = 0.8507).

There were significantly more tweets that mentioned the abuser in tweets with #whyistayed (X-squared = 3.9389, df = 1, p-value < 0.05), but while there are still more tweets that mention abuse in tweets with #whyileft, there is not significantly more (X-squared = 2.7989, df = 1, p-value = 0.09)

Mentioning the abuser was not dependent on the use of #whyileft (X-squared = 0.035419, df = 1, p-value = 0.8507) nor was it dependent on the use of *#whyistayed* (X-squared = 0.011013, df = 1, p-value = 0.9164).

While mentioning the abuser was found to be independent on the use of *left* or *stayed* in the hashtag, there were statistically more tweets that mentioned the abuser when looking at tweets with the initial hashtag placement. In the next section, I use the themes of the *#WhyIStayed* campaign to see if the way they talk about the abuser might be motivating more initial hashtag placement.

3.4.3 Themes of staying in an abusive relationship

According to the system proposed by Weathers et al. (2016), there are themes of topics in the first day of the campaign, including: "(lack of) resources, responsibility for abuse, fear, and gender-linked power," where gender-linked power is split into power-to and power-over. In this section, I will show how the tweets in this corpus fit into these five themes, and how some themes may be more involved in others in driving the use of initial placement hashtags as a delay device. Since the general corpus consisted of 25% of initial hashtag placement, any higher percentages for initial hashtag placement for different themes indicates that Twitter users are using initial hashtag placement more often for different topics. Particularly, 50% of tweets in the survivor tweet corpus that involved fear used initial hashtags. These tweets were coded using the descriptions in sections 3.4.3.1 - 3.4.3.4, and 164 tweets were categorized as having multiple themes. Only 22 tweets failed to fit neatly into the four categories set out in the previous work, so these were labeled as being in the ambiguous category.

3.4.3.1 Lack of Resources

Using the description from Weathers et al. (2016), disclosure tweets were categorized as "lack of resources" if it involved insufficiencies, including internal and external needs. External needs involves finances, housing, etc, as shown in (83).

> (83) #WhyIStayed I moved to another state when he changed jobs. I had no money. I had nowhere to go.

Internal needs involved social support, like in (84), where the tweeter talks about not having "emotional support".

> (84) Because he was my emotional support and I made me feel like I owed him #WhyIStayed

One form of social support that was lacking for many tweeters dealt with religion, and the lack of support from the church and church personnel, as in (85) and (86).

> (85) because I was supposed to be "good christian wife" #WhyIStayed

> (86) Because the pastor said I didn't know "one iota about a man's heart" and the marriage counselor thought I could "help" my abuser #WhyIStayed

But the lack of social support could come from other social constructs, like the law:

(87) #WhyIStayed Because my ex is a cop and he always told me that he would take my kids away..

(88) #whyistayed he threatened me into not calling cops by sayn he wold stab himself and frame me for it

Others discussed the lack of resources for their children, like in (22).

(89) Because my children were young and I was unemployed. #WhyIStayed #domesticviolence @HouseOfRuthMd

Finally, tweeters could also lack the mental resources to understand that they did not deserve the violence (90), or the knowledge to know their experience was indeed abuse, as in (91) and (92).

(90) I didn't know I deserved better #WhyIStayed

(91) #WhyIStayed He didn't hit me and push me that often, and I didn't know there were other forms of abuse #DomesticViolence #KnowTheSigns

(92) #whyIstayed I had a Ph.D. and smart women don't get abused ...

There were 218 tweets that were categorized in the *lack resources* theme, and 91 (41.74%) of them had an initial hashtag. There were 173 tweets with initial placement hashtags, and 52.60% of them were categorized in the *lack resources* theme.

Table 3- 9: Initial placement and lack resources theme

	Initial placement	No initial placement	Total
Lack resources	91 (41.7%)	127 (58.3%)	218 (100%)
Different theme	82 (36.4%)	143 (63.6%)	225 (100%)

Initial hashtag placement was not found to be dependent on the lack of resources theme (X-squared = 1.0929, df = 1, p-value = 0.2958). However, there is a greater percentage of initial hashtag placement with the theme of lack resources than in the other four themes.

3.4.3.2 Responsibility for abuse

Tweets with the theme of responsibility for abuse discussed the blame for the abuse in some way, including "a variety of reasons, including male biology, difficult child- hoods, sports careers, and military service" (Weathers et al. 2016), as shown in (93).

> (93) Abused as a child, I thought he just needed someone to love him unconditionally. I thought I could fix him. Married almost 9 yrs #WhyIStayed

Also to blame in many tweets is the survivor, as shown in (94) and (95).

(94) #WhyIStayed I thought I provoked/caused the beatings

(95) He pushed me down on the bed, squeezing my arms, screaming in my face "You make me do this!" I believed him. #WhyIStayed @WhyIStayed

There were 35 tweets in this corpus that fit into this category, 10 of which had an initial hashtag.

Table 3- 10: Initial placement and responsibility of abuse theme

	Initial placement	No initial placement	Total
Responsibility for abuse	10 (28.6%)	25 (71.4%)	35 (100%)
Different theme	143 (35.0%)	245 (65.0%)	408 (100%)

As with the first theme, initial placement was not found to be dependent on the theme of responsibility of abuse. However, unlike the theme of lack resources, the percentage of initial hashtag placement for the theme of responsibility for abuse is less than the other four themes combined, as shown in the last row of Table 3-10, and 28.6% is only a few percentage points more than the average of 25% of initial hashtag placement in the general corpus.

3.4.3.3 Fear

Some tweets in this campaign talked about fear and shame as a motivating reason for staying in or leaving their abusive relationship (Weathers et al. 2016). Tweets that were categorized in this way could be talking about general fear explicitly, as in (29) and (30):

(96) #WhyIStayed fear for my life #WhyILeft I had no life

(97) Because I was afraid! #WhyIStayed #domesticviolence @HouseOfRuthMd

but they could also talk about fear in a covert way, as in (31).

(98) #WhyILeft I saw my face in the floor surrounded by crystals and see my self into 3 future places; hospital, psychiatric or cementery [sic]

The fear of shame and embarrassment was also categorized with this theme, like in (99).

(99) I was so scared and embarrassed to have to tell people why we weren't together anymore #WhyIStayed

There were 75 tweets that had the fear theme, and 37 (49.33%) had an initial placement hashtag. This is almost half of the tweets with the theme of fear used in initial hashtag placement.

Table 3- 11: Initial placement and fear theme

	Initial placement	No initial placement	Total
Fear	37 (49.3%)	38 (50.7%)	75 (100%)
Different theme	136 (37.0%)	232 (63.0%)	368(100%)

According to the chi-squared test for independence, these initial hashtag placement was approaching significance with dependence on the theme of fear (p=.06), unlike the other categories which were not even close to significant. This percentage is much larger than the percentage of the other themes combined, as shown through Table 3-11.

3.4.3.4 Gender-linked power

Gender-linked power, as written in Weathers et al. (2016) is based on the power dynamic ideas in Yoder & Kahn (1992). In this older work, it defines the gendered differences of power, where women are more likely to have *power-to* change how one "feels over one's own thoughts, feelings, and behaviors" whereas men are more likely to have *power-over* and control of another person (Yoder & Kahn 1992). In this dissertation, I categorized tweets as having gender-linked power if it was coded as being power-over or power-to.

Tweets were coded as having the power-to theme when it talked about controlling feelings, like

the phrase "I made me feel" in (100), "I realized" in (101), and "I was in denial" in (102).

> (100) Because he was my emotional support and **I made me feel** like I owed him #WhyIStayed

> (101**) I realized** love doesn't settle, and neither should I #WhyILeft

> (102**) I was in denial** of what was starting to happen. #WhyIStayed

Tweets with power-over themes were coded as such when they were talking about controlling

and manipulating people, as in (103) and (104).

> (103) He emotionally manipulated me into feeling guilty for wanting to leave him. #WhyIStayed

> (104) #WhyIStayed he burned all my other bridges, I was afraid I would be alone

There were 268 tweets that are coded as either having power-to, power-over, or both, and 105 of

these featured an initial placement hashtag, with 39.2%. This is more than we see in just the

power-to tweets, as shown in Table 3-12, where power-to has 36.4% of the tweets with initial

hashtag placement.

Table 3- 12: Gender-linked power and initial hashtag placement

	Initial placement	No initial placement	Total
Power-to	56 (36.4%)	98 (63.6%)	154 (100%)
Different theme	117 (40.5%)	172 (59.5%)	289 (100%)

The percentage of initial hashtag placement for power-over tweets is greater than power-to and the all of the gender-linked themed tweets, as shown in Table 3-13.

Table 3- 13: Gender-linked power and initial hashtag placement

	Initial placement	No initial placement	Total
Power-over	62 (40.5%)	91 (59.5%)	153 (100%)
Different theme	111 (38.3%)	179 (61.7%)	290 (100%)

Once again, initial hashtag placement was not dependent on gender-linked themes, but both percentages of initial hashtag placement for the gender linked theme were greater than the percentage of initial hashtag placement found in the general corpus.

There was a greater percentage of initial hashtag placement in power-over than the percentage of initial hashtag placement in the survivor corpus, while power-to was less than the survivor corpus in general and power-to theme.

3.4.3.5 Summary of Themes and Placement

Having looked at five themes as variables with the tweet's content that might be motivating the initial hashtag placement, I show that the theme that featured the largest percentage of initial hashtag placement was fear, as shown in the fourth row in Table 3-14. Almost half of the tweets that fit into the category of fear featured an initial hashtag placement.

Weathers et al. (2016) did not indicate which of these themes were most frequent. By using these themes as categories for content, I was able to not only see what most of the people were talking about when they used these hashtags (lack of resources was the most tweeted theme, as shown in row 1 of Table 3-14), but also which tweets prompted more initial hashtag placement (fear had the highest percentage of initial hashtag placement, as shown in row 4 of Table 3-14). Not as many people tweeted about responsibility of abuse since Table 3-13 shows this category has the fewest tweets in this corpus. Furthermore, this percentage of initial hashtag placement for responsibility of abuse was only 28.6%, as shown in row 5 of Table 3-14, which is only a little more than the percentage for the general corpus, and much less than the 39.1% of initial hashtag placement found in the entire survivor corpus.

Table 3- 14: Tweet themes and initial hashtag placement

	Initial hashtag placement	No initial hashtag placement	Total
Lack of resources	91 (41.7%)	127 (58.3%)	218 (100%)
Gender-linked power-to	56 (36.4%)	98 (63.6%)	154 (100%)
Gender-linked power-over	62 (40.5%)	91 (59.5%)	153 (100%)
Fear	37 (49.3%)	38 (50.7%)	75 (100%)
Responsibility of abuse	10 (28.6%)	25 (71.4%)	35 (100%)
Ambiguous themes	6 (27.3%)	16 (72.7%)	22 (100%)
Multiple themes	71 (43.3%)	93 (56.7%)	164 (100%)
Total	173 (39.1%)	270 (60.9%)	443 (100%)

In the next section, I will define these categories and discuss the significance (or lack thereof) of these themes in relation to using initial hashtag placement as a delay device.

3.4.3.6 Themes and Phases

Since there were more tweets with initial hashtag placement in Phase 1 than Phase 2 of the campaign (see Section *3.3.5)*, I investigate what themes in tweet content are more prevalent during the different phases. The themes of lack of resources and responsibility of abuse both have more tweets in Phase 1. The themes of fear and gender-linked power-to feature appear more often in Phase 2. Half of the tweets with the theme gender-linked power-over appeared in Phase 1. Investigating the themes in relation to phase of the campaign can reveal what content was more prevalent in Phase 1, when there were more initial hashtag placement, and what content categories are more prevalent in Phase 2, when there was less tweeting.

There were more lack of resources tweets in Phase 1 (119 tweets) than in Phase 2 (99 tweets), as shown in Table 3-15.

Table 3- 15: Phase of the campaign and lack resources theme

	Phase 1	Phase 2
Lack resources	119	99
Different theme	97	128

The theme of lack resources is dependent on the phase of the campaign (X-squared = 5.3861, df = 1, p-value <0.05).

Similarly, there were more responsibility of abuse tweets in Phase 1 (24 tweets) than in Phase 2 (11 tweets), as shown in Table 3-16.

Table 3- 16: Phase of the campaign and responsibility of abuse theme

	Phase 1	Phase 2
Responsibility of abuse	24	11
Different theme	192	216

The theme of responsibility of abuse is dependent on the phase of the campaign (X-squared = 5.1409, df = 1, p-value <0.05). Both lack resources and responsibility of abuse were tweeted about more during phase 1 of the campaign, the phase where there were more delaying initial hashtag placement.

Since fear was the theme in last section that has the higher percentage of initial hashtag placement, I expected there to be more tweets in Phase 1, but only 29% of these tweets were in Phase 1. There were more fear tweets in Phase 2 (53 tweets) than in Phase 1 (22 tweets), as shown in Table 3-17. The theme of fear is dependent on the phase of the campaign (X-squared = 12.716, df = 1, p-value < 0.001).

Table 3- 17: Phase of the campaign and responsibility of abuse theme

	Phase 1	Phase 2
Fear	22	53
Different theme	194	174

Another theme that was more prevalent in Phase 2 of the campaign was power-to. There were more power-to tweets in Phase 2 (99 tweets) than in Phase 1 (55 tweets), as shown in Table 3-18.

Table 3- 18: Phase of the campaign and power-to theme

	Phase 1	Phase 2
Power-to	55	99
Different theme	161	128

The theme of power-to is dependent on the phase of the campaign (X-squared = 15.286, df = 1, p-value<.0001).

This is different from the pattern in gender-linked power-over. The number of tweets in Phase 1

and Phase 2 are evenly split, as shown in Table 3-18.

Table 3- 19: Phase of the campaign and power-over theme

	Phase 1	Phase 2
Power-over	77	76
Different theme	194	174

The theme of power-over is not dependent on the phase of the campaign (X-squared = 0.14419,

df = 1, p-value = 0.7041), and this is odd since the other four variables are dependent on phase.

The results of these inquiries are shown in Table 3-20. As the first two rows of Table 3-20 show,

lack of resources and responsibility of abuse were higher in Phase 1, while fear and Gender-

linked power-to were higher in Phase 2. Gender-linked power-over were about the same in both

phase 1 and phrase 2.

Table 3- 20: Tweet themes and initial hashtag placement

	Phase 1	Phase 2	Total
Lack of resources	119	99	218
Gender-linked power-to	55	99	154
Gender-linked power-over	77	76	153
Fear	22	53	75
Responsibility of abuse	24	11	35
Ambiguous themes	17	5	22
Multiple themes	73	91	164
Total	216	227	443

The impact of these difference is seen by using a chi-squared test: The themes that were more popular in Phase 1 were lack of resources (X-squared = 5.3861, df = 1, p-value < 0.05) and responsibility of abuse (X-squared = 5.1409, df = 1, p-value < 0.05). The themes that were

popular in Phase 2 are fear (X-squared = 12.716, df = 1, p-value < 0.001) and gender-linked power-to (X-squared = 15.286, df = 1, p-value<.0001). Gender linked power-over was the only category that was not significantly different in each phase. Four of the five variables are dependent on the phase of the campaign.

3.5 Conclusions

This chapter shows that hashtags in the initial placement act in a way similar to the spoken discourse markers in Chapter 2. Delaying hashtags are similar to delaying discourse markers that are present in spoken language, further solidifying the similar form and function of some DM-like hashtags and DMs.

In this chapter, initial hashtag placement is not dependent on the mention of abuser, but there is a large percentage of tweets that mention abuser that also feature an initial hashtag placement.

Another finding from this inquiry was that there were significantly more tweets that mentioned the abuser in tweets with *#whyistayed* (X-squared = 3.9389, df = 1, p-value < 0.05), but not significantly more with tweets that mention abuse in tweets with *#whyileft* (X-squared = 2.7989, df = 1, p-value = 0.09).

Initial hashtag placement was found not to be dependent on the themes discussed in Weathers et al. (2016); however, we see that when the tweet showed up in the campaign (either phase 1 or phase 2) it is dependent on theme for all of the categories except for gender-linked power-over.

Analysis and comparison of this topic specific twitter corpus to a general corpus shows that discourse marker hashtags are acting as a hesitation in this corpus of survivor tweets. Hashtags with an early placement are significantly more common in the disclosure corpus than they are in the large corpus of tweets. Because of their placement in front of some personal revelation, it appears that these hashtags are acting a delay device.

There were more initial hashtag placement used in the first phase of the campaign, which means more people hesitated sharing personal information when the disclosure conversations were just beginning.

The variables of initial hashtag placement is dependent on when in the campaign it appeared, with more initial hashtag placement in the first phase of the campaign. This means people used this initial hashtag placement as a delay device when the conversations about abuse were just beginning.

Initial hashtag placement is a delay device for avoiding abruptly conveying stories of violence and horror. Tweeters are using the initial placement to talk about such personal information in the same way others use discourse markers in face-to-face disclosures, especially those conveying stories of surviving violence. Because of the character constraints of tweets, tweeters do not have the space to delay in the same way as one would in other disclosures; instead, initial placement is an innovative way to delay while joining in the larger conversation.

CHAPTER 4

TAG REFRAMING

4.1 Introduction

The first function of hashtags that I will examine is tag reframing. Tag reframing refers to enhancing the meaning of the tweet, usually by clarifying the intended message in the tweet or contextualizing the tweet's message. While previous chapters have linked the form and function of some hashtags and DMs, this chapter will investigate how some hashtags act in similar ways to discourse markers, but their written nature differentiates them in a remarkable way from spoken DMs.

The pragmatic role of clarifying the message in computer mediated language aids in avoiding miscommunication online, especially since some tweets are written in an ambiguous way without the hashtag. In Figure 4-1, Paul Ryan, under the username @SpeakerRyan, tweeted about the Manchester tragedy, using the trending hashtag #Manchester. The tweeter @IrisRimon shines light on the ambiguity of Ryan's original tweet content without the hashtag. The tag reframing flavors the message of this tweet by adding context to what tragedy he is talking about. If Paul Ryan would have instead ended it in #budgetcuts, this tweet would mean something very different.

Figure 4- 1: Tweet by Paul Ryan

In this chapter, I investigate the role of hashtags that act in a similar fashion to *#Manchester* in

Figure 4-1, where the hashtag can clarify the message or even change the meaning of the tweet,

a function I will call *tag reframing*.

Some examples of intentionally vague online messages include subtweeting, which is tweeting

something negative about another person indirectly (Edwards and Harris 2016), and

vaguebooking, which is a similarly passive-aggressive Facebook message that may or may not be

directed towards someone (Child and Starcher 2016). An example of subtweeting is in Figure 4-

2, where the referential expressions *you* and *yourself* do not have a known referent.
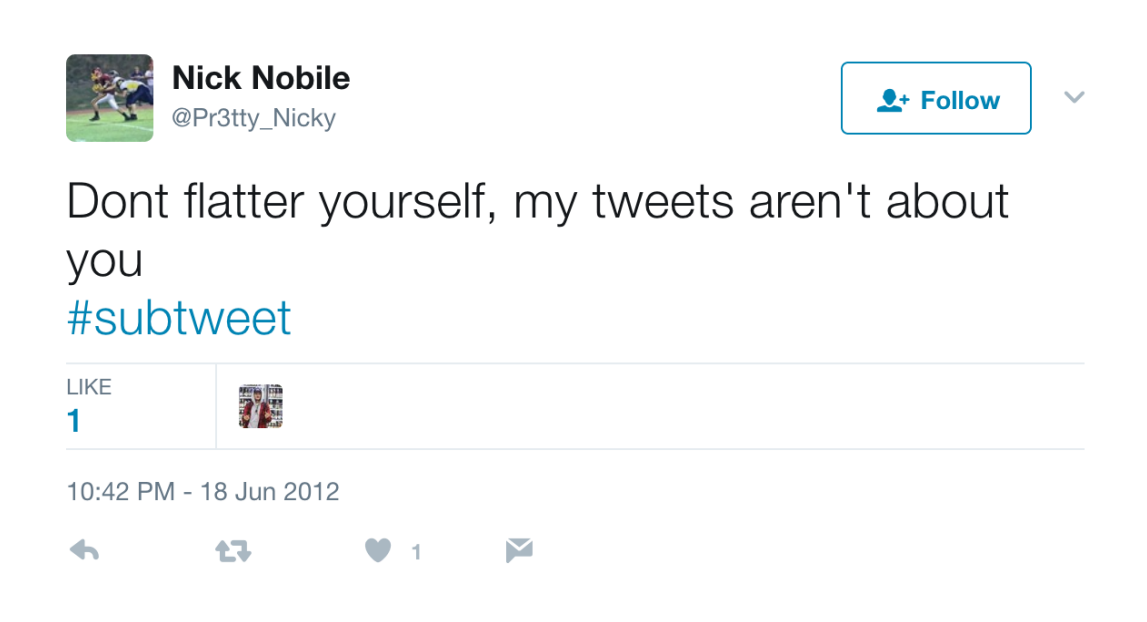
Figure 4- 2: Subtweet

In Figure 4-2, the user @Pr3tty_Nicky denies subtweeting about the unknown second person in a previous tweet, but in this tweet, since the referent of referential expressions *you* and *yourself* are unknown, this user is actually subtweeting. The hashtag *#subtweet* is specifying why the second person should not flatter himself or herself- using *#subtweet*, it is saying that this message is directed at someone specific. The tweet in Figure 4-2 is a subtweet with or without the hashtag since the referential expressions *you* and *yourself* are still ambiguous, but the hashtag is clarifying the message further, ensuring the intended audience understands the meaning of the tweet.

The tweet does not have to be purposefully ambiguous for the hashtag to aid with understanding of the tweet content. In fact, Blattner et al. (2016) describes these types of hashtags as, "a summary of the larger tweet, a modern-day, electronic 'cliff-notes'". In fact, these hashtags clarify the message of the tweet so well, Blattner et al. notes the hashtag helped second language learners understand the tweet content without being fluent in their L2.

When someone adds information that adds clarifying information, like context or specific details about a noun phrase, this is called repair. Tree and Schrock (1999) discuss this by stating, "A repair entails a change of state from the original version to the corrected version." Pragmatic repair is not fixing some wrong information, but strengthening the information to ensure felicitous communication. Clark and Marshall (1981) mention that speakers repair definite references "to make them more likely to succeed" where the hearer will better understand the speaker given the need for mutual knowledge. DMs can signal repair, this clarifying information or repair of information, as shown in (105) and (106) when *oh* and *I mean* introduce this clarification.

> (105) And then he cited four presidents, um Jefferson Eisenhower Kennedy, and uh someone else for um. . . for being adulterers, but- but being very great presidents, **oh** and Roosevelt. (Tree and Schrock 1999)

> (106) she appears to be perfectly happy - . **I mean** she can't be a hundred per cent happy, nobody is, but she appears to be happy (Tree and Schrock 2002)

These hashtags act as discourse markers like those that signal a clarifying repair or those that signal a clarification, similarly using copresence heuristics as discussed in Clark and Marshall (1981). Some hashtags are similar to discourse markers that signal repair in that they can signal

for a clarification of referent identification, similar to how *oh* can function as a precursor to a

clarifying repair (Tree and Schrock 1999). Repair in CMC is not uncommon: Sanna-Kaisa

Tanskanen and Johanna Karhukorpi (2008) focus on repair in email. The clarification found in

this chapter of this dissertation are similar to the repairs in email found with Sanna-Kaisa

Tanskanen and Johanna Karhukorpi (2008). However, unlike previous studies of repair in CMC,

I show that tag reframing can also be further divided into the types of repair outlined in Clark

and Marshall (1981). Unlike pragmatic repair in Clark and Marshall (1981), however, hashtags

do not have a second turn before the reframing hashtag. Instead, the clarification happens in the

same turn as the initial ambiguous message, usually at the end of the tweet. Because of this, the

hashtags that enhance the meaning of the tweet using mutual knowledge will be called tag

reframing.

4.2 Literature on repair

First, to understand how tag reframing acts in a way to clarify messages, we must refer to the

literature regarding repair and mutual knowledge. Repair can include giving a clarification or

context for a statement; it does not involve fixing wrong information. Repair is needed because

we cannot be in other people's minds to always know who or what they are talking about. Clark

and Marshall (1981) argued that mutual knowledge is essential for successfully identifying the

referent being talked about in a finite amount of time (else our conversations about even the

simplest ideas would take eons). In fact, Clark and Marshall discuss the idea that we use

particular types of repairs to specify information about a definite noun phrase. Repair relying on

compresence heuristics and repair using copresence can help people interpret the speaker

meaning of definite expressions. The copresence heuristics as outlined by Clark and Marshall

include physical copresence, linguistic copresence, indirect copresence, and community membership. These will be vital in the analysis of hashtags, since our categories include temporal copresence (instead of physical copresence), linguistic copresence, indirect copresence, and community membership.

Physical copresence relies on physical information that is available to both interlocutors. For instance, if a desk is broken in the same room as I am, I can physically point it out to my audience. Linguistic copresence is what is often called anaphora, where a pronoun can be used when the hearer already knows the antecedent, such as "Jane gave **the boy** his ball. **He** thanked her" where *the boy* and *he* are two referential expressions for the same person. Indirect copresence does some really interesting thing by relying on the hearer to provide the connection between the definite referent and the mention of an aspect of it. An example of this is "Bob bought **a book** the other day, and **the cover** was torn" where *the cover* is referring to a part of *a book*. The heuristic of community membership relies on both interlocutors knowing the same information because they are members of the same community; this is how proper names work since both people have to know the referent of the name. For example, the statement "I saw Bob yesterday" only works if the speaker and hearer are both part of a community where Bob is a salient character.

While Clark and Marshall describe how repair works with definite expressions in spoken language, is repair any different for CMC? In spoken language, repair is usually used after a miscommunication to clarify the message. However, turn-taking in CMC is often not as immediate as it is with verbal communication, since people could interpret things incorrectly without having a repair until the next time that the speaker gets back online, i.e. much online

communication is asynchronous. The problem that can occur with this asynchronous communication is that people cannot react immediately to that miscommunication with clarifying information. In email, another form of computer-mediated language, repair occurs during the same turn as the original message (Tanskanen and Karhukorpi 2008); however, it does come later than the original message in that it appears at the end of the email. Hashtags, which also typically appear at the end of a tweet and usually in the same tweet, are similar in the placement. The previous studies in repair, including both Clark and Marshall (1981) and Tanskanen and Karhukorpi (2008), set a foundation for this study in hashtags.

4.3 Data and Methodology

4.3.1 Corpus collection methods

In particular, to study the pragmatic uses of hashtags, a corpus of 1791 tweets was created and adapted from the 2012 Twitter Stratified Random Sample (Illocution Inc 2013). The corpus available from Illocution Inc are free to download, but the data only includes the tweet: no identifying information or demographic information on any of these twitter users was gathered. After downloaded this data set, the data was cleaned up to only include tweets that included a hashtag, and from there, I had 1791 tweets in this corpus in text file format.

From this text file, a corpus with 1791 tweets was created in an Excel workbook. On different Excel sheets, tweets were put into the category of tag reframing if the hashtag clarifies the tweet content. Descriptive information was gathered about the data, as was information about the usage of it. The use of each hashtag was determined through context and subject to

interpretation, and some hashtags were tallied for more than one category of information packaging.

4.3.2 Description of the data

The 2012 data set under analysis consists of 1791 tweets with 2,344 hashtags altogether. This means that the average amount is 1.309 hashtags per tweet. While it is normal for tweets to have only one hashtag since the average is so close to 1, the tweet with the most hashtags features eight different hashtags to package its speaker meaning.

Not including the mandatory # character, the average hashtag is 9.81 characters. The longest is thirty-three characters, and the four shortest hashtags only include two other characters after #. Four characters and eight characters are the most frequent lengths, since thirty-three hashtags fall into each of these categories. While format of hashtags mandates the inclusion of #, for the post part, hashtags greatly vary in length.

Like character count, hashtags vary in the amount of words used to convey speaker meaning, but if more than one word is used, no spaces are included. For instance, the hashtag with the most words is seven words long, but most, though, consist of one word. Sometimes, no words at all are used. In one of the shortest hashtags, no words are even used, as the two characters are a less than sign < and the number *3*, which is a symbol for a heart. Also, in the non-word category, we see that some hashtags utilize onomatopoeia, like *zzzz* to stand for the sound of snoring. Some consist of just abbreviations for a number of words, like *omg* for "oh my god" and *np* for "now playing".

4.3.3 Methodology

From the Excel sheet, each of the tweets were coded depending on how well they fit into the categories of copresence using the following definitions and example tweets from opportunistically gathered tweets.

Reframing hashtags were coded as community membership when they relied on knowledge that comes from being a part of the same speech community. Since people can be members of different communities, repairs using community membership are not always clear to any individual coder, especially on Twitter, where not all of the audience is familiar with the same references; however, repairs using community membership are directed to those who would get the reference. The tweets (107) - (108) examples of community membership that were used to manually code the corpus:

> (107) I'm at the point where, if Amazon Prime doesn't have what I want... go into mass confusion and slight panic #Firstworldproblems

> (108) Today the Colonels fly out to Oregon State to take on the #16 Beavers tomorrow! We wish them safe travels and good luck. #GeauxColonels

> (109) 4 Papers, 3 days. #challengeaccepted #allthecaffeine

Community membership repairs can be geared to a wide audience by using salient cultural entities, like the phrases in hashtag form: #Firstworldproblems in (1) and #challengeaccepted in (3). The hashtag #Firstworldproblems functions here to let the audience know that the person is quite aware of the triviality of the complaint, yet the person still feels the need to

express it. The hashtag #challengeaccepted was a popular phrase from the TV show *How I Met Your Mother*, which was often uttered by the character Barney before he attempted some outlandish task. Other hashtags are geared to a specific speech community: the hashtag *#GeauxColonels* is targeted towards others who would understand the faux-French spelling of *go*, used by people in Louisiana- it is signaling being French-like and Louisiana affiliated. Using these hashtags and their different scopes in signaling community, the new data set was coded.

Reframing hashtags were coded as temporal Copresence when people used hashtags while live-tweeting during an event; the hashtag points to an event in time that corresponds to a physical entity. The tweets (110) and (111) represent data manually coded as temporal copresence:

> (110) Her outfit omg give me #newgirl

> (111) you know its a boring game when musberger raises his voice for #lsu crossing midfield like its a touchdown #BCS

In (110), if #newgirl was not included, the audience would be clueless as to what outfit the writer is speaking of. Members of the audience watching the TV show *New Girl* at the time of this message would understand what the referent of outfit is; thus, this part of the audience would ascertain the overall speaker's meaning of this tweet. We see this again in (111), where #BCS repairs the expression *its* using temporal copresence of the game at the time of the live-tweet.

Reframing hashtags were coded as linguistic copresence when the hashtag specifies or clarifies a part of the utterance using a coreferent in the hashtag that clarifies something about the noun phrase or part of the utterance. We see this in example (112) and (113) :

(112) Now I have my own I realize how useful it is! #samsunggalaxytab

(113) I feel like a kid at Christmas, when I wake up tomorrow I get to go pick up Spencer!! #puppylove

In (112), the pronoun *it* and the hashtag *#samsunggalaxytab* are two referential expressions talking about the Samsung tablet. In (113), the hashtag *#puppylove* clarifies the meaning of Spencer.

Reframing hashtags were coded as indirect copresence when people used hashtags to clarify involving an indirect mention in the tweet, where the audience has to supply the connection to the hashtag.

(114) Can't wait to kick off season 5 already! #jerseyshore

(115) Everybody is up to no good these days.. #secrets

In (114), the hashtag #jerseyshore specifies what show *season 5* belongs to in this indirect way, in a similar way to inferables. In (115), *#secrets* specifies what kind of *no good* is going on here. These examples show the range of how hashtags can clarify the message of a tweet using indirect copresence.

Using these four definitions and corresponding example tweets, the new data set was manually coded for fitting into this categories. One potential problem is that no coder can be a part of all the different speech communities, so some tweets cannot be coded as community membership when these tweets are signaling to a specific community. Also, no coder can be watching, listening, and present at the same time and place as referenced by when or where the tweets are written, so the ability to code of this data for temporal copresence is also limited. Because of these restrictions, there is the category of ambiguous.

4.5. Results

Of 2,344 hashtags in this corpus, 1760 of these hashtags were tag reframing. As can be seen in Table 4-1, tag reframing is the most common hashtag in this corpus. Furthermore, it appears that while tag reframing is the most common type of hashtag in this corpus, people have preferences for using different categories based on co-presence heuristics, as we can see in Table 4-1, where the counts of each type are arranged from most to least frequent.

Table 4- 1: Tag reframing types

	Count	Percent Reframing tags
Linguistic Co-presence	1279	72.6%
Community membership	301	17.1%
Temporal Co-presence	44	2.5%
Indirect Co-presence	31	1.8%
Ambiguous Category	105	6.0%
Total	1760	100%

In row 1, linguistic copresence is the most frequent with 73%. The large percentage of tag

reframing with linguistic copresence may be due to the written nature of tweets and the ease in

access to anaphora. Users of CMC have to rely heavily on linguistic cues because it is a form of

written communication.

One might predict because of the prevalence of certain catch phrases and inside jokes,

community membership as a type of tag reframing would be more popular. However, only 301

hashtags clarify something about the speaker's meaning using information that is only accessible

to membership of the target community, as shown in row 2 of Table 4-1. Hashtags that repair using community membership co-presence are not the most common type of tag reframing. As seen in Table 4-1 row 2, community membership consists of 17.1% of tag reframing.

According to row 2 in Table 4-1, community membership is more common than tag reframing that rely on temporal co-presence and indirect co-presence, but not as common, according to the corpus for this study, as tag reframing using linguistic co-presence.

According to Table 4-1 row 3, temporal co-presence is the second least common heuristic used to for tag reframing, according to this study. Only 44 hashtags used temporal co-presence as a means for tag reframing.

A number of hashtags were difficult to categorize, including tweets that feature codeswitching, target a small community, or discuss a very local event. These appear in the Table 4-1 row 5, labeled as *ambiguous*, which consists of 105 hashtags (6% of tag reframing). These hashtags are impossible for any one person to code because he or she cannot be in all places and belong to all communities.

4.4 Analysis

There were 1760 hashtags (75.1%) in the corpus that were found to clarify part of the message. The reframing tags can be further subdivided using copresence heuristics as discussed in Clark and Marshall (1981). Unlike the study by Clark and Marshall (which only studies repairs on definite noun phrases), the hashtags can clarify a portion or the entire message, and it is not just limited to definite expressions.

These hashtags were usually found at a clause end (either the beginning or the end) and were independent of the syntax. However, instead of having an insult or expression hesitation, these clearly acted as referential expressions to others in the tweet or clarified part of the message in some way. We will see in the following subsections the four categories within repair on how hashtags can package more information. These categories are (i) community membership (section 4.4.1), (ii) physical copresence (section 4.4.2), (iii) linguistic copresence (section 4.4.3), and (iv) indirect copresence (section 4.4.4).

4.4.1 Community Membership

Repairs that involve community membership rely on the knowledge that one must know as a member of the same speech community in face-to-face communication. Similarly, hashtags that reframe the tweet based on community membership also rely on knowledge of the community to enhance the meaning of the tweet. Since people can be members of different communities, repairs using community membership are not always clear, especially on Twitter where not all of the audience is familiar with the same references; however, repairs using community membership are directed at a subset of their followers.

4.4.1.1 Community Membership relevant literature

Community membership reframing tags are of interest to other scholars, who have written on specific hashtags in their studies. Sharma (2013) shows how users employ *blacktags* to signal race and race related issues, which creates a sense of unity in a community, while Caleffi (2015) notes that hashtags create "communities of people interested in [a specific] topic." Both note the importance of community and how one joins the larger conversation on topics through specific

117

hashtags, but ignore how the community membership actually allows the intended audience to decipher the speaker's meaning.

4.4.1.2 Community Membership analysis

Community membership repairs can be geared to a wide audience by using salient cultural entities. For instance, one tweet states:

(116) Went to five guys today #winning

The hashtag #winning is a hashtag based on quotes from Charlie Sheen, when during some interviews where he exhibited erratic behavior- he used the term *winning* often to express emotion; it functions here to let the audience know that the person is quite aware of the trivialness of the accomplishment, yet the person still feels excited about it. It became very popular that year, but the ironic meaning behind it at the time was encoded in being part of the community and this knowledge of usage.

A hashtag can also target a smaller subset of the followers by having a very specific community reference, like the one in the following example:

(117) RT @vballproblems: #525 That awkward moment when you shank a really easy ball... #vballproblems

In (117), the hashtag *#vballproblems* is using community membership, specifically the volleyball player community, to clarify some element of the tweet, and in this case, it is specifying what type of ball they're talking about: the actions described in the tweet involve playing volleyball,

but only volleyball players would understand the awkwardness described in the tweet. We see a similar pattern in (118) and (119) where the speakers talk about problems of other communities:

(118) [...] I wish a had a butt :c #whitegirlproblems

(119) Just took an hour in the shower detangling my hair #CurlyHairProbz

The "Problems" hashtags are not the only ones, though, that can help the audience decipher the speaker's meaning in some way by signaling community membership. In (120), the hashtag *#team* clarifies the reference to *hockey* earlier in the tweet.

(120) Time to watch the hockey :) come on #teamGB

If someone following the Olympics reads this, they would know that *the hockey* in the tweet message is not talking about a game being played locally; instead, the use of *team* before *gb* signals that they are talking about the Olympics; we see this same pattern in (121) where the tweet features *#teamusa*:

(121) MEN'S GYMNASTICS!!! Let's go #TEAMUSA!

It is traditional for fans of sports teams to refer to their favorite national team using this combination of *team* and abbreviation for the country. These messages are specifically geared to the subset of the audience that would understand this cultural reference.

Tweets that feature reframing tags that use community membership rely on some shared knowledge of the community. Not every message is geared towards the entire set of followers of that tweet; in fact, often people will not be pedantic with the tag reframing to give enough information for informed audience members to decipher the meaning while leaving out the less culturally knowledgeable followers.

4.4.2 Temporal Co-presence

 Temporal co-presence involves live-tweeting, where someone tweets during some type of organized activity. Live-tweeting is likely only to occur in special instances, like televised events and conferences, where someone may use the event name to clarify the meaning of their tweet in a character-constrained space.

Temporal copresence is similar to physical copresence (Clark and Marshall,1981), where people might be in the same physical location (like a conference or party), or where people may physically have a television or radio tuned into a larger event (like watching the season premiere of a favorite show). However, unlike physical copresence, the speaker is perhaps in a different physical location than the audience of the tweet, as is the case with live-tweeting a football game. The speaker and audience must have timing in common to have successful clarification from this additional information.

4.4.2.1 Temporal Co-presence relevant literature

Temporal co-presence occurs with live-tweeting. Schirra et al. (2014) studied live-tweeting during the TV show *Downton Abbey,* particularly in what prompts live tweeting, and found user engagement involved both occasional live-tweeting and more often "listening" to the Twitter conversation about the show, where the Twitter user is only passively reading the tweets. Pittman and Tefertiller (2015) looked at how live-tweeting as a social activity happens more often as a synchronic activity, as opposed to live-tweeting during asynchronous streaming. Because it is an activity that occurs in spikes during specific times, temporal copresence may be less common than the other types of tag reframing, because of time constraints.

Counts and Geraci (2005) used the term physical co-presence in discussing creating a "shared experience" through the use of older versions of social media. Because of the shared experience and synchronic activity, this category may lend itself more to a temporal co-presence since it relies on something physically there with the speaker and hearer at the same time, but in different physical locations.

When tweeters use tag reframing, they are clarifying their message based on engaging in a social activity at the same time as other users. With the older social networking sites discussed in Counts and Geraci (2005), users organized events and reconnected after events using social media, which is not as synchronic as live-tweeting, thus not truly physically co-present as tag reframing currently allows.

4.4.2.2 Temporal Co-presence analysis

In (122), #madman reframes the message with temporal co-presence since it is used to let other watchers of the show *Mad Men* know that the information in this tweet is talking about the current intimate scene that is occurring at the moment of the tweet.

(122) Someone hurry up and invent Viagra and give one to Don Draper. #madmen

If #madmen was not included, the audience might be clueless as to who Don Draper is, and why he needs a little blue pill. Members of the audience watching *Mad Men* at the time of this message would understand that who is the referent of Don Draper; thus, this part of the audience would ascertain the overall speaker's meaning of this tweet. We see this again in (123), where *#britawards* gives the physical context for the first part of the tweet:

(123) Wish someone would sing this song for me #brunomars #britawards

While *#brunomars* clarifies the *someone* in the first part of the tweet, *#britawards* specifies which of Bruno Mars' performances is the desired singing of the speaker. The speaker is talking about a specific performance that is happening at the time of this live tweet.

Some reframing tags rely on more than one form of co-presence to clarify the message. In (124), the referents *Willie*, *Mike Boogie*, *Janelle*, and *Dan* are referencing people from the television show *Big Brother*. Both *#BB14* and *#BBAD* give physical context that reframes the message of the tweet, as shown in (18):

(124) #BB14 #BBAD so Willie is already playing Mike Boogie and Janelle is in full

coaching mode with her team. Dan looks like he misses his wifey!

By using something that is physically present to both interlocutors (the current episode of *Big

Brother* on TV) at the same time, the audience will understand who these people are and what

has been happening during the episode to motivate this tweet. These hashtags create ease in

communication without consuming too much space by referring something salient to the

interlocutors.

4.4.3 Linguistic Co-presence

Tag reframing that involves linguistic co-presence specifies or clarifies some part of the

utterance, but this involves anaphora. In section 4.4.3.1, the relevant literature is discussed,

while in 4.4.3.2 provides an analysis of the tweets in this corpus that fit into this category.

4.4.3.1 Linguistic Co-presence relevant literature

Linguistic co-presence, sometimes called anaphora, is where something can be brought up again

in pronoun form (anaphor) because we already know the antecedent (the first referential

expression). Levinson (1987) describes anaphora as, "one linguistic element, lacking clear

independent reference, can pick up reference through connection with another linguistic

element". While many have analyzed the pragmatic, semantic, and syntactic constraints of

anaphora and antecedent relationships (Chomsky 1981; Levinson 1987; Huang 1991; Huang

2000; Kratzer 2009), the main concern of this section is showing how the tag reframing

capitalizes on anaphora to clarify or reframe the meaning of the tweet.

4.4.3.2 Linguistic Copresence analysis

Linguistic co-presence specifies, reframes, or clarifies a part of the utterance, but this involves actually having an ambiguous term or referent in the message of the tweet and a coreferent in the hashtag that clarifies something about the noun phrase or part of the utterance. In (125), we see that *#TIGERS* defines the referential expression *boys*:

> (125) Well done boys we deserved this #TIGERS #WORLDSERIES

The second hashtag *#WORLDSERIES* acts as a physical co-presence reframing tag, giving a physical context to the teams and the game. The hashtag *#TIGERS* is linguistic copresence since it renames the first referent, clarifying which team they think *deserved* to win.

In (126), the referential expression *the sun* could be talking about the sky's sun, but it can also reference a person if his or her nickname was *the sun*.

> (126) I wish the sun wouldn't fucking come out tomorrow. #hatethesunnewspaper

Instead, *the sun* is particularly discussing the newspaper version.

In (127), the hashtag *#OneDirection* specifies who the referent for *they* is in the first part of the tweet.

(127) They look soo cutee ! #OneDirection<3

While linguistic co-presence utilizes two different referential expressions for the same referent, indirect co-presence make a statement about referent or experience then reframe the meaning by indirectly referring to part of that referent or experience, as we see in section 4.4.4.

4.4.4 Indirect Co-presence

Indirect co-presence is similar to linguistic co-presence in that it relies on information that is already available in the tweet, but unlike linguistic co-presence, the co-referent was not directly mentioned. Section 4.4.4.1 presents the relevant literature to the analysis of tweets that fit into this category, shown in 4.4.4.2.

4.4.4.1 Indirect Co-presence relevant literature

Indirect co-presence in tweets relies on mentioning some aspect of it, and the hearer has to provide the connection in some way. When this happens in spoken language, it is called inferables in Prince (1981). Prince situates this in the study of given and new information, and describes this process as, "A discourse entity is Inferrable if the speaker assumes the hearer can infer it, via logical-or, more commonly, plausible-reasoning, from discourse entities already [e]voked or from other [i]nferrables."Huang (2000; 2007) uses the term bridging anaphora to describe this process, and in his work points out how the two referring expressions are not coreferential, but rather "linked." We see this link occurring in tweets, especially where the

125

tweeter makes the tweet content purposefully ambiguous, so the message of the tweet is clarified with information that is inferred from the words in the tag reframing hashtag.

4.4.4.2 Indirect Co-presence analysis

Repairs that use indirect co-presence specify aspects of some earlier referent. Usually, this type of co-presence deals with the beginning part of a statement that is very generalized like, "I bought a used book," and it finishes with, "and a page was ripped" where page is indirectly referring to the book. For the following examples, we see a similar pattern.

(128) This shower is going to be painful #sunburn

(129) It's Jimmy Heller time! #Prototype2

For instance, in (128), we see how *#sunburn* is specifying what will painful about *this shower* is indirectly identifying when he states, "This shower is going to be painful".

The tweet content "This shower is going to be painful"" is general; the shower could be painful for a different number of injuries. Someone could use the same tweet content with the hashtag *#hotwaterheaterbroken* if the reason for the painful shower is freezing temperatures of water. We see a similar pattern in (130) where the speaker is talking about a video game.

(130) It's Jimmy Heller time! #Prototype2

The tweet message "It's Jimmy Heller time!" talks about a video game character, but the audience might think Jimmy is a friend or cousin, without the hashtag *#Prototype2*, which

126

includes the video game name. It has this part/whole relationship discussed with bridging anaphora.

4.6. Conclusion

Tag reframing occurs when a hashtag flavors the meanings of a tweet, usually by clarification or contextualization. There are four types of tag reframing, which correspond to the co-presence heuristics from Clark and Marshall (1981): (i) by tying into community membership; (ii) by alluding to temporal or physical co-presence; (iii) by referring to early wording via linguistic co-presence; (iv) and through bridging-type connections of indirect co-presence. Not all hashtags that change the meaning of the tweet could be easily categorized into this schema; since not everyone is part of the same sub-communities or in the same place or time as everyone else, many tweets remain ambiguous as to what particular category best fits the tag reframing. Linguistic co-presence was the most used category of tag reframing, while indirect co-presence and temporal co-presence were the two least popular categories.

Clark and Marshall (1981) study how people use vertical repair in conversations where the repairs get stronger to ensure the definite noun phrase is successful. This work only looks at individual tweets, so I cannot say one type of tag reframing is more successful than others. However, I might expect the stronger copresence heuristics to be used more often since tweets do not always have the turn-taking that comes with spoken conversations.

If I follow that expectation, then I would expect there to be many temporal copresence tweets given that physical co-presence, the counterpart from Clark and Marshall (1981), is one of the strongest repairs. However, this was not the case since only 2.5% of the tag reframing hashtags

used temporal co-presence. One reason for less hashtags clarifying using temporal copresence

may be the means of communicating with CMC: typically, when people use CMC, they are not

in the same physical location as the audience, so referring to salient physical entities is difficult,

since it depends on pointing to a particular time where that physical entity will be salient to the

speaker and audience.

Out of the categories of tag reframing, linguistic copresence was the most frequent. Since tweets

are written, it makes sense, but it highlights the differences in use of written and spoken DMs.

These two differences in the use are interesting because their frequency may be indicative of how

written DMs work on social media, where anaphoric relationships are more available than

physical deixis to ensure felicitous communication.

CHAPTER 5

USING REFRAMING HASHTAGS TO UNDERSTAND *#NASTYWOMAN* TWEETS

5.1 Introduction

As mentioned in Chapter 1, DMs can reveal speaker's attitude, according to criterion (iii). This chapter will look at the different attitudes surrounding *#NastyWoman*, a hashtag that was used with varying attitudes before election day. In this chapter, I use tag reframing hashtags and a learning algorithm to classify the attitude and use of tweets in this big data set, the problems that arise with this method of classification, and the combinations of some collocates in the tweets. Using these, I give an analysis of the use of *#NastyWoman* in the 2016 presidential election.

As discussed in Chapter 4, one function of hashtags is to reframe the content of a tweet. While others have noted how hashtags are involved in community allegiance (Caleffi 2015, Sharma 2013), in section 4.4.1, I presented analysis of the tag reframing hashtags that used community membership to clarify the meaning of the tweet and the attitude expressed with the use of *#NastyWoman*. This chapter uses tag reframing hashtags involved with community membership to decipher the meaning of *#NastyWoman* tweets.

5.2 Background

During the 2016 United States presidential election, there was a prevalence of gendered adjectives (Browning & Fleckenstein 2017). In particular, a corpus study revealed that adjectives describing Clinton were focused on her "unlikeable" personality while adjectives describing Trump were focused on his fitness for office. One of the gendered, personality-linked adjectives aimed towards Hillary Clinton, the first major female candidate for president, was the adjective *nasty*. Section 5.2.1 will look at the two ways *#Nastywoman* is used by the different communities, while 5.2.2 will look at the nature of insults, slurs, and reclaimed epithets.

5.2.1 Ambiguity in the use #Nastywoman

The adjective *nasty* was notably used by the male candidate Donald Trump while he was criticizing Clinton during the third debate. After this occurred, many people began using the hashtag *#NastyWoman* in response to that debate statement. One tweet by a Trump supporter calls it a "Pro-Hillary term" :
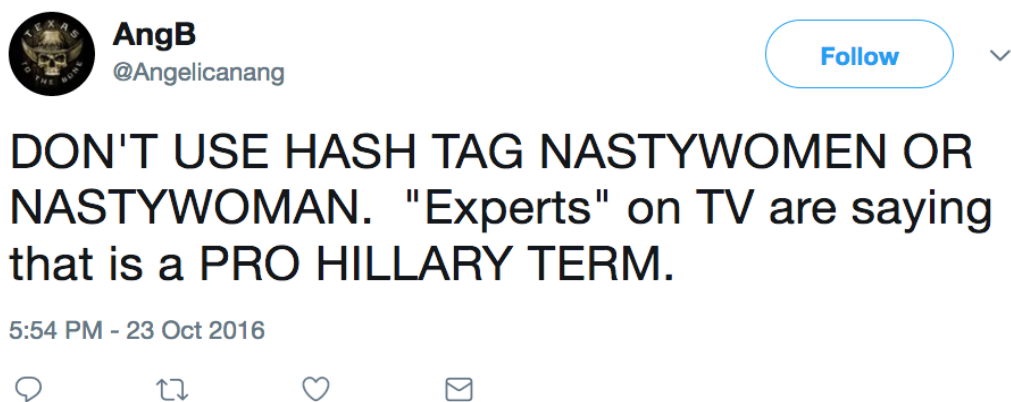
Figure 5- 1: Tweet about #NastyWoman

This person even avoided using the hashtag because this person wanted to avoid supporting the female candidate. The user @Anglican is only partially right: this hashtag can be used with a positive, empowering attitude by Clinton supporters, but it can also be used by Trump supporters, as well.

Two polar opposite uses of the hashtag #NastyWoman arose after this debate, which also account for a majority of the #NastyWoman uses. The first way people used *#NastyWoman* was in the same ilk as Trump, in a demeaning, insulting way (discussed in section 5.2.2.1). The second way people used it included how some people reclaimed this term to empower other community members and identify with Clinton (discussed in 5.2.2.2).

This chapter will use tag reframing hashtags that are associated with the Clinton campaign and her supporters, like *#ImWithHer* to assist with the coding of this data, while using tag reframing hashtags associated with the Trump campaign, like *#MAGA*, to assist with the coding of the insulting versions of *#NastyWoman.*

In some tweets, it was easy to see how the tweeter intended the meaning of *#NastyWoman,* like in (131) where the content of the tweet is clearly insulting the female candidate, and (132) where the content is clearly using this term in a reclaimed way:

> (131)Thank #JesusChangesEverything and the evil #ClintonFoundation and cartel led by #NastyWoman #HillaryClinton exposed #LockHerUp

> (132) I've joined the nasty party and change my screen name. lol #ImWithHer #NastyWoman and proud to be one!

However, there were some tweets, like that in (133) without the original hashtags, that could be interpreted very differently with different tag reframing hashtags, as shown in (134) and (135), with hashtags associated with the different communities:

(133) Because our children are so important to Michele & #HillaryClinton #Nastywoman

(134) Because our children are so important to Michele & #HillaryClinton #Nastywoman #draintheswamp #VoteTrump

(135) Because our children are so important to Michele & #HillaryClinton #Nastywoman #ImWithHer #VoteClinton

In (133), it is difficult to decipher if the hashtag *#Nastywoman* is used pejoratively towards Michelle Obama and Hillary Clinton, or if it is used in an empowering way. However, with the hashtags for (134), the tweet would be interpreted as being associated with someone who is part of the Trump community; thus, it would be used more like how Trump used it towards his opponent. On the other hand, the meaning of *#NastyWoman* in (135) is seen more as an empowering, reclaimed way. This supports the findings in the previous chapter. The tag reframing hashtags can be used for interpreting the meaning of *#NastyWoman*, especially with the help of a computer learning algorithm.

5.2.2 Slurs, Insults, and reclaimed epithets

This chapter looks at how tag reframing can be used to interpret the hashtag *#NastyWoman* as being insulting or reclaimed, so to understand that, I present information about insults and slurs in 5.2.2.1 and reclaimed pejoratives in 5.2.2.2.

5.2.2.1 Slurs and insults

After Trump used the adjective nasty to describe Clinton during one of the 2016 presidential debates, some tweeters used the #NastyWoman hashtag, in an insulting way, as in (136):

> (136) @HillaryClinton, your'e [sic] truly an old #nastywoman through and through. You make me sick.

Jucker and Taavitsainen (2000) explains how insults are concerned particularly with the perlocutionary effect, and talks about three types in this diachronic study: flyting, name-calling, and flaming. In (136), we see something that is akin to name-calling, especially since it is tweeted directly to @HillaryClinton, the Twitter handle of the female candidate. However, it is a flaming type of insult, as described in Jucker and Taavitsainen (2000) in that it is similar to name-calling, but it's "explicit purpose [is] to insult other participants and to provoke insults from others." The message in (136) is insulting Clinton and provoking her supporters. However, unlike flaming, the pejorative tweets in this corpus are more creative and less rule-bound, which are two characteristics of flaming, according to Jucker and Taavitsainen.

Cepollaro (2015) gives an account of insults using presupposition- when a speaker uses a slur, the audience "share[s] the speaker's derogatory attitude." In this case, the audience members may be complicit in the insulting usage of *#NastyWoman.*

Reclaimed versions of these insults are discussed very differently in scholarly works, as shown in section 5.2.2.2.

5.2.2.2 Reclaimed slurs

In a reaction to Trump's insult aimed at Clinton, many tweeters of *#NastyWoman* reclaim the term:

> (137) This proud #nastywoman has had her say! #ImWithHer
>
> (138) She slayed it! Love that #NastyWoman! #MadamePresident #ImWithHer

A few similar reclaimed words are used in an empowering way in-group: *queer* (Brontsema 2004) and *bitch* (Sutton 1995, Kleinman et al. 2009). This is similar in some ways to taboo intensifiers; according to Constant et al. (2009), taboo intensifiers usually "involve emotions like anger, frustration, and aggression" but can be used in a positive way to express solidarity. However, slurs are different in that they can express heightened emotion (both positive and negative, depending on the user, in a similar fashion to that of taboo intensifiers), but they also target a specific marginalized community (Yoon 2015). Yoon (2015), which looks at the semantic expressive nature of slurs, states that the semantics of reclaimed slurs " do not seem to range on the negative interval" and instead, "the more negative the original slur was, the more positive/intimate the in-group slur seems to become." Because of bipolar nature of slurs and

their reclaimed counterparts, we would expect to find a bipolar use of *#NastyWoman* in this corpus.

Kleinman et al. (2009) talks about a potential problem with reclamation: during the 2008 election, when Tina Fey and Amy Poehler used *bitch* in a reclaimed way to advocate for their candidate, Tracy Morgan responded using *bitch* pejoratively towards Tina Fey, and this "reveals that any attempt at 'reclaiming' 'bitch' can lead someone, most likely a man, to return the word as a slur." This is a pattern seen in the 2016 use of *#NastyWoman* where the supporters of the male candidate used the hashtag in a pejorative way after the reclamation of this term.

Reclaimed epithets are often used as indicators of some type of power dynamic. Brontsema (2004) describes the patterns with the reclamation as not having a clear endpoint, but rather a fluctuation in different uses, and successful reclamation is not always the eradication of the pejorative form. In the case for *#NastyWoman*, using reclaimed *nasty* serves as a rejection of the cultural values imposed on Clinton. It allows tweeters to claim their own agency in their identifying terminology.

5.3 Data and Methodology

In section 5.3.1, the gathering of a big data corpus is described. In section 5.3.2 and 5.3.3, two potential ways that this data was classified are presented. Section 5.3.2 provides information on how the data was manually classified, which was a tedious and time consuming process, whereas in 5.3.3, the process of using a learning algorithm and tag reframing hashtags to quickly and efficiently classify the data is detailed. In 5.3.4, I give a description of the data used for this analysis.

5.3.1 Gathering the tweets using R

Tweets were gathered using an R script from Tatman (2015). R script is a piece of code in a programming language, that can be inputted in an interface like Rstudio. Many use R to write code that calculates statistics, but it can also be used by corpus linguists to collect data and analyze the data. The R script worked with the Twitter API to collect the tweets into a .csv file that worked with Excel.

Since the Twitter API only allows access to 14 days of tweets at a time, tweets for each day were collected individually, to ensure getting every tweet for that day. This process was repeated for 18 days, from October 21, 2016 until November 8, 2016. Only one day of tweets were collected in each iteration so the entire day's tweets were gathered in separate files. After each day was saved as individual files, these 18 files were manually combined into a large csv file for analysis.

5.3.2 Coding data using Excel for training data

Manually coding this big data set began by taking the csv file, reading the tweet, and identifying how *#nastywoman* was being used in the tweet, and this manually coded data became the training data for the algorithm. The data was at first manually coding for how *#nastywoman* was being used; if it was being used in an empowering, positive way, it was labeled as "yes.reclaimed." If it was being used in a negative, degrading way, it was labeled as "no.reclaimed." If it was used in in a way that was hard to code because of content, it was

labeled as "ambiguous." There was a number of tweets that were difficult to code, like those

selling items (9), using a number of trending hashtags (10) and trending topics (11) to cash in

on the popularity of the tag, but these were not excluded from the final data set.

> (139) We got all your election watches ready here! #Meteor2016 #MAGA
>
> #NastyWoman https://t.co/X4LZx2QeLc

> (140) RT @LTCartoons: 20%off sitewide #Sale ends Mon 1159pmPT Est 1997 #Happy
>
> #Shopping #christmas #chanukah @zazzle #sneakers #nastywoman

> (141)-"How to learn 'Supreme Cleave'"-"Learn Elvish at Home"-"How to defeat
>
> #NastyWoman Sorcereresses"-"How to Resist Domination'"

From here, 3,000 tweets were manually classified, but it became increasingly difficult because of

the sheer size of the file. Even using sorting options on Excel to code retweets in the same way

as the other, similar tweets did not automate the process. Using a search spreadsheet equation

to look for certain hashtags as a way to group reclaimed instances of *#nastywoman* and

pejorative instances of *#nastywoman* was unsuccessful because some tweets had both *#maga*

and *#imwithher*. The next step involved a truth table to code the data- if it had more than one

hashtag affiliated with the Clinton campaign (without any that was clearly associated with the

Trump campaign), it was coded as "yes.reclaimed," while those using more than one hashtag

affiliated with the Trump campaign (and none of those positively associated with the Clinton

campaign) was coded as "no.reclaimed." While this allowed me to code 17,795 tweets, it was

137

inefficient and problematic when there were several hashtags of different campaigns in one tweet, as shown in (7). This problem was solved by using a learning algorithm that used reframing hashtags, which is described in 5.3.3. This coded data was used as the training data to input into the learning algorithm.

5.3.3 Coding data using Python

A python script that was set up with three parts was utilized to automate the classifying process. The main python code involved a first part where a python script was used that trained from an already labeled set of tweets and assigned a value, positive or negative, based on being reclaimed or not. Each time it encountered a hashtag that was in a reclaimed tweet from the training, the positive value of that hashtag increased. Each time it encountered a reframing hashtag that was in an insulting version of *#NastyWoman*, the value of that hashtag decreased. This encoded the information into a .json file, which the second part of the main python code used.

The next part was to take another set of data that was already coded to run through the trained algorithm. It looks at the content of the tweets, and it counts up the values, both positive and negative, assigned to hashtags and assigns the sum to the tweet. The algorithm prints out a label, reclaimed or not reclaimed, based on this value for the tweet.

One potential problem is that there might be a number of tweets that use the same reframing hashtag for different community memberships. There was a need for something in the program to account for the ambiguity of some of these hashtags, where they might be used many times, but the number might be closer to 0 because of the fluctuation in different uses for different communities. Because of this, the program was adjusted to include a usefulness threshold, which labels hashtags that don't appear very often or that appear often but in different ways as "Ambiguous".

This part of the python code allows the programmer to set a threshold, which has an initial value of 10. Comparing this new column of "reclaimed", "insulting", and "ambiguous" with the original hand-coded label gives an idea of what is trained well and what is not. If the program labels it as "Reclaimed" and it matches the existing hand-coded column, then it is working well. If the program labels it as "Insulting", and it matches the existing hand-coded column, then it is working well. When the threshold is set at a higher number, there is lower probability that the script will label tweets as "Reclaimed" or "Insulting" incorrectly, but more tweets will be labeled as "Ambiguous". If there are false positives, then the programmer can set a higher threshold in the configurations part of the python code.

The program is set up to where more hand-coded data can be inputted into the first part and rerun part two while continuing to tweak the threshold until it accounts for the majority of the data without printing out falsely labeled data.

Finally, the third part of coding the data involves inputting the entire data set of tweets to get the full data set coded for content. It prints it out as a .csv file, where the script coded the tweets as "Reclaimed", "Insulting", and "Ambiguous" in the last column. From here, the data can be analyzed in Rstudio, an interface that uses the computer language R, for count and statistical information. In the next section, descriptive information about this data set is presented.

5.3.4 Description of the data

There are 55,933 tweets collected over 18 days right before the 2016 presidential election. There are 1,083,782 words total in this corpus, with an average of 16.172 words per tweet. There were 7,198,108 characters in this corpus, with an average of 128.7 characters per tweet.

Some tweets included tag reframing hashtags that were associated with multiple campaigns, but the algorithm was able to account for this complication since it counted the values across the tweet content. An example of a tweet that used hashtags that are associated with both campaigns, originally discussed in (3), appears in the screenshot below in Figure 5-2:

Figure 5- 2: Tweet with different tag reframing hashtags

This tweet, and others like it, are difficult for the algorithm to code because of the use of

hashtags associated with different communities being used in the same tweet. The tweeter

@missdonnasgavel might talking about how much Hillary Clinton and Michelle Obama have

done for children, or @missdonnasgavel might be saying this sarcastically, thinking that they

have not done anything for children. This is where the training data was so vital to the

investigation in the usage of *#NastyWoman*, as shown in section 5.4.

5.4 Training data and tag reframing hashtags

5.3.4 Training Hashtags

In the first pass of training, a file with 324 tweets coded for reclaimed/insult (where there were 161 reclaimed and 163 insulting) was used as the first set of training data. There were 330 tag reframing hashtags used in this first training set, 12 that were used frequently for one community, resulting in a large positive (for the reclaimed version of *#NastyWoman*) or a large negative (for the insulting version of *#NastyWoman*) number assigned to the hashtags during the training phase. The tag reframing hashtags with the largest absolute values appearing in the .json file were *#imwithher*, *#maga*, and *#draintheswamp*, as shown in Table 5-1.

Table 5- 1: Tag reframing hashtags in first pass of training with highest magnitude of values

Tag reframing hashtags	Number assigned	status
#maga	31	insulting
#draintheswamp	-29	insulting
#crookedhillary	-20	insulting
#imwithher	86	reclaimed
#dumptrump	10	reclaimed
#strongertogether	10	reclaimed

An example of how the hashtag *#draintheswamp* was used in conjunction with a pejorative *#nastywoman* in the training data appears in (142), while *#MAGA* is used in (143) with a pejorative version of *#nastywoman.* In (144), reclaimed *#nastywoman* with *#Imwithher.*

(142) @michellemalkin: Hillary isn't just a #nastywoman. She's a ruthless hatemonger. #draintheswamp #FlyoverQuotable @CR https://t.co/RzNL58fLr3

(143) "Yea...@HillaryClinton has the right temperament...#nastywoman #MAGA with @realDonaldTrump

(144) I filled out my ballot and dropped it in the drop box today. Feels great to be a #nastywoman #ImWithHer

The hashtag *#maga* stands for *Make American Great Again*, a campaign slogan for Donald Trump, while *#draintheswamp* was a slogan based on one of Trump's campaign promises.

The hashtag *#imwithher* was used as a campaign slogan for the Hillary Clinton campaign, emphasizing the fact that she was the first female presidential candidate running on the Democratic Party's ticket.

In an attempt to code more data, a larger, similarly balanced csv file was created with 3,187 tweets that were manually labeled reclaimed and 3,187 tweets that were manually labeled insulting, with a total of 6,374 tweets total. From here, the following breakdown of tag reframing hashtags emerged as appearing frequently for a community:

Table 5- 2: Tag reframing hashtags in second pass of training with highest magnitude of values

Tag reframing hashtags	Number assigned	Status
#maga	-737	insulting
#neverhillary	-567	insulting
#alwaystrump	-467	insulting
#wrong	-465	insulting
#imwithher	1843	reclaimed
#vote	168	reclaimed
#nastywomenvote	167	reclaimed
#dumptrump	148	reclaimed

This, when ran against the larger, uncoded file, resulted in coding almost 10,000 more tweets in categories other than *Ambiguous*, as shown in Table 5-3:

Table 5- 3: Breakdown from coding with tag reframing hashtags

Tag reframing hashtags	After first training	After second training
Reclaimed	10,163 tweets	16,226 tweets
Insulting	3,563 tweets	7,350 tweets
Ambiguous	41,027 tweets	32,350 tweets

In Table 5-3:, we can see how this algorithm labeled 13,726 tweets with the first set of 324 training tweets as either *Reclaimed* or *Insulting*. Also shown in Table 5-3 is how it labeled 23,576 tweets as *Reclaimed* or *Insulting* with the final set of 6,374 training tweets. However, even after training algorithm with thousands of training tweets, there were over 30,000 of tweets that were still labeled as *Ambiguous*, which will be addressed in section 5.5.

5.3.4 Problems with Training

There are many reasons why tweets were labeled as *Ambiguous*, including: not enough hashtags, too many untrained hashtags, and presence of hashtags on opposing communities.

One of the reasons the algorithm labeled some tweets as *Ambiguous* is because *#NastyWoman* was the only hashtag used in the tweet, as shown in (145) and (146):

(145) RT @JeffreyGuterman: . @RepBrianBabin: I think sometimes a lady needs to be told when she's being nasty. #nastywoman https://t.co/WJzNcQB

(146) November 8th where you at. I'm ready for that #NastyWoman POTUS https://t.co/Ui0GXdeTaL

Because the training data was balanced for reclaimed and insulting versions of #NastyWoman, this hashtag itself had the value of 2, which is not enough to make it past the threshold of 4. There were 22,507 tweets that had less than two hashtags in the tweet. This means that 22,507 of the 32,350 *Ambiguous* tweets (69.6%) were labeled that because of lack of other reframing hashtags.

One problem is affiliated with the gathering of training data: while some tweets had many hashtags, the tag reframing hashtags in the tweet may have been assigned a smaller absolute value or not appeared in the training data at all. Examples of this are in (147) where the hashtags *#Dover*, *#Laconia*, *#Manchester*, *#Nashua*, *#NH*, *#NewHampshire*, *#Portsmouth*, *#Bedford*, and *#Derryäó_* don't show up in the training data.

(147) RT @TheNicoleComer: She IS a #NastyWoman #NASTYHILLARY #Dover #Laconia #Manchester #Nashua #NH #NewHampshire #Portsmouth #Bedford #Derryäó_

A similar problem that involved not having more instances of some hashtags show up in the training. In (148), #TrumpTruth was assigned -1 and #ShameOnHer was assigned -2, and while this tweet does seem geared towards Trump supporters, the threshold was not large enough to

assign this combination of tag reframing hashtags as *Insulting*. Instead, (145) was assigned as ambiguous.

> (148) RT @JaneDoeCountry: Can't wait for NOV 8! #TrumpTruth #ShameOnHer #NastyWoman #WarHawkHillary #PodestaEmails13 #Veritas #VoteTrumpPence16

Problems with training data and threshold values account for at least 9,842 tweets of 32,350 *ambiguous* tweets.

Another potential issue with getting an algorithm to code the data based on reframing hashtags is that there were too many hashtags on opposing sides of the election. An example of this is in (149), where #MAGA was assigned the negative number -737 with the training data, and #ImwithHer was assigned the positive number 1843.

> (149) What a #NastyWoman! #MAGA #ImWithHer #MakeAmericaGreatAgain #2A #NRA https://t.co/i6H34u6OO4

Because the hashtags *#MAGA* and *#MakeAmericaGreatAgain* as tag reframing hashtags for the Trump supporter community while *#ImWithHer* was used by the Clinton supporter community, the algorithm assigned this tweet as ambiguous.

5.6 Frequency and Collocations

5.6.1 Frequency of hashtags that are tag reframing

In section 5.4, I showed the numbers assigned to tag reframing hashtags with the algorithm. While this data shows the relative frequency of these hashtags in training data that is reclaimed or pejorative (depending on if the number assigned is positive or negative), it does not say much about the actual frequency of these hashtags in the larger corpus. In Table 5-4, we see that #imwithher, a hashtag associated with the Clinton campaign, appeared 7,161 times in the corpus for this chapter, which is much more frequent than #maga, one of the more frequent hashtags of the community of Trump supporters.

Table 5- 4: Tag reframing hashtags in second pass of training with highest magnitude of values

Tag reframing hashtags	Number assigned in algorithm	Frequency in corpus
#maga	-737	1396
#neverhillary	-567	843
#alwaystrump	-467	471
#wrong	-465	593
#imwithher	1843	7161
#vote	168	1250
#nastywomenvote	167	505
#dumptrump	148	1259

Other hashtags that showed up with frequency in the training data, like #*alwaystrump* and #*wrong* are not that frequent in the corpus, as shown in Table 5-4.

5.6.2 Collocations

This section looks at what frequently collocated with *nasty* and *#nastywoman*. I also looked at collocations within the hashtags themselves of these hashtag parts: never, nasty, hillary, donald, clinton, trump.

5.6.2.1 Collocation of nasty

First, frequent collocations of #nastywoman on the right included *just*, *#imwithher*, and *who*, as shown below in Table 5-5:

Table 5- 5: Collocations to the right of #nastywoman

Left collocation		frequency
#nastywoman	just	1584
#nastywoman	#imwithher	1065
#nastywoman	who	733

An example of a tweet where *#nastywoman* collocates with *just* is shown in (150). In (151), *#nastywoman* collocates with *#imwithher*. In (152), *#nastywoman* collocates with *who*.

(150) This #NastyWoman just voted.

(151) @IMKristenBell you go #Nastywoman #ImWithHer

(152) RT @TeenVogue: RT if you're a #NastyWoman who's headed to the polls to vote for your rights

Collocations to the left were *a* and *this*, where *#NastyWoman* was part of a noun phrase, and *#imwithher*, where *#nastywoman* was another tag reframing hashtag, as shown in Table 5-6.

Table 5- 6: Collocations to the left of #nastywoman

Right collocation		frequency
a	#nastywoman	9552
#imwithher	#nastywoman	2045
this	#nastywoman	1843

In example (153), like in example (151), *#nastywoman* collocates with *#imwithher*, but the words appear on different ordering. In example (154), the determiner *a* appears before *#nastywoman,* while in (155) the determiner *this* appears before the hashtag.

(153) Hey! Grab him by his Ballots #ImWithHer #NastyWoman
https://t.co/kwwEe3xjI6

(154) RT @HuffPostWomen: RT if you're a #NastyWoman https://t.co/I3mBfrv8lP

(155) This #nastywoman voted early for Hillary Clinton

Other frequent collocations to the left included *the #nastywoman* (frequency: 1393-- example in

(156)), *that #nastywoman* (frequency: 770-- example in (157)), and *one #nastywoman*

(frequency: 657-- example in (158)), where the hashtag is part of a noun phrase.

(156) Filling out my ballot like the #NastyWoman I am

(157) I love that #NastyWoman https://t.co/Y6QkPX0ugV

(158) "RT (AT)TeamPelosi: From one #NastyWoman to another, you were an
inspiration last night, (AT)HillaryClinton. -NP"

Looking at the collocations of *nasty* in the hashtags, *#nastywoman* was the most common

hashtag with *nasty* in it, which makes sense since the corpus was based on this hashtag. The

next three collocations of *nasty* include *women*, *womenvote*, and *womenunite*, as shown below in

Table 5-7:

Table 5- 7: Collocations of nasty in hashtags

Collocation of *nasty*	frequency
#nastywomen	1048
#nastywomenvote	505
#nastywomenunite	178

An example of *#nastywomen* appears in (159), while *#nastywomenvote* appears in (160). In (161), this user tweeted with *#nastywomenunite*.

> (159) Guess where you can spot the biggest number of #NastyWomen? At their local polling place on November 8th #NastyWoman #DebateNight #ImWithHer
>
> (160) You'd better believe it. #NastyWomenVote #NastyWoman https://t.co/rxgG5Fdkvt
>
> (161) "i voted today and i'm NASTY. miss park, if you're nasty. #nastywomenunite #nastywoman https://t.co/JglZ442REM"

From the left and right collocations of *#nastywoman*, this hashtag was being used often as a noun phrase (appearing after articles and before just and who). The hashtag *#nastywoman* appeared in both pejorative and reclaimed versions, which is expected given the results in section 5.5.1. The hashtag *#nastywomen* appears frequently in this corpus, as well as

#nastywomenvote and #nastywomenunite, all three hashtags appeared in the context of reclaimed #nastywoman.

5.5.3.2 Collocations of Temporal Terms

There were a number of temporal terms that showed up in these community oriented hashtags. One hashtag piece that showed up on tweets for either side involved #never, like #neverhillary and #nevertrump. The frequencies of these are shown in Table 5-8, where #nevertrump showed up much more frequently than #neverhillary.

Table 5- 8: Collocations of never in hashtags

Collocation of *never*	frequency
#nevertrump	2234
#neverhillary	843

The hashtag #nevertrump, as shown in (162) appeared 1,391 more times than #neverhillary, as shown in (163).

(162) 34% are idiots who believe sensationalized and misleading headlines rather than getting facts. **#NeverTrump** #NastyWoman https://t.co/x9vWi6qz0F

(163) **#NeverHillary** I will never vote for that lying, corrupt #NastyWoman https://t.co/W5vcOVI7Lk

The hashtags *#neverdonald* and *#neverclinton* are not used at all in this corpus of over 55,000 tweets. So, what is going on with the names of the two major candidates in this election in this corpus of *#NastyWoman* tweets?

5.5.3.3 Names in hashtags

The names of the candidates also showed up often in these hashtags that are affiliated with their campaigns. The presence of the candidates' names is to be expected, but how do people name the candidates in tweet form? To investigate this, the search term "#*NAME*" was entered into AntConc, where *NAME* was the name part for each inquiry, while * was the wildcard.

The hashtag #hillaryclinton appeared 1279 times, and an example of this appears in (164).

(164) RT @bannerite: I'm a #NastyWoman and I voted for #HillaryClinton https://t.co/cH8dIJB6Fk

To see what shows up with Hillary Clinton's first name, *#neverhillary* was the most frequent, while *#crookedhillary* and *#hillyes* were also in the top frequencies, as shown in Table 5-9.

Table 5- 9: Collocations of *hill* in hashtags

Collocation of *hill*	frequency
#neverhillary	843
#crookedhillary	522
#hillyes	224

The example from (163) on #neverhillary appears as (165) here. An example of #crookedhillary

being used in a tweet appears in (166), while *#hillyes* appears in (167).

      (165) **#NeverHillary** I will never vote for that lying, corrupt #NastyWoman
https://t.co/W5vcOVI7Lk

      (166) All hired by **#crookedhillary** #nastywoman campaign! https://t.co/Ph11ybj1AG

      (167) Madam President if you're nasty. #NastyWoman #HillYes
https://t.co/xHyUCoZHwL

To see what shows up with only Hillary Clinton's last name, *#clintonkaine* was the most

frequent, while *#clintonfoundation* and # were also in the top frequencies, as shown in Table 5-

10.

Table 5- 10: Collocations of *clinton* in hashtags

Collocation of **hill**	frequency
#clinton	734
#clintonkaine	132
#clintonfoundation	94
#nastywomenforclinton	35

The hashtag #clinton, shown in example (168), was the most frequent hashtag that involved her last name. The second most frequent was *#clintonkaine*, shown in (169) where the tweeter is talking about which set of candidates the person voted for. The hashtag *#ClintonFoundation* is not necessarily talking about Hillary Clinton herself, but using this hashtag, many talked about how the Clinton Foundation was involved with scandals, as shown in (170). In (171), the hashtag *#nastywomenforclinton* was used with other empowering hashtags.

(168) It's becoming more obvious #FBI Director Comey HAS a double standard against #Clinton #NastyWoman in not releasing #Trump ties to Russia!!

(169) I'm a #NastyWoman and I approved this message **#ClintonKaine** #StrongerTogether #pantsuitnation https://t.co/aVlLYFyR9L

(170) Money Grubbing Clintons **#ClintonFoundation** #followthemoney #EmailScandal #nastywoman #NeverHillary https://t.co/baVyJd6XkU

(171) Thank you!! #NASTYWOMENFORHILLARY2016 are #StrongerTogether. #NastyWomenUnite #nastywoman **#nastywomenforclinton** https://t.co/nIAm3onHGE

People were less likely to include just her last name as part of a tweet, compared to her first name. We see a much different pattern when investigating the parts of Donald Trump's names in hashtags.

The hashtag *#donaldtrump* appears 286 times in this corpus, and an example of it used in a tweet is in example (172).

(172) Man of marred mettle @NewtGingrich defended #DonaldTrump like an amateur. He miserably failed to call @MegynKelly https://t.co/KoFLiy7MU1

Next, I investigated what shows up with Donald Trump's last name. The hashtag *#nevertrump* was the most frequent, while *#dumptrump* and *#trumpbookreport* were also in the top frequencies, as shown in Table 5-11.

Table 5- 11: Collocations of *trump* in hashtags

Collocation of **trump**	frequency
#nevertrump	2234
#dumptrump	1259
#trumpbookreport	789

The example of #nevertrump in a tweet appeared in (162) and again here as (173). In (174), *#dumptrump* was used with *#NeverTrump*. Example (175) features *#TrumpBookReport*, a hashtag that was used by the opposing side to liken his lack of preparation for debate to children presenting a book report with little preparation.

(173) 34% are idiots who believe sensationalized and misleading headlines rather than getting facts. **#NeverTrump** #NastyWoman https://t.co/x9vWi6qz0F

(174) Another reminder of who this man is before you vote. #badhombres #nastywoman #mexicansbelike #NeverTrump #DumpTrump https://t.co/lEKa6rdZwk

(175) Blueberries for who? And where was the mother? WRONG. #TrumpBookReport #NastyWoman

However, we see far less frequency in hashtags that just have his first name. To see what shows up with Donald Trump's first name, *#donaldduck* was the most frequent, while *#dirtydonald*

160

and *#predatordon* were also in the top frequencies with only 8 instances each, as shown in

Table 5-12.

Table 5- 12: Collocations of *donald* in hashtags

Collocation of *donald*	frequency
#donaldduck	55
#dirtydonald	8
#PredatorDon	8

An example of the hashtag *#donaldduck* used appears in (176), where the tweeter is talking

about the stunt where Donald Duck impersonators were sent to Trump Tower to mock the

Donald Trump campaign. The hashtags *#dirtydonald* (177) and *#predatordon* (178) are used

as referential expressions for Donald Trump.

> (176) So 1st #crookedhillary shit on #honestabe now **#donaldduck** what more proof do
> you need this #nastywoman hates america! #VeritasProject

> (177) Love that #nastywoman. using **#dirtyDonald** s own words to rake in the dough.
> Genius! #voteblue #ImWithHer https://t.co/ejEBxS4XUU

> (178) If looks could kill...she'd be lying on the floor. #EvilDon **#PredatorDon**
> #NastyWomenVote #NastyWoman #AlSmithDinner #HillYes #ImWithHer

Comparing Table 5-11 and Table 5-12, people included just Trump's last name more often than people included just Trump's first name.

Also, looking at the name use in the frequent hashtags in this corpus, people used Hillary Clinton's first name more often than her last name (if only a part of her name showed up in the hashtag). The opposite pattern is true for Donald Trump, where people were more likely to include just his last name, not just his first name. Uscinski and Goren (2010) found a similar pattern with the use of first names in the 2008 democratic primary for presidential election. The coverage of Hillary Clinton used her first name four times more than did the coverage for Barack Obama.

Since this imbalance of first name use happened in both 2008 and 2016 in the elections with Hillary Clinton, this difference in first name use for the female candidate may be due to a purposeful distinction between former President Bill Clinton and Hillary Clinton. Another possible reason for this difference in first name use may be because the candidate actually used her first name on campaign paraphernalia and campaign branding. However, Angeli (2015) found that this imbalance of first name use was present in European elections, as well. Because of this, this imbalance in use of first names and surnames may be due to a gender inequality in the treatment of the candidates.

5.6 Results

In section 5.6.1, I present the results of the analysis in regards to classifying the tweets using *#NastyWoman* as reclaimed or pejorative. In 5.6.2, I give the frequency of some tag reframing hashtags used in the algorithm. In 5.6.3, I show some of the collocations present in the data and hashtags.

5.6.1 Reclamation

Using this data, I was able to code 23,576 tweets using tag reframing hashtags, and study the distribution of tweets over time during the 18 days prior to the election. Quickly after Donald Trump used the adjective *nasty* to describe Hillary Clinton, the hashtag *#NastyWoman* was used in a reclaimed way, shown below in Figure 5-3 as day 1. However, on day 1, others used *#NastyWoman* and other hashtags associated with the Trump campaign to use this hashtag in a manner that was similarly insulting to the female candidate. Both of the two different ways that it was used show different attitudes of the speakers, and other hashtags that showed community membership were used to decipher this attitude. In Figure 5-3, the two attitudes are shown in two different colors, and the fluctuation in use over time as depicted over the 18 days indicates a struggle in the prevalent attitudes expressed using this hashtag.

Distribution of tweets per day
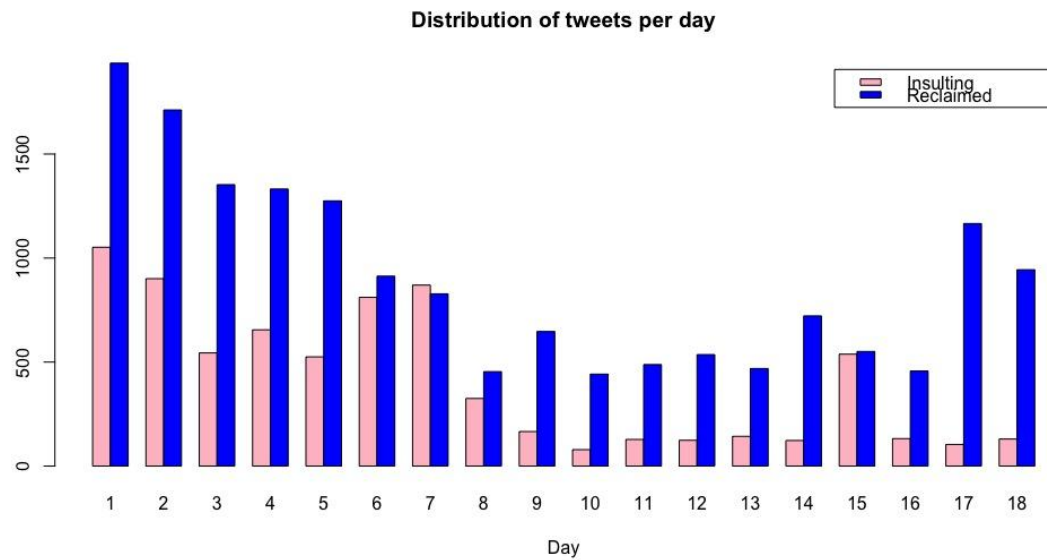
Figure 5- 3: Distribution of tweets per day

While some people may suggest that sexism is no longer relevant because a female candidate made it to the position of presidential candidate, the fact that people had to take back the power of the term *#NastyWoman* suggests that this is not the case, and this graph in Figure 5-3 shows the power struggle in framing this hashtag and using this hashtag for community membership. There were more reclaimed *#NastyWoman* used around days 1 through 5, while days 6 and 7 had more uses of the insulting versions of this hashtag. There was another rise in frequency of reclaimed *#NastyWoman* around days 9 through 14. The insulting and reclaimed versions of *#NastyWoman* were used in varying degrees over the 18 days before the election.

The reclaimed uses of *#NastyWoman* allowed people to show their support for Clinton, identify with the female candidate, and reject the cultural values imposed on Clinton during this election cycle. In using this hashtag to join the larger conversation about the election, the reclaimed

164

version of *#NastyWoman* allows tweeters to claim their own agency in their identifying terminology.

While people often used *#NastyWoman* as a reframing tag with community membership, it was the other hashtags used in the tweets that really clarified the meaning of the tweet, so much so that the hashtags that were clearly associated with the Trump community could clarify the meaning of the tweet and show that the hashtag *#NastyWoman* is being used in a pejorative way. Likewise, other hashtags clearly associated with the Clinton campaign can clarify an otherwise ambiguous tweet, to where readers would know that *#NastyWoman* was being used in a positive, reclaimed way. Since these two communities, with tag reframing hashtags that were based on campaign slogans and promises, used *#NastyWoman* very differently based on the community membership, the algorithm was able to assign a label to the tweet, and automate the process of identifying how *#NastyWoman* was being used in each tweet.

5.7 Conclusions

This chapter shows that tag reframing hashtags using community membership, as described in 4.4.1, can be used to interpret a large data set when the two communities use a term in polar opposite ways.

The hashtag *#NastyWoman* was used by Trump supporters in a negative way, and many hashtags affiliated with that community can be used to code tweets that use more than one hashtag in the tweet as pejorative. On the other hand, Clinton supporters used *#NastyWoman* in a reclaimed, empowering way, and tag reframing hashtags like *#ImWithHer* can be used to interpret the meaning of *#NastyWoman* as such. The reclaimed version strengthens emotion or

165

intimacy, akin to the racial slurs discussed in Yoon (2015). The patterns of reclamation indicate an ongoing power struggle with the use of this term.

By using such a large data set including all the tweets, we can get an idea about the power struggle in the use of reclaimed *#NastyWoman* and pejorative *#NastyWoman*. Using tag reframing hashtags and a large corpus of tweets, I was able to analyze the opposing uses of *#NastyWoman*, and this work contributes to the larger issues with the gendered use of adjectives in the 2016 presidential election.

CHAPTER 6

CONCLUSION

This chapter summarizes the findings in this dissertation and how hashtags fit into the larger
category of discourse markers. It also proposes future studies into the functions of topic marking
and expression of emotion.

6.1 Summary of Findings

In chapter 1, I show how hashtags are discourse markers akin to previously studied DMs, that
fit into the frameworks of characteristics presented in Schiffrin (1987), Fraser (1996), Fraser
(1999), and Schourup (1999). Particularly, the definition of DMs is that they are types of extra
sentential words that are used to organize within a discourse, while revealing context
connections and speaker attitudes. I continue this by setting up the following as criteria for
DMs: DMs must (i) be positioned outside the clause structure; AND one of the following: (ii)
work to signal the coherence relation or position of the utterance to some other part of the
discourse AND/OR (iii) show the speaker's attitude toward an utterance, AND/OR (iv) signal a
lower register (i.e. informality). For each of these criteria, I show through examples and
counterexamples of how DMs fit the criteria, and then show how hashtags also.

In Chapters 2 and 3, I show how hashtags in tweet disclosures act like delaying DMs present in
spoken disclosures. I present environments where these delay devices appear, like when the
abuser or perpetrator is mentioned, where more delays occur. In my previous work with
hashtags, many of the hashtags studied appeared in the tweet final position, and we saw this in

Chapter 3 where 25% of hashtags appear in the final position. However, in Chapter 3, when my data was narrowed to just tweets disclosing violence and abuse, many tweets features a tweet-initial hashtag. This initial placement of the hashtag acts as a hesitation or delay device before information that might be difficult to reveal, akin to *so* (Buysse 2012), *well* (Jucker 1993)*, ano* (Wang 2011), and *nage* (Wang 2011). Furthermore, I looked into the motivations in content for the initial placement hashtags, and initial placement hashtags are used more often when the tweet mentions the abuser in some way. This means that people felt the need to delay when talking about the abuser using a number of different referential expressions.

When the hashtag campaign for domestic abuse awareness began, people were more likely to use initial hashtag placement. Because of the two peaks in usage of the campaign hashtags, the campaign was split into two phases, and I show that people use the hashtag more during the first phase of the campaign, when the conversations about abuse were in the early stages. This difference in placement of hashtag was not motivated by the first uses of the hashtag since before the campaign and the first use of the hashtag with the campaign both feature the hashtag used in the final position.

In Chapter 3, I looked at a number of other motivations in the tweet content for initial hashtag placement, and I found that there was a greater percentage of hashtags used when people used *#whyistayed*. I did not see this same pattern with *#whyileft*. Both mentioning abuser involved a larger percentage initial hashtag placement compared to the general corpus. Some themes, like fear and lack of resources, also had a greater percentage of initial hashtag placement, since fear

was at 49% while lack of resources was around 42%. This is in contrast with the theme responsibility of abuse, which only had 28.6%, which was only marginally larger than the 25% from the general corpus. The tweet content affected how often people used initial hashtag placement when talking about different aspects of survival.

In Chapter 4, I present the next function of hashtags as discourse markers: tag reframing. In this, I show evidence on how these hashtags clarify or reframe the tweet content with a tweet-final hashtag, and in particular, these clarifications happen using the co-presence heuristics discussed in Clark and Marshall (1981). These tag reframing hashtags clarify or reframe the message: (i) by tying into community membership; (ii) by alluding to temporal co-presence; (iii) by referring to early wording via linguistic co-presence; (iv) and through bridging-type connections of indirect co-presence. Some of these heuristics were more salient than others because of the media, like linguistic co-presence, which was the most used category of tag reframing.

This work does not make claims that some types of tag reframing are stronger than others in the same way Clark and Marshall (1981) did with types of repair because their work looked at entire conversations whereas this dissertation looks at individual tweets. However, I somewhat expected the stronger copresence heuristics to be used more often since tweets do not always have the turntaking that comes with spoken conversations, and because of that, I was expecting the frequency in this corpus to reflect the ordering of vertical repair.

Instead, I found a small percentage of temporal copresence tweets, the physical copresence counterpart from Clark and Marshall (1981), which was odd since that is one of the strongest repairs. One reason for less hashtags clarifying using temporal copresence may be the means of communicating with CMC: typically, when people use CMC, they are not in the same physical location as the audience, so referring to salient physical entities is difficult, since it depends on pointing to a particular time where that physical entity will be salient to the speaker and audience.

Linguistic copresence was the most frequent tag reframing, which is indicative of the written nature of hashtags. These two differences in the use are interesting because their frequency may be indicative of how written DMs work on social media, where anaphoric relationships are more available than physical deixis to ensure felicitous communication.

In Chapter 5, I used a corpus of over 55,000 tweets to study reclaimed *#NastyWoman* and pejorative *#NastyWoman*. According to criterion (iii) DMs show the speaker's attitude toward an utterance, and this chapter investigates two very different attitudes expressed through the use of *#NastyWoman*. Using tag reframing hashtags and a learning algorithm, I analyzed the opposing attitudinal uses of *#NastyWoman* and how this contributed to the gendered use of adjectives in the 2016 presidential election.

In the next section, I provide applications building on these findings from this dissertation research, and in section 6.3, I discuss future research into hashtags as discourse markers.

6.2 Future Research

There are two other categories that hashtags fall into besides serving as discourse markers: topic marking and expression. To further study these pragmatic functions of hashtags, one would need to create a specific corpus, like those in Chapter 2 and Chapter 3 in this dissertation, since the following functions are not as common as reframing tags, discussed in Chapter 2.

6.3.1 Introduction of new topics and referents

There is a great deal of research in linguistics on how speakers use various techniques to mark the importance of some content words using extralinguistic material or discourse markers, and this is important since this is one of the functions of hashtags. In the following example, we see how this type of hashtag can occur in the middle of a tweet, and how the hashtag symbol highlights the term after it, making it more prominent to the audience.

(179) Some people really enjoy #exercise. Those people are #crazy.

In fact, this use of hashtags may be related to theories in information structure, specifically those related to the pragmatic terms of topic and focus. As mentioned in Chapter 1, Erman (2001) shows how *you know* "introduc[es] new topics, or topical aspects, thereby moving the text forward". This is related to other research in linguistics: Lambrecht (1994) defines topic as "the matter of [already established] current interest which a statement is about and with respect to which a proposition is to be interpreted as relevant" and focus as "the new information conveyed about a topic"; using these terms, he discusses how prosody can be a means of marking topic and focus in a sentence, for emphasis.

171

Since hashtags function similarly, we will see that one of the pragmatic functions of hashtags is marking a prominence of a new referent (or new topic) or marking an already salient topic for the audience of that tweet. In (180), the word exercise follows the hashtag symbol, and this word is elevated to a keyword.

(180) Some people really enjoy #exercise. Those people are #crazy.

While some may search to see what others are saying about exercise, this Twitter user has the parallel hashtags before exercise and crazy to emphasize these two terms, similar to the act of adding a change of prosody in spoken discourse or using bold in written communication. In the following tweet, we see how the first hashtag acts as an emphatic maker for pragmatic purposes, while the second one acts as tag reframing (see Chapter 4 for this explanation).

(181) Go ahead and give #DonDraper a glass of scotch and is overdue Emmy for being a beauty #madmen

One noticeable difference between these two hashtags includes positioning: while #DonDraper is needed in this sentence, #madmen is independent of the syntax. However, placement is not the only difference since the hashtag before DonDraper is promoting the prominence of the name Don Draper, while the second is the vehicle for a repair using indirect copresence while clarifying the referent Don Draper.

Felicitous use of this type of hashtag is dependent on the content and number of hashtags used. Most examples of this type come before a noun or adjective, as seen in (182) and (183):

(181) Some people really enjoy #exercise. Those people are #crazy.

(183) #delicious home made blackberry kiwi #mojito

Also, unlike the hashtags that signal repair, these emphatic hashtags can only be used a few times. In fact, in this corpus, only one tweet featured this type of hashtag used consecutively, as seen below with #*socialist* and #*revolution*:

(184) #LEFTIST TACTICS: #Change is code-word for creating a #socialist #revolution. #Marxist It's not what you think!

In contrast, in the following meme, we see an example of the emphatic hashtag used where every word in the tweet follows a hashtag:

Figure 6- 1: Hashtag overuse

In this example, the author chooses to not follow the felicity conditions for this type of hashtag to create humor. This is not the same felicity condition for tag reframing hashtags: many of these can be added to the end of the sentence without affecting the felicity of the tweet.

Part of the reason why it is infelicitous to have more than two emphatic hashtags may have to do with how the hashtags are packaging information. In fact, this use of hashtags may be related to some theories in information structure, specifically those related to the pragmatic terms of topic and focus. Lambrecht (1994) defines topic as "the matter of [already established] current interest which a statement is about and with respect to which a proposition is to be interpreted as relevant" and focus as "the new information conveyed about a topic"; using these terms, he discusses how prosody can be a means of marking topic and focus in a sentence, for

emphasis. If we look at example (185), and put into context with this person's earlier tweets, in (186) and (187), we see that the idea of exercise is a topic that is salient to this Twitter users followers:

(185) Some people really enjoy #exercise. Those people are #crazy.

(186) Who loves #spinning as a #workout? Really? Loves?

(187) You know that song that goes "breaking up is hard to do"? Well, we think #workingout is even harder.

Since this person talks about working out in (186) and (187), exercise is something relevant and already established in other tweets, which fits the earlier definition of topic. This would make the hashtag symbol itself a topic marker in (185). The new information, or focus in (185) would be *crazy* since it acts as a "complement of the topic" (Lambrecht 1994).

The function of hashtags involving marking an importance of some key words in the tweet is linked to the original creation of hashtags, but this use has evolved to go beyond just creating a searchable key word: it is used to add emphasis or act in a similar way as capitalization in DML. To study this function of hashtags more rigorously, a larger context may be needed with many of the hashtags that mark topic. Therefore, one will have to collect a number of tweets from different users that use hashtags in this manner.

6.3.2 Expression of emotion or evaluation, like interjections

Some hashtags act more like interjections, or expressive discourse markers, in that they express

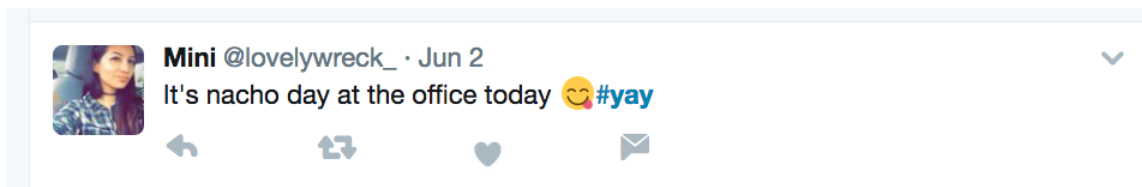some type of emotion or evaluation, as shown in Figure 6-2:

Figure 6- 2: Hashtag interjection

Norrick (2009) discusses "secondary interjections like *damn*, *fuck* and *shit*" as those that

pragmatic markers that can express strong emotions, and a similar hashtag appears in Figure 6-

3.

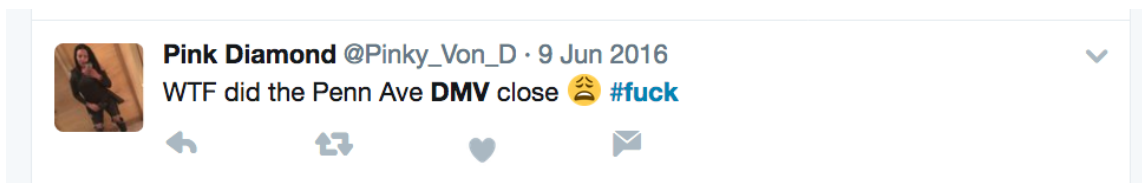*Figure 6- 3: Hashtag strong emotion*

Zappavigna (2015) discusses how these hashtags "enact evaluative meaning" and while she

shows these hashtags do fit nicely into the systems of attitude (using the Martin and White

(2005) terms affect, judgement, and appreciation as categories), these categories fail to capture

how these hashtags are related to other discourse markers; however, by also analyzing these

hashtags by belonging to this class of secondary interjections, it would recognize the relationship

between these expressive hashtags and the other hashtags that function like discourse markers.

Using the systems of attitude as Zappavigna (2015), an example of hashtags involved in

expression of emotion appears as (188) and (189) below, where #yay expresses a positive

emotion of excitement while a negative emotion is expressed with

#SHOOOOOOOOOOOOOOOOOOOTMMMEEEEEEEEEEEEEEEEEEE:

> (188) Might have myself a job at Ice Cream Land (: #yay}

> (189) STILL AWAKE FOR THIS GODDAMN LIT CRIT
> #SHOOOOOOOOOOOOOOOOOOOTMMMEEEEEEEEEEEEEEEEEEE}

Judgment hashtags and appreciation hashtags are two other categories of expression hashtags,

based on Zappavigna (2015). In (190), *#yourteamsucks* is judging an aspect of the marlins fans

in a negative way, so it would fit into this judgment category.

> (190)I feel sorry for marlins fans #yourteamsucks

However, in (191), the *hashtag #useless* fits into the appreciation category because it determines

the quality of the refs.

> (191) In every sport the refs suck #useless

By situating the previous work of Zappavigna (2015) into the area of discourse markers, it

creates a broader analysis for hashtags in that all of them are acting like discourse markers in

different ways. Instead of just noting the expressive nature of the hashtags in the previous work,

the area of discourse markers would account for how hashtags act in larger discourse.

6.3 Applications

There are a number of applications that scholars in other areas can pursue based on the findings of this dissertation.

While this dissertation does not tell someone how to use hashtags prescriptively, advertisers and public relation people can use this research to use hashtags felicitously. In the case of DiGiorno pizza and the #WhyIStayed campaign, the person using the Twitter account for this company failed to look into why #WhyIStayed was trending, as shown in figure 6-4:
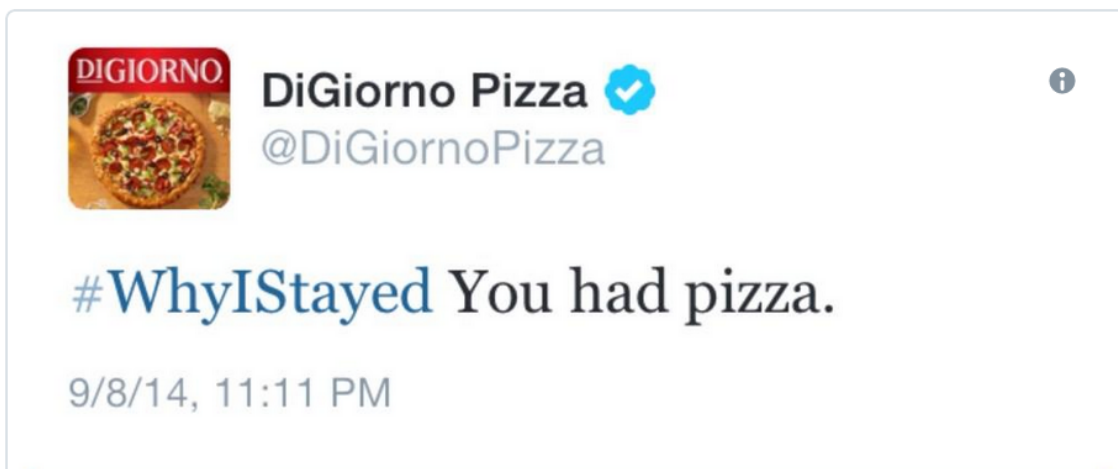
Figure 6- 4: DiGiorno Hashtag

Using the hashtag used for survivor stories, presented in Chapter 3, DiGiorno tried to use a trending hashtag to advertise for the company, but this backfired when people following along with the conversation saw this faux-pas. Knowing how to use hashtags in a felicitous way can assist with advertising and reaching the larger audience available on Twitter without hurting

the company image. Also, knowing how people actually use hashtags can help advertisers and public relations people sound like they know how to use hashtags, even if they are not "native" hashtag users. This could help the tweets sound like natural conversation (think actually listening to your favorite DJ talk about a product they love to a friend as opposed to bad commercials on the radio advertising a product with stilted conversation).

Advertisers are not the only ones who can learn to use hashtags like natural hashtag users. Blattner et al. (2016) talks about how Twitter and hashtags can be used as a tool for learning socio-pragmatics for second language learners, and the work presented in this dissertation can be used to assist with maintaining felicitous use of hashtags in their L2.

Another future study can involve how felicitous hashtag use may be able to allow others to detect the use of bots on the Twitter. If the bots are using them in a way that is dissimilar to the functions mentioned in this dissertation, it can create a natural diagnostic for detecting which tweets are created by an actual person and which are computer generated.

Other fields can use the research presented on tag reframing in Chapter 2 and applying it in Chapter 3 to write a learning algorithm to assist with looking at other phenomena, like reclaimed epithets, via Twitter corpora. Others wanting to study more political tweets could use a specific corpora, like that in Chapter 3, and the information on tag reframing from Chapter 2, to write their own linguistic query. Someone who works in sentiment analysis may be able to use an algorithm that uses tag reframing hashtags, like the algorithm in Chapter 3, to study how people feel in tweets. A sociolinguist could look at the corpus like those collected for Chapter 3 and 4, and collect demographic information on the users to see where people were tweeting

about domestic abuse and political issues. Since the survivor stories in Chapter 3 via tweets and

Chapter 5 via interviews are quite similar, scholars in mental health can look into the usefulness

of tweeting about trauma, akin to journaling and talk therapy.

This dissertation used four corpora to study how hashtags are being used on Twitter,

particularly with respect to information management and delaying disclosure, and other scholars

might use this dissertation to get ideas on how to vary corpora, especially collection

methodology, to investigate their own research queries.

REFERENCES

Angeli, Maria. 2015. Media Reinforcing Patriarchy in Politics. *Gender, Law and Institutions*, 35.

Anthony, Lawrence. 2014. AntConc (Version 3.4.3) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/.

Baiat, G. Esfandiari, Coler, M., Pullen, M., Tienkouw, S., & L. Hunyadi. 2013. Multimodal analysis of "well" as a discourse marker in conversation: A pilot study. In *Cognitive Infocommunications* (CogInfoCom), 2013 *IEEE 4th International Conference* on (pp. 283-288). IEEE.

Baldwin, Tyler & Joyce Y. Chai. 2011. Beyond Normalization: Pragmatics of Word Form in Text Messages. *IJCNLP*, 1437-1441.

Biddle, Sam. 2011. How the Hashtag Is Ruining the English Language (Updated). *Gizmodo*. Gizmodo.com. http://gizmodo.com/5869538/how-the-hashtag-is-ruining-the-english -language (13 June, 2017).

Blattner, Geraldine, Amanda Dalola & Lara Lomicka. 2016. Mind Your Hashtags: A Sociopragmatic Study of Student Interpretations of French Native Speakers' Tweets. Essay. In Congcong Wang & Lisa Winstead (eds.), *Handbook of research on foreign language education in the digital age*, 33–58. Hershey, PA: Information Science Reference, an imprint of IGI Global.

Brinton, Laurel J. 1996. *Pragmatic markers in English: Grammaticalization and discourse functions*. New York: M. de Gruyter.

Brinton, Laurel J. 2010. Discourse Markers. In Andreas H. Jucker & Irma Taavitsainen (eds.) *Historical Pragmatics*, 285-314 Berlin/New York: Mouton.

Brontsema, Robin. 2004. A queer revolution: Reconceptualizing the debate over linguistic reclamation. *Colorado Research in Linguistics*, *17*(1), 1-17.

Francis, W. N., & Kucera, H. 1961. Brown corpus of standard American English. Brown University, Providence, RI, 6.

Browning, Darcey & Kristen Fleckenstein. 2017. Gendered adjectives and the 2016 Presidential Election: A corpus based study. Paper presented at the Cultural Constructions Conference, Arlington, Texas.

Bruns, Axel, Tim Highfield & Jean Burgess. 2013. The Arab Spring and social media audiences: English and Arabic Twitter users and their networks. *American Behavioral Scientist* 57(7). 871-898.

Bruns, Axel & Jean E. Burgess. 2011. The use of Twitter hashtags in the formation of ad hoc publics. In Proceedings of the 6th European Consortium for Political Research (ECPR) General Conference 2011, University of Iceland, Reykjavik. http://eprints.qut.edu.au/46515 (19 November, 2015).

Buysse, Lieven. 2012. So as a multifunctional discourse marker in native and learner speech. *Journal of Pragmatics* 44(13): 1764-1782.

Caleffi, Paola-Maria. 2015.The 'hashtag': A new word or a new rule. *Skase Journal of Theoretical Linguistics*, 12(2).

Constant, Noah., Constant, Noah, Christopher Davis, Christopher Potts, & Florian Schwarz. 2009. The pragmatics of expressive content: Evidence from large corpora. *Sprache und Datenverarbeitung, 33*(1-2), 5-21.

Cepollaro, Bianca. 2015. In defense of a presuppositional account of slurs. *Language Sciences*, 52, 36-45.

Chen, Yiya, & Agnes Weiyun He. 2001. Dui bu dui as a pragmatic marker: evidence from Chinese classroom discourse. *Journal of Pragmatics* 33(9): 1441-1465.

Child, Jeffrey. T., & Starcher, Shawn C. 2016. Fuzzy Facebook privacy boundaries: Exploring mediated lurking, vague-booking, and Facebook privacy management. *Computers in Human Behavior*, 54, 483-490.

Chomsky, Noam. 1981. Lectures on government and binding. *Dordrecht: Foris.*

Chunara, Rumi, Jason R. Andrews & John S. Brownstein. 2012. Social and News Media Enable Estimation of Epidemiological Patterns Early in the 2010 Haitian Cholera Outbreak. *The American Journal of Tropical Medicine and Hygiene* 86(1).39-45.

Clark, Herbert H. & Catherine R. Marshall. 1981. Definite reference and mutual knowledge. *Elements of Discourse Understanding,* 10-63, Cambridge University Press, Cambridge.

Clark, Herbert H. 2002. Speaking in time. *Speech communication, 36*(1), 5-13.

Clark, Rosemary. 2016. "Hope in a hashtag": the discursive activism of# WhyIStayed. *Feminist Media Studies*, 16(5), 788-804.

Counts, Scott & John Geraci. 2005. Incorporating physical co-presence at events into digital social networking. In CHI'05 extended abstracts on Human factors in computing systems (pp. 1308-1311). ACM.

Cravens, Jaclyn D., Jason B. Whiting, & Rola O. Aamar. 2015. Why I Stayed/Left: An Analysis of Voices of Intimate Partner Violence on Social Media. *Contemporary Family Therapy,* 37(4), 372-385.

Crystal, David. 2011. *Internet linguistics: A student guide.* Routledge.

Davis, Bud. 2013. Hashtag Politics: The Polyphonic Revolution of # Twitter. *Pepperdine Journal of Communication Research* 1(1). 4.

de Moor, Aldo. 2010. Conversations in context: a twitter case for social media systems design. In *Proceedings of the 6th International Conference on Semantic Systems* (p. 29). ACM.

Dorgeloh, Heidrun. 2004. Conjunction in sentence and discourse: sentence- initial and and discourse structure. *Journal of Pragmatics* 36. 1761-1779.

Dresner, Eli & Susan. C. Herring. 2010. Functions of the Nonverbal in CMC: Emoticons and Illocutionary Force. Communication Theory 20(3). 249.

Edwards, Autumn, & Harris, Autum. J. (2016). To tweet or 'subtweet'?: Impacts of social networking post directness and valence on interpersonal impressions. *Computers in Human Behavior*, 63, 304-310.

Erard, Michael. 2008. *Um. . .: Slips, Stumbles, and Verbal Blunders, and What They Mean.* New York, NY: Anchor Books.

Erman, Britt. 2001. Pragmatic markers revisited with a focus on you know in adult and adolescent talk. *Journal of Pragmatics, 33*(9), 1337-1359.

Erzikova, Elina., Gandy, Lisa, Hall, Adrienne, & Kuntz, Kara. 2016. The 2014 NFL Player Ray Rice Domestic Abuse Case: An Analysis of Factors That Contributed to Tweet Popularity During the Scandal. In 10th European Conference on Information Systems Management: ECISM 2016 (p. 75-82).

Fraser, Bruce. 1988. Types of English discourse markers. Acta Linguistica Hungarica, 38(1-4), 19-33.

Fraser, Bruce. 1996. Pragmatic markers. *Pragmatics* 6: 167-190.

Fraser, Bruce. 1999. What are discourse markers?. *Journal of Pragmatics* 31(7): 931-952.

Fraser, Bruce. 2009. Topic Orientation Markers. *Journal of Pragmatics*, 41(5), 892-898.

Freelon, Freelon, Deen, Charlton D. McIlwain, & Meredith D. Clark. 2016. Beyond the hashtags:# Ferguson,# Blacklivesmatter, and the online struggle for offline justice.

Gunraj, Danielle N., April M. Drumm-Hewitt, Erica M. Dashow, Sri Siddhi N. Upadhyay, and Celia M. Klin. 2016. Texting insincerely: The role of the period in text messaging. *Computers in Human Behavior*, 55, 1067-1075.

Han, Donghong. 2011. Utterance production and interpretation: A discourse-pragmatic study on pragmatic markers in English public speeches. *Journal of Pragmatics* 43: 2776-2794.

Harvey, Kevin. 2012. Disclosures of depression: Using corpus linguistics methods to examine young people's online health concerns. *International Journal of Corpus Linguistics* 17(3). 349-379.

Haselow, Alexander. 2011. Discourse marker and modal particle: The functions of utterance-final *then* in spoken English. *Journal of Pragmatics*, 43(14), 3603-3623.

Horeck, Tanya. 2014. #AskThicke: "Blurred Lines," Rape Culture, and the Feminist Hashtag Takeover. *Feminist Media Studies* 14(6).1105-1107.

Horwitz, Sally. 2007. Sally Horwitz - June 18, 2007. *Voice/Vision Holocaust Survivor Oral History Archive*. http://holocaust.umd.umich.edu/horwitz/ (30 March, 2015).

Huang, Yan. 1991. A neo-Gricean pragmatic theory of anaphora. *Journal of Linguistics*, *27*(02), 301-335.

Huang, Yan. 2000. *Anaphora: A cross-linguistic approach*. Oxford University Press on Demand.

Illocution Inc. 2014. http://illocutioninc.com/site/products-data.html (5 April 2013).

Jucker, Andreas. H. 1993. The discourse marker *well*: A relevance-theoretical account. J*ournal of Pragmatics* 19(5): 435-452.

Jucker, Andreas. H., & Irma Taavitsainen. 2000. Diachronic speech act analysis: Insults from flyting to flaming. *Journal of Historical Pragmatics*, *1*(1), 67-95.

Jucker, Andreas H. & Yael Ziv. 1998. Discourse markers: Introduction. Essay. In Andreas H. Jucker & Yael Ziv (ed.), *Discourse markers: descriptions and theory*, 1–12. Amsterdam: Benjamins.

Kalman, Yoram M., & Darren Gergle. (2014). Letter repetitions in computer-mediated communication: A unique link between spoken and online language. *Computers in Human Behavior*, 34, 187-193.

Kleinman, Sherryl, Matthew Ezzell, & Corey Frost. 2009. Reclaiming critical analysis: The social harms of "bitch". *Sociological Analysis*, *3*(1), 46-68.

Kratzer, Angelika. 2009. Making a pronoun: Fake indexicals as windows into the properties of pronouns. *Linguistic Inquiry*, *40*(2), 187-237.

Lambrecht, Knud. 1994. *Information structure and sentence form: A theory of topic, focus, and the mental representations of discourse referents.* Cambridge: Cambridge University Press.

Louwerse, Max M., and Heather Hite Mitchell. 2003. Toward a taxonomy of a set of discourse markers in dialog: A theoretical and computational linguistic account. *Discourse processes*, 35(3), 199-239.

Levinson, Steven. C. 1987. Pragmatics and the grammar of anaphora: A partial pragmatic reduction of binding and control phenomena. *Journal of Linguistics*, *23*(02), 379-434.

Matson, Erin. 2016. Feminist Over-Sharing in the Wake of the Ray Rice Scandal. In *Scandal in a Digital Age* (pp. 201-207). Palgrave Macmillan US.

McCreadie, Richard, Ian Soboroff, Jimmy Lin, Craig Macdonald, Iadh Ounis, and Dean McCullough. 2012. On building a reusable Twitter corpus. In Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval (pp. 1113-1114). ACM.

McMinn, Andrew J., Yashar Moshfeghi, and Joemon M. Jose. 2013. Building a large-scale corpus for evaluating event detection on twitter. In Proceedings of the 22nd ACM international conference on Information & Knowledge Management (pp. 409-418). ACM.

Merrigan, Gerianne. 2000. Negotiating Personal Identities Among People With and Without Identified Disabilities: The Role of Identity Management. In Dawn O Braithwaite & Teresa L Thompson (eds.), Handbook of Communication and People with Disabilities: Research and Application. United States: Lawrence Erlbaum Associates, Publishers.

Müller, Simone. 2005. *Discourse markers in native and non-native English discourse.* Amsterdam: John Benjamins.

OVC TTAC. 2015. Victim Impact: Listen and Learn-- Presenter's Toolbox. Office for Victims of Crime. https: //www.ovcttac.gov/victimimpact/presenters_toolbox.cfm (30 March, 2015).

Page, Ruth. 2012. The linguistics of self-branding and micro-celebrity in Twitter: The role of hashtags. *Discourse & Communication* 6(2). 181-201.

Petrovic, Sasa, Miles Osborne & Victor Lavrenko. 2010. The edinburgh twitter corpus. In Proceedings of the NAACL HLT 2010 Workshop on Computational Linguistics in a World of Social Media (pp. 25-26).

Pittman, Matthew, & Alec C. Tefertiller. 2015. With or without you: Connected viewing and co-viewing Twitter activity for traditional appointment and asynchronous broadcast television models. *First Monday*, 20(7).

Potts, Liza, Joyce Seitzinger, Dave Jones & Angela Harrison. 2011. Tweeting Disaster: Hashtag Constructions and Collisions. Proceedings of the 29th ACM International Conference on Design of Communication, 235-240.

Prince, Ellen F. 1981. Toward a taxonomy of given-new information. *Radical pragmatics.*

Romano, Manuela. 2014. Evaluation in emotion narratives. *Evaluation in Context* 242. 367.

Romero, Daniel M., Brendan Meeder & Jon Kleinberg. 2011. Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter. Proceedings of the 20th international conference on World wide web, 695-704. ACM. http://dl.acm.org/citation.cfm?id=1963503 (19 November, 2015).

Schirra, Steven, Huan Sun, & Frank Bentley. 2014. Together alone: motivations for live-tweeting a television series. In Proceedings of the 32nd annual ACM conference on Human factors in computing systems (pp. 2441-2450). ACM.

Schiffrin, Deborah. 1986. Functions of *and* in discourse. *Journal of Pragmatics*, *10*(1), 41-66.

Schiffrin, Deborah. 1987. *Discourse markers*. Cambridge: Cambridge University Press.

Schiffrin, Deborah. 2001. Language and public memorial: 'America's concentration camps.' Discourse and Society 12: 505- 534.

Schiffrin, Deborah. 2001. Language, experience and history: 'What happened' in World War II. Journal of Sociolinguistics 5 (3): 323- 352.

Schiffrin, Deborah. 2002. Mother and friends in a Holocaust life story. *Language in Society*, 31(3), 309-353.

Schiffrin, Deborah. 2003. Linguistics and oral history: The discourse connection. In D. Tannen (Ed.) Discourse and Beyond, Washington D.C: Georgetown University Press.

Schiffrin, Deborah. 2003. 'We knew that's it:' Retelling the turning point of a narrative of danger. Discourse Studies.

Schiffrin, Deborah. 2006. Discourse marker research and theory: revisiting and. Essay. In Kerstin Fischer (ed.), *Approaches to discourse particles*, 315–338. Amsterdam: Elsevier.

Schourup, Lawrence. 1999. Discourse markers. *Lingua*, 107(3-4), 227-265.

Scott, Kate. 2015. The pragmatics of hashtags: Inference and conversational style on Twitter. *Journal of Pragmatics*, *81*, 8-20.

Sharma, Sanjay. 2013. Black Twitter?: Racial hashtags, networks and contagion. N*ew Formations: a journal of culture/theory/politics* 78(1).46:64.

Shaw, Risa S. 2007. Meaning in context: The role of context and language in narratives of disclosure of sibling sexual assault. Cincinnati, Ohio: Union Institute & University dissertation.

Small, Tamara A. 2011. What the hashtag? A content analysis of Canadian politics on Twitter. Information, *Communication & Society* 14(6). 872-895.

Stvan, Laurel. 2006. Diachronic change in the uses of the discourse markers why and say in American English. *Linguistic Insights - Studies in Language and Communication* 25:61-76.

Sutton, Laurel. A. 1995. Bitches and Skanky Hobags: The Place of Women in Contemporary Slang. In Kira Hall and Mary Bucholtz (eds.) *Gender Articulated: Language and the Socially Constructed Self.*

Tanskanen, Sanna-Kaisa & Johanna Karhukorpi. 2008. Concessive Repair and negotiation of affiliation in email discourse. *Journal of Pragmatics* 40(9). 1587-1600.

Tatman, Rachael. 2015. # go awn: Sociophonetic Variation in Variant Spellings on # Twitter. *Working Papers of the Linguistics Circle,* 25(2), 97-108.

Tree, Jean E. Fox, and Josef C. Schrock. 1999. Discourse markers in spontaneous speech: Oh what a difference an oh makes. Journal of Memory and Language, 40(2), 280-295.

Tree, Jean E. Fox, and Josef C. Schrock. 2002. Basic meanings of you know and I mean. Journal of Pragmatics, 34(6), 727-747.

Uscinski, Joseph E., & Lilly J. Goren. 2011. What's in a name? Coverage of Senator Hillary Clinton during the 2008 Democratic Primary. *Political Research Quarterly*, 64(4), 884-896.

Vandergriff, Ilona. 2013. Emotive communication online: A contextual analysis of computer-mediated communication (CMC) cues. Journal of Pragmatics, 51, 1-12.

Van Olmen, Daniël. 2013. The imperative of say as a pragmatic marker in English and Dutch. *Journal of Germanic Linguistics* 25(3): 247-287.

Vine, George. 1983. George Vine - July 5, 1983. Voice/Vision Holocaust Survivor Oral History Archive. http://holocaust.umd.umich.edu/vineg/ (30 March, 2015).

Wang, Yan. 2011. A Discourse-Pragmatic Functional Study of the Discourse Markers Japanese *Ano* and Chinese *Nage. Intercultural Communication Studies* 20 (2).41-61.

Weathers, Melinda R., Jimmy Sanderson, Alex Neal, & Kelly Gramlich. 2016. From Silence to #WhyIStayed: Locating Our Stories and Finding Our Voices. *Qualitative Research Reports in Communication,* 17(1), 60-67.

WTNH News8. 2012. Rape victim speaks out about attack. *YouTube.* https://www.youtube.com/watch?v=-BcFdw25WHI/ (30 March, 2015).

Yang, Guobin. 2016. Narrative agency in hashtag activism: The case of# BlackLivesMatter. Media and Communication, 4(4), 13-17.

Yus, Francisco. 2011. *Cyberpragmatics: internet-mediated communication in context.* Amsterdam: John Benjamins Pub. Co.

Yoon, Suwon. 2015. Semantic constraint and pragmatic nonconformity for expressives: compatibility condition on slurs, epithets, anti-honorifics, intensifiers, and mitigators. Special issue on Slurs at *Language Sciences* 52: 46-69.

Zappavigna, Michele. 2012. *Discourse of Twitter and Social Media: How We Use Language to Create Affiliation on the Web.* (Continuum Discourse Series). London: Bloomsbury Publishing.

Zappavigna, Michele. 2015. Searchable talk: the linguistic functions of hashtags. *Social Semiotics* 25(3). 274-291.