Why is that a tag?!:
User-generated tag structure in online fan fiction archives.

by

Ella Hitchcock

A thesis submitted in partial fulfilment of the requirements for the degree of

Master of Arts in Digital Humanities

and

Master of Library and Information Studies

Digital Humanities and Library and Information Studies
University of Alberta

Abstract

Established in 2007 by the Organization for Transformative Works, the Archive of Our Own (AO3) is a fan fiction archive that hosts over three million fan works, consisting mostly of fan fiction. It has become an active hub of fan activity, making it an ideal object of study as a current representation of fan attitudes and trends. This thesis will examine the linguistic features of user-generated tags in fan fiction archives across three different types of media: television, print, and film. It proposes that there are similarities in user-generated tags that cross organizational boundaries on AO3 and that the linguistic features of user-generated tags change across category (hereafter called fandom) boundaries in ways that show user engagement with the type of media for which they are writing. Data was collected from three related fandoms: *Hannibal (TV)*, *Hannibal Lecter Tetralogy - Thomas Harris*, and *Hannibal Lecter (Hopkins Movies)*. It was then run through feature analysis software to determine how tag structure changes across fandom boundaries. This analysis is aimed at revealing relationships between fandoms through the similarities and differences of tags used in each fandom and showing how users engage with the property through the complexity of the tag structures in each fandom. The results revealed that while user-generated tags on AO3 contain distinct linguistic features, there are minimal differences in tag content between the three fandoms, and that while the majority of the tags analysed consist of short, simple noun phrases, there are enough longer, more complex tags to indicate that users engage and communicate with other users through the tags themselves.

Acknowledgements

I would like to thank my supervisors, Dr. Ali Shiri and Dr. Geoffrey Rockwell for their guidance and support, without them this thesis would not have been finished. I would also like to thank my first reader, Dr. Harvey Quamen for taking the time to provide much needed feedback.

I would like to thank my friends and family for supporting me through this time and all the great people at both UAL and the Internet Archive who have been so supportive and patient with me.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

I have been reading and writing fan fiction for over a decade. In that time fan spaces have evolved, growing larger and more mainstream. My first introduction to fan fiction and the communities that surround them were sites like Livejournal and Fanfiction.net. Here I could moan about the *Sherlock* Series Two finale with other people, discuss *Game of Thrones* character motivations with other fans of the show, and most importantly, think and write about my favorite characters and shows. Even then though, I would occasionally get frustrated over how difficult it was to find the fiction that I wanted to read. Browsing long lists of fan fiction, so called 'fic', could be fun, but sometimes searching pages and pages of comments for a good recommendation, or following link after link on a Livejournal blog looking for that one fic got a little tiring. I eventually followed the crowd from Livejournal to newer spaces like Tumblr and AO3 and immediately noticed differences in how these spaces were organized. Instead of more conventional modes of organization such as indexes or navigation menus, these sites were built around tags created by users, not administrators. The users were the ones creating searchability within fan spaces. This sparked my interest in how fans could form their own spaces online and how they organized those spaces when given the freedom to do so.

This thesis is divided into five chapters, organized as follows. Chapter Two will contain a brief history of fan fiction, the shift of fan activity to digital spaces, the development of social tagging, and how social tagging has been used in fan spaces. Chapter Three contains the data collection methods, data cleaning methods, data analysis methods and theoretical framework underlying this research. Chapter Four will be an analysis of the data collected based on analytic methods laid out in chapter three and contain summaries of the three data sets and various data

visualizations. Finally, Chapter Five discusses the results of this research and how this research fits into the wider discussion of fan spaces and the development of social tagging practices in digital communities.

Fandom studies is a relatively young field and the study of fandom in digital contexts is even younger. This despite the fact that over the past ten years media that was previously relegated to the fringes of popular culture has become part of mainstream culture. This is most obvious with the popularity of the Marvel Cinematic Universe, but even less popular media is gaining an audience due to the ubiquity of digital platforms. Fan activity has been part of the digital landscape from the early days of Usenet to the blogging of Livejournal and the more explicitly fan oriented platforms like fanfiction.net and AO3. While there has been research conducted on the ways in which fandom has adapted itself to digital spaces, there has been little research on the ways that fans are organizing their work and their communities in practice. This research aims to make a contribution to this thus far understudied area of fandom studies. The objectives of this research are to explore the tagging environment of AO3, and to examine how users tag their work in a space built by and for themselves. To this end, the research questions are as follows:

1.) What linguistic features exist in user-generated tags across fandoms on AO3?

2.) What similarities and differences exist in user-generated tags across fandoms (television, books, and movies) on AO3?

3.) What can we infer about discourse in the fan fiction communities from these user-generated tags?

It should be noted that this study has limitations. The study was conducted on only three fandoms and as such this study cannot be said to be a comprehensive review of AO3. It is also

true that while the popularity of the Hannibal television show in fan spaces (see chapter 3) makes the Hannibal fandom a good choice for study given its relatively small size, the community *is* still relatively small, which makes drawing sweeping conclusions difficult. Despite this, I hope that the data analysed in this thesis and the conclusions drawn further the field of fan studies.

Chapter 2: Literature Review

Section 1. A short history of fandom from print to digital.

Section 1.1. What is fan fiction?

Fan fiction is defined as works of fiction written by fans of a particular entertainment property that uses the setting or characters of that property to tell an original story (Kem, 2005). Fan fiction in its modern form has been around since the 1970s, starting with the *Star Trek* fan magazines and fan conventions, which allowed fans to share their stories, art, and theories. This social aspect of fandom has been amplified by the internet and fan fiction has largely moved into the digital sphere in the last fifteen years (Brydges et al., 2017). Due to a wide variety of topics and genres, fan fiction is often organized by the intellectual property fans are writing about, as exemplified by the fandoms that this thesis will be examining. Fan fiction writing is often a social activity and a collaborative one, as fans collaborate and communicate through social media and commenting on fan fiction (Jensen, 2014).

This thesis focuses on fan activity and fan fiction in western culture, I would be remiss if I did not at least mention fandom and fan fiction in other cultures. Japanese anime and manga have a rich culture of fan creation, both online and offline. Doujin, or fan created comics, are a huge part of anime and manga fandoms and are shared in digital and physical form.

Section 1.2. A short history of fandom and fan fiction in the west

Fandom has a long and storied history. Depending on who you ask fandom started at a few different points. Sherlock Holmes aficionados would claim that Sherlock Holmes is “the first fandom” (Coppa, 219) as it was one of the first western fan communities that fits our modern definition of fandom: a community of people coming together to share and discuss their interests

in a particular piece of culture or media. Others see the rise of the *Star Trek* fan community as the first large scale fandom, not only for the level of organization involved (such as fan conventions and the distribution of fanzines), but for the large amount of fan creativity and production that took place in the community. The *Star Trek* fandom was the first to distribute and share fanworks on a large scale (or at least the first whose sharing efforts garnered large amounts of attention from mainstream culture). The rise of modern fan studies coincided with the airing of *Star Trek: The Original Series* and thus the fan productivity of the *Star Trek* fandom has had a large amount of influence over fan studies. The *Star Trek* fandom saw the first fan conventions, the first fanzines, the creation of 'shipping' (the romantic pairing of two characters), and it popularized fan fiction writing as a method of fan engagement.

Section 1.3 History of fandom and fan fiction on the internet

The distribution of fan fiction before the internet was in many ways more decentralized than it is now. Fans would pass individual stories around small social groups, trade stories at conventions, and submit to fanzines that would circulate their work, but there were no online archives or fan blogs yet. Fans were siloed by location, social connections, and television program schedules. This separation did have the side effect of creating networks of fans, but in general sharing fanworks was much more difficult.

The internet, even before it was ubiquitous, was an ideal platform for fandom and fan communities. Fans flocked to the spaces created in the early days of networked computer (Usenet | Fanlore.org). Early discussion spaces like Usenet provided fans with access to discussion boards on a variety of topics in the 1980s and 1990s. More importantly, Usenet provided fans with a shared space that they could use to share their work (Usenet | Fanlore.org). As the internet gained a wider audience in the mid 1990s, fandom spread across various

websites. Various fan forums and archives sprung up, from small fan sites to what would eventually become large fan hubs like fanfiction.net. Fandom moved towards a more centralized model in the early 2000s, prompted in large part by the increasing popularity of personal blogs. Started in 1999 as a personal blogging platform, Livejournal (LJ) had already been in use for a few years as a personal blogging platform when fans began to migrate in the early 2000s. It quickly became a fandom hub, hosting hundreds if not thousands of fan blogs. Part of the popularity of LJ can be attributed to the site's social functionality. Users could create and personalize their own blogs, connect with other users through the use of friend lists, commenting, and RSS feeds (Livejournal | fanlore.org).

While LJ was rising in popularity as a more general platform for fandom, fanfiction.net was becoming a central hub for fan fiction online. Started in 1998 as a place for fans to aggregate their fan fiction, the site is still one of the largest fan fiction archives on the internet in terms of number of works hosted (fanfiction.net | fanlore.org). While it is still an active hub of fan activity today, it has dropped in popularity over the years. Various factors are responsible for this drop in popularity, not least of which is the large amount of control that site administrators maintain over the content posted on the site.

For a long time, fanfiction.net was the premier fan fiction repository online. The site was started in 1998 by Xing Li, a Los Angeles based developer and X-files fan. It was designed to be an archive, not a social space (fanfiction.net | Fanlore.org). This made sense at the time, as other community spaces existed (Usenet groups, Livejournal, small fansites, etc.). Community building infrastructure was not the primary focus of the site, although it was later enabled through discussion forums and user groups (fanfiction.net | Fanlore.org). Despite its popularity, fanfiction.net's cataloguing system has always been a sticking point for fans. Fanfiction.net uses

an authority-controlled system to organize the material hosted on the site and the site maintains a large amount of control over what gets posted by users. Over its history there have been many controversies surrounding the type of content that is allowed on the site. The site does not host fan fiction works based on particular creators (e.g., George R. R. Martin or Anne Rice), respecting requests from authors that fans not remix their work. The site also has a loose ban on RPF (Real Person Fiction), that is, fan fiction written about real people. The most controversial restriction on content was the banning of explicit and NC-17 works, which took place twice (once in 2002 and again in 2012) (fanfiction.net | Fanlore.org). This ban drove a large number of fans to look for other places to post their work.

Several events shook fandom in the mid to late 2000s. In 2007, Livejournal, then one of the largest centres for fan communities, permanently suspended over 500 blogs over perceived inappropriate content in the user's interest lists, a move that later became known as 'Strikethrough' in the fan community (due to the suspended blog names being crossed out). This move affected several large fan communities and raised questions about the safety and stability of fan communities online. 'Strikethough' was followed by another second suspension campaign, 'Boldthrough' that affected even more fan blogs (Livejournal | fanlore.org). These suspensions occurred around the same time as the rise of Fanlib, a website that, in theory, was a space for fans to share their work. However, the site was partnered with various copyright owners and the terms of service stated that by posting fanwork on the site users gave Fanlib the right to "use, reproduce, distribute and publicly display Your Content on the website or through its services (such as email notification and RSS feeds) free of charge." (Fanlib Terms of Service Section 8). These terms, along with the fact that the site was partnered with copyright owners, meant that Fanlib was seen by the fan community as a group of outsiders attempting to profit off

of fan labour. These events brought into focus the need for a space for fans to post and share their work free of outside pressure, were a major factor in the fan migration to Tumblr and were directly responsible for the creation of AO3.

The idea (and the name) of AO3 was proposed by Astolat in 2007. Astolat, a Big Name Fan (fandom's term for a prominent individual either in a fandom or in the general fan community), wrote a blog post titled *An Archive of One's Own*. In the post she stated the need for an archive run by fans for fans and outlined an idea of a fan space that would eventually become the Archive of Our Own, now better known as AO3 (Astolat, 2007). The most important feature of this new space would be that it would be "run BY fanfic readers FOR fanfic readers" (emphasis mine). Other important features were that it be community supported, highly searchable, user oriented, allow comments, and that it "would allow ANYTHING -- het, slash, RPF, chan, kink, highly adult" (Astolat, 2007). In the context of the mature content ban on fanfiction.net and the deletion of fan blogs on Livejournal, the complete freedom of such a platform quickly became extremely popular. AO3 came online in 2007 through the efforts of the Organization for Transformative Works (OTW), a non-profit organization established to "serve the interests of fans by providing access to and preserving the history of fanworks and fan culture in its myriad forms" (About the OTW | AO3). These values are reflected in the structure of AO3 (Fiesler, Morrison, and Bruckman, 2016). The features that Astolat outlined in her initial post have become the defining features of AO3, the site code is open source, it is maintained by donations, any and all content is allowed, and the multiple iterations of the robust tagging system have ensured a level of findability that is not present on other fan archives.

Section 2. Taxonomies used in online content

Section 2.1. Discussion of the online organization methods

To understand the impact of AO3's organization system, some further context is needed. The internet started out as a series of networked computers developed by the United States Department of Defence in the 1960s. These networks were closed systems that sent packets of information from one node in the network to another, information was not hosted 'online' in any sense (Hafner & Lyon, 1998). Even after the internet went live and started gathering a large user community, the focus was on uploading content. Users could upload content but not edit, share, and collaborate once the content was uploaded. The network was shared, but not collaborative. This changed with the transition of Web 1.0 to Web 2.0, the so-called 'social web'. Web 2.0 saw the rise of blogging, which then led to the rise of social media, mirroring the ways in which the internet evolved from focusing on products to focusing on consumers. In their retrospective on the first Web 2.0 conference, *What is Web 2.0*, O'Reilly discusses how companies survived the transition from Web 1.0 to 2.0 "embraced the power of the web to harness collective intelligence", that is, turned to the collective activity of their users to create content (O'Reilly, *What is Web 2.0*).

At their most basic, taxonomies are structured systems used to organize information in useful ways. The ways in which taxonomies are structured depends on the type of information they are organizing and the purpose for which they are built. For the purpose of this discussion, we will be discussing cataloguing taxonomies ranging from authority control to collaborative tagging and how these structures are used in fan spaces.

As a system that relies the most on established cataloguing methods and implements a hierarchical, top-down cataloguing structure, authority control is often the most useful system for organizing vast amounts of information from disparate sources. However, authority control

comes with a number of caveats. The first, and most relevant to this discussion, is the lack of self-expression that is provided to users of this system. In an environment in which the user is the focus, authority-controlled classification systems impose unnecessary restrictions on self-expression. On the opposite end of the spectrum is collaborative tagging, a system in which users create the cataloguing system from the bottom-up.

Tagging is a method of information classification that allows users to annotate digital items with keywords (tags). It has become a ubiquitous part of the Internet landscape in the years since it was introduced to the Web with the shift toward Web 2.0 which encouraged users to create their own content. This shift gave rise to collaborative tagging (Moulaison, 2008). As the Internet has become more focused on social networking, collaborative tagging (also known as social tagging) has become more and more popular.

Collaborative tagging is an organizational structure in which users generate their own keywords to classify their work instead of choosing keywords from a predefined list to classify their content. These user-generated keywords provide information about the item that can be used for findability purposes and content analysis by other users. These keywords can be highly idiosyncratic, but they are also public, allowing other users to use these tags as a findability aid (Macgregor & McCulloch, 2006; Goh, Chua, Lee, & Razikin, 2009) and as a content indicator (Choi & Syn, 2012).

Guy and Tonkin (2006) state, “the tagging terms used in those [collaborative tagging] systems are imprecise. It is the users of a folksonomy system who add the tags, which means that tags are often ambiguous, overly personalized, and inexact” (Guy & Tonkin 2006, 1-2). This in-exactitude causes problems for traditional information retrieval methods, as tagging systems usually have no controlled vocabulary and very little in the way of hierarchical structure

(Marlow et al., 2006). However, it has been shown that when such tagging systems are formalized, they can enhance retrieval results (Morrison, 2007; Lawrence & Schraefel, 2006) and improve collection accessibility (Lu, Park, & Hu, 2008). Regularities in user activity also develop in tagging environments, and the emergence of stable patterns has shown that minority tags can co-exist with popular ones without disrupting the system as a whole (Golder & Huberman, 2006). Research has also shown that even well-organized authority control systems can be enhanced by the incorporation of collaborative tagging (Daglas et al., 2012).

An interesting and relevant avenue of research into collaborative tagging has also commented on the presence of statistical patterns found in language, the most relevant of these being Zipf's Law (Zipf, 1935). Zipf's law states that the frequency of a word is inversely proportional to its frequency rank, that is, that the most frequent word (or tag, in this case) will occur roughly twice as much as the next most frequent tag. It is characterized in part by the presence of a 'long tail' of data that trails far from the point of highest frequency. The presence of Zipf's has been observed in other collaborative tagging systems and is sometimes used as evidence of their utility (Cattuto, Loreto, and Pietronero, 2007; Mathes, 2004; Jorgensen, 2004; Guy and Tonkin, 2006).

Collaborative tagging practices have become an important part of cataloging systems across the Internet, from academic databases to social sites such as del.ici.ous and Twitter (Daglas et al., 2012; Choi & Syn, 2016). There is a growing body of evidence pointing to the development of consistent tagging practices in collaborative tagging over time (Eynard, et. al., 2013; Kipp, 2007). Kipp (2007) conducted a study on tags on three social bookmarking tools, Del.icio.us, Connotea, and Citeulike and concluded that users were separating tags into a few different categories. Most notably, they observed that there was a separation between non-subject

tags that were used for affective purposes (describing emotional states) and subject tags that were used for more descriptive purposes (used for findability). The prevalence of non-subject tags that was discovered by this study can be used as evidence that users see classification as a more holistic process than it is traditionally considered. While there is a large body of literature on social tagging in social media contexts there is very little on its use in more insular environments, such as fan spaces.

Section 3. The use of tagging in online fan spaces

Tagging has been used in a few different ways in fan spaces. Organization systems used in fandom vary considerably in their approach to tagging (Johnson, 2014). The three popular fan spaces that focus on fan fiction over the last 15 years, Livejournal, fanfiction.net, and AO3, use varying levels of authority control to achieve their goals and enforce various levels of hierarchical structure.

Section 3.1. Fanfiction.net and authority control

As discussed in the first section of this chapter, fanfiction.net was designed as a more traditional archive rather than as a social space and this is reflected in its classification structure. The site does not use a tagging system at all, instead using an authority-controlled system (Johnson, 2014). Fan works are organized by media type (see Fig. 1) and then by fandom within media type. Figure 1 depicts the main navigation bar of fanfiction.net. As we can see, fanfiction categories (organized alphabetically from left to right) are laid out for the user, which defines the ways in which users navigate the site. This layout also affects the experience of various fandoms, as this organization separates fandoms that have multiple iterations across various types of media. For example, the Harry Potter franchise is split into two categories based on media type: ‘Movies’ and ‘Books’. This forces authors to choose one category or the other and thus limit

their audience.

Figure 1: Organization of content on fanfiction.net homepage

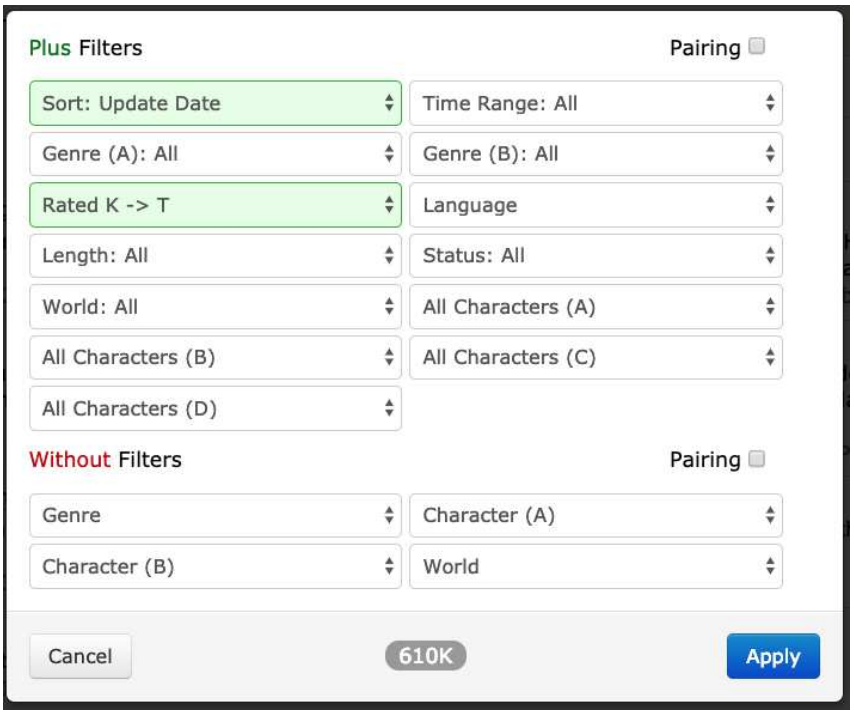


Figure 2: Search filters within a fandom on fanfiction.net

While the homepage search bar site presents a search option to users, the majority of searching within fandoms takes place on fan work list pages. The filtering options available on fan work lists are a combination of general and specific (see Fig. 2). Filters such as ‘Language’, ‘Rating’, ‘Length’, and ‘Status’ are general, while filters like ‘All Characters(A-D)’ and ‘World’

are specific to a fandom. This filtering system is limited by a number of things, the most obvious of which is the character filters. A fan fiction author may include more than four characters in her work, yet the filtering system will only allow four characters in the work description. This means that if a user wants to find a list of works based around a favourite character, they might not be able to find all of the works that include that character because the author did not include them in her work description.

Another limitation of this authority-controlled filtering system is the 'Genre' category. The 'Genre' filters reflect a mainstream idea of genre. Only five genres out of the twenty-one available in the filter reflect fan culture (these genres are 'Angst', 'Hurt/Comfort', 'Friendship', 'Family', and 'General'), the rest of the genres that fans can used to categorize their work are reflections of traditional genres (such as 'Horror', 'Western', 'Sci-Fi', 'Romance', or 'Adventure'). In an environment where narrative is often less important than character and genre conventions are often ignored, this authority-controlled cataloguing system is insufficient.

Section 3.2. Livejournal's tagging system.

From an information science perspective, Livejournal uses the most traditional tagging system, that is, a completely uncontrolled system. Users can write whatever they want in their tags. There are no restrictions on what users can tag their content with and there is no effort to standardize the tags in any way (Johnson, 2014). While this method is suited to fandom in some ways, it fails in others. Tags in fan spaces on Livejournal are most often used to signify occurrences and spaces that are already standardized in the fan community. A tag like 'kinkmeme' is a good example of this. Kinkmemes are blogs in which fans share fan fiction. They are fandom-specific and consist of a single long comment thread. Users request a topic, pairing, or situation and other users write fan fiction based on these requests. These 'prompts'

and ‘fills’ as they are called, are all contained in one long thread of comments. In this way users can freely respond to each other.

Section 3.3. AO3 and tags

AO3’s organization system sits squarely in the middle; it is a hybrid between Livejournl and fanfiction.net. It is not a completely free tagging system, but neither is it completely authority controlled. While the site does organize its homepage by media type to a certain extent, users also have the option to browse all fandoms presented to them (See Fig. 3).

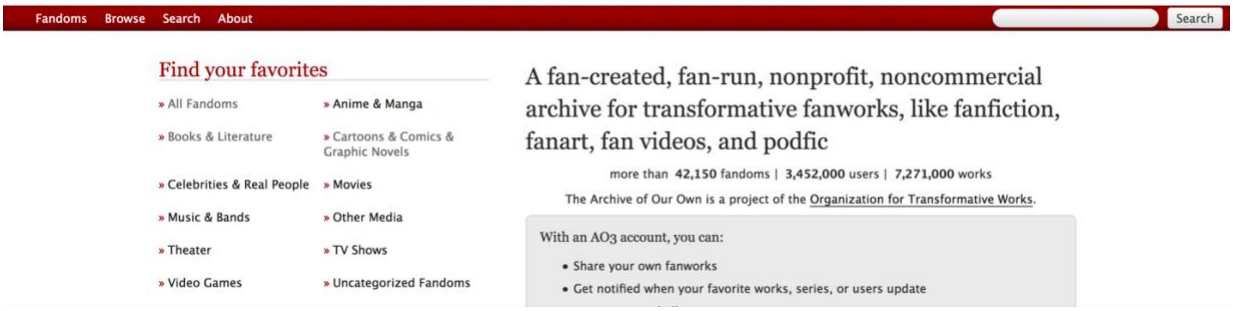


Figure 3: Archivofourown.org Homepage

From the homepage users can navigate to particular fandom’s works page, which lists all of the works associated with the fandom organized by Update Date. Figure 4 depicts the top of the *Hannibal (TV)* fandom works list, with the most recent works shown, as well as the filtering system at the right side of the screen.

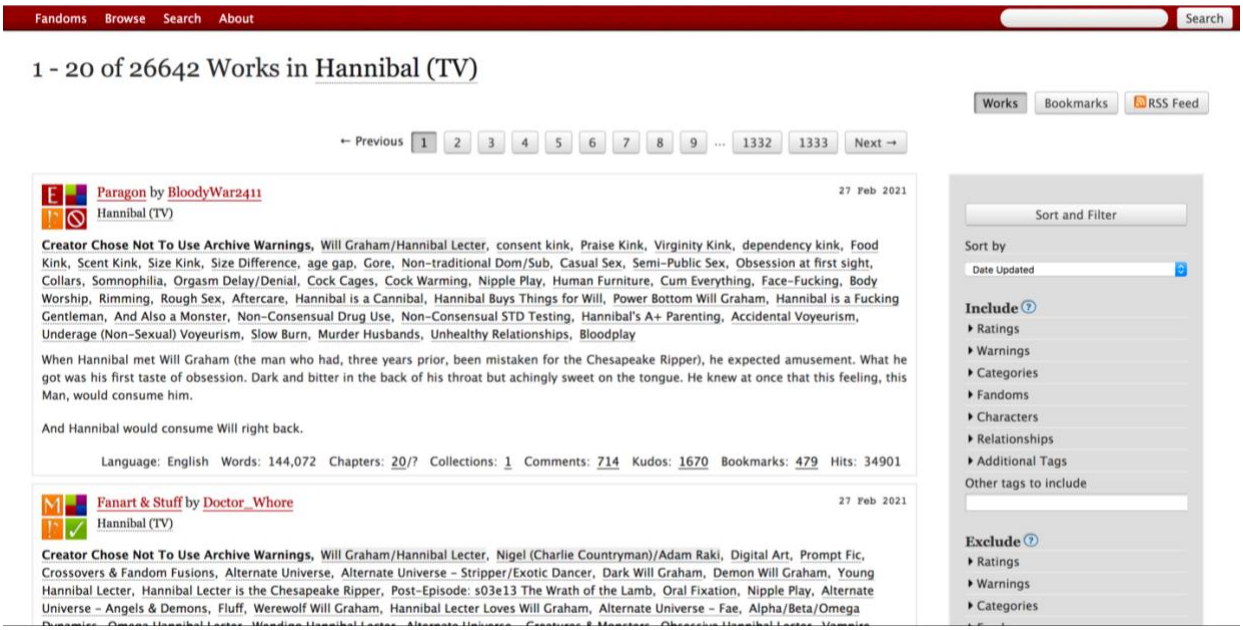


Figure 4: Hannibal (TV) work list on Archiveofourown

Works are listed with a large amount of metadata, all of which is filterable. Figure 5 shows the filtering system as it is represented on AO3. The tagging system on AO3 is split into several different categories. Aside from ratings, fandom categories and archive warnings, tags are split into three main categories: relationship tags, character tags, and additional (or 'freeform') tags. The relationship and character tags are relatively straightforward, indicating which characters are present in a work (e.g., *Will Graham*, *Hannibal Lecter*) and what relationships are included (e.g., *Will Graham/Hannibal Lecter*). The freeform tags are more complicated, and it is here that AO3's tagging system show's its versatility.

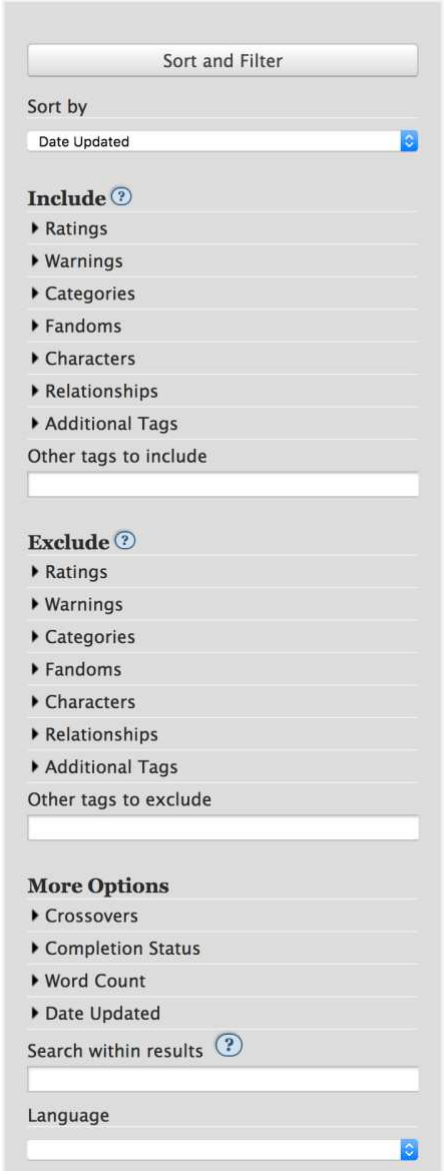


Figure 5: Filter system on AO3

Users on AO3 create their own tags, which are then reviewed by volunteers called "Tag Wranglers". These wranglers are recruited from the user-base of the site and thus are often familiar with the fandoms they wrangle. They are responsible for categorizing 'freeform' tags on the back end of the site without changing the text of tags on the public facing side of the site. This allows users to use their own language while creating a useful metadata system. For example, a user might tag their work as 'angsty af'. A tag wrangler would then categorize the 'angsty af' tag under 'angst' on the back end of the site, allowing the user to keep the form of the tag while ensuring that the work can be found under the 'angst' tag if other users are browsing the tag. This tag wrangling process allows for search functionality without impacting user's ability to freely tag their work (Tag FAQ | AO3; Fiesler et al., 2016; Johnson, 2014). AO3's hybrid system also supports the creation of more complicated tags that can't necessarily be catalogued. Tags like 'Will's killing game is strong' or 'We all know how this is going to go but that's half the fun' are completely valid and are left alone by tag wranglers. This freestyle tagging method allows authors to express themselves in a variety of ways in that they have access to controlled vocabulary terms but are not limited by them. Authors have access to metadata beyond a simple

summary of the work and these tags offer information to users browsing through fandom as to character dynamics and tropes present within the fan fiction work.

The uses of freeform tags on AO3 are extremely varied. Gursoy, Wicket, and Feinburg (2018) created a basic typology for the categorization of freeform tags on AO3, separating the tags into 'Expressive' (tags that are stylistically more informal, drawing the reader into a conversation [498]) and 'Declarative' (tags that assert the inclusion of specific elements in a fan fiction work [498]) categories. For the purposes of their study on metadata in fan work repositories these categories were enough. However, there is more complexity that can be examined in freeform tags on AO3. Like the tags on Tumblr, tags on AO3 are used for a wide range of functions. The most basic of these is findability; users label their work with tags like *Coffee Shop AU* or *Slow Burn*, categorizing their work with labels that can be used by other users to find their work.

Section 3.4. Tumblr and its effect on classification methods on AO3

Any discussion of AO3 tags would be incomplete without mentioning Tumblr. Tumblr is a blogging site that in many ways has taken the place of Livejournal in fan spaces as a blogging and social media platform. As of the end of 2018, Tumblr was one of the largest blogging sites and a popular destination for fandoms. There is a large amount of crossover between AO3 and Tumblr, many authors will link to their Tumblr blogs in their work on AO3 and fan fiction contests and prompts often start on Tumblr with the work being posted to AO3 for greater distribution purposes. Tumblr is perhaps the best comparison to AO3 in how tags are used, to the point that AO3 tags have been called 'Tumblr-style' tags (AO3 News, 2013). Tags on Tumblr serve many of the same functions as tags on AO3 and follow many of the same patterns (Attu & Terras, 2017). Tags on Tumblr are often long form, reflecting a more stream of consciousness

and conversational style that has been widely adopted and integrated into the AO3 tagging system (Hoch, 2018).

The sense of community on AO3 is diffuse. The sheer number of fandoms present in the archive is a barrier to a cohesive sense of community; however, within a fandom, the sense of community can be very strong, especially in smaller fandoms. Tags on AO3 are used to create a sense of community on a website that does not have forums, or other community building infrastructure. Aside from commenting, there is no way for users to interact with each other, which, for an archive, makes sense. However, as Jenkins (2007, 2013, etc.) argues, fan culture is participatory; community grows through the interactions between fans. The freeform nature of tags on both sites allows tags to act both as a boundary and a bridge. Tags can be 'fandom specific' (such as 'Somebody help Will Graham'), thus acting as a boundary between fandoms or 'cross fandom' (such as 'angst'), casting a wide net across multiple fandoms. This double nature is implemented by the fact that tags on Tumblr and AO3 are one of the most productive search methods on both sites. The searchability and interactive function of tagging are important uses of tagging that have been brought over from Tumblr (Hoch, 2018). The point being that tags are purposeful and meaning making, that they can reflect users' attitudes and the choices that they make about their material, both for descriptive and affective purposes.

Chapter 3: Research Methodology

Section 1. Theoretical framework and methodology theory

As of this writing there has been a significant lack of academic research on organizational paradigms in online fan spaces, specifically around user-generated tags. The subject of this thesis straddles multiple disciplines thus the theoretical framework of this thesis draws from different theories for its theoretical basis. These theories include metadata and paratext theory, fan studies, and semiotics, all of which have contributed to the theoretical background of this thesis.

Section 1.1. Semiotics and Eco's Open Work

Semiotics is broadly defined as the theory and study of signs as they are created, used, and interpreted in all of their forms (Nöth, 1990). This thesis is situated within the field of semiotics by approaching freeform tags as a form of sign interpretation as they function both for findability and expressive purposes (Gursoy et al., 2018). Umberto Eco states in *The Open Work* that an open text is a text that is interpretable by readers in many different ways (Eco, 1989). This study draws strongly on Eco's work, as the metadata that users create represents an act of interpreting the canonical text for which they are writing fan fiction. AO3 tags represent a way of expressing that interpretation through their use of freeform tags. If we define a property as an open text, tags can be seen as an extension of individual user interpretation of that text. Through the study of these tags on a fandom level, we can then draw conclusions about how individual interpretation can affect the community as a whole. Freeform tags are a representation of an author's interpretation of their own work and an interpretation of the community in which they are writing. Related to this idea, a theory put forward in Herzog (2012) theorizes that author's notes (optional notes that authors can include either at the beginning or end of their work on

AO3) are a way of interpreting (and justifying their interpretation of) the source material and that these notes can be a source of community building and connection with their audience. This analysis is not limited to author's notes; it can be applied to other types of user-generated metadata especially if we view user-generated tags on AO3 as paratext.

Section 1.2. Metadata and paratext

There is a growing body of work on metadata as a form of paratext in fan fiction spaces (Hill & Pecoskie, 2014; Leavenworth, 2014). Genette's original idea of paratext was limited to print, explaining how metatextual information (such as the title, author's notes, etc.) directed the reader to engage with the actual text in certain ways. While there have been studies that explore fan paratext more generally through author's notes and comments, there is little literature surrounding how tags act as paratext except for the idea that they do (Leavenworth, 2014). In her 2014 article "The paratext of fan fiction", Roberta Leavenworth contextualizes freeform tags on AO3 as a form of paratext, priming the reader for the type of content they can expect and signalling "how the text should be read and how it should not be read." (Leavenworth, 47; emphasis original).

Section 1.3. Fan studies

A large body of critical work has sprung up around fan fiction in the last forty years. In his 1992 book *Textual Poachers*, Henry Jenkins discussed fan works as 'remixing', hypothesizing that fan works represent how fans engage with and change the material that they are fans of, thus 'remixing' mass media into something new. This theory has been examined over the last twenty years and is still a prevailing theory with fan studies. This thesis builds on this theory, positing that the theory of 'remixing' not only applies to the content of fan works, but to the metadata that fans create for their work. The discussion surrounding fandom in a digital

environment is ongoing, including discourse on community, the nature of fan works, and the relationship of fans to copyright holders (Pearson, 2010).

It is impossible to discuss fan studies without discussing Henry Jenkins and his seminal work *Textual Poachers*. Much of the media landscape has developed into what Henry Jenkins calls "participatory culture" (Jenkins, 2013). This participatory model of culture "sees the public not simply as consumers of preconstructed messages but as people who are shaping, sharing, reframing, and remixing media content in ways that have not been previously imagined." (Jenkins 2013, 2). These ideas have influenced multiple generations of fan scholars and this thesis owes much to his work. Jenkins introduced the idea of fans as "textual poachers", media consumers who "construct their cultural and social identity through borrowing and inflecting mass culture" (Jenkins 1992, 23). Fan fiction is an example of this phenomenon, as it rewrites story canon by interacting with and reinterpreting the media that fan fiction writers consume. Chaney and Liebler (2007) define canon as "the official storylines and back stories invented by the creators of television shows, movies, and books." (Chaney & Liebler 2007, 1). Fan fiction remixes canon in the content of the work, and AO3 metadata allows users to perform remixing in the metadata itself, allowing tags to consist not only of statements about the text used for findability, but also for interpretation and interaction with other users. AO3 encourages users to add their own freeform tags that can represent their feelings on their own work much more accurately than an authority-controlled system would allow.

Section 1.4. Methodological Framework

This research applies the process of textual analysis to paratextual material. This is only possible due to the form of the tags, as the length and complexity of some of the tags lend themselves to this process. Despite my previous familiarity with fan communities, I was unsure

what my data would look like and thus structured my data collection, coding, and analysis as an iterative process. I hoped that building the research processes from the data would give new insight into the communities with which I was already so familiar. This led me to grounded theory as a theoretical approach, grounded in content analysis as a methodology due to the qualitative nature of the data.

Once the data was collected, a cursory examination of the data revealed a large number of repeated tags, as well as a significant amount of linguistic variety within the tags themselves. Based on these observations I decided to proceed with a mixed methods approach, performing both quantitative analysis and qualitative analysis on the data. Frequency analysis represented the quantitative side of the data analysis and the qualitative side consisted of content analysis. This two-tiered approach allowed me to both examine the distribution of the dataset in greater detail and focus on specific textual trends.

Section 1.5. Why these fandoms?

Why the *Hannibal Lecter* franchise? The *Hannibal Lecter* franchise fandom is a compelling object of study not only because of its popularity within the larger fan community, but because its small size belies its popularity. So much so that the fandom attracted attention from the mainstream media, especially once the television iteration of the franchise, *Hannibal*, first aired in 2013. The character of Hannibal Lecter has been an enduring presence in pop culture since he was popularized by Anthony Hopkins in the film, *The Silence of the Lambs* (1991). Created by Thomas Harris in his 1981 novel *Red Dragon*, the character was immediately iconic and is now consistently ranked highly in 'Top Villain' lists of television, film, and literature (100 Heroes & Villains | American Film Institute).

The first book in the Hannibal Lecter book series, *Red Dragon*, was released in 1981 (Harris, 1981), the second, *The Silence of the Lambs* was released in 1988 (Harris, 1988), the third book, *Hannibal* was released in 1999 (Harris, 1999), and the last book, *Hannibal Rising*, was released in 2006 (Harris, 2006). The first two books, written by Thomas Harris, predate the film and television adaptations, although the latter two books were written after *The Silence of the Lambs* film adaptation had established Hannibal Lecter in popular culture.

There have been five film adaptations of the book series. The first adaptation was the film *Manhunter* (1986), which adapted the first book in the series, *Red Dragon*. The next film adaptation was *The Silence of the Lambs* (1991) which adapted the second book in the series. A film adaptation of the third book (released in 1999), *Hannibal* was released in 2001. This was followed by a second film adaptation of *Red Dragon* (titled *Red Dragon*) that was released in 2004. The last film adaptation, *Hannibal Rising*, was released in 2007, one year after the book was released.

All of these adaptations were followed by the television series *Hannibal*, which ran from 2013-2015 on CBS. The series was created and run by Bryan Fuller. While previous film adaptations adhered quite closely to the literary source material in content, the television series detailed characters and events not previously explored, as it covered events prior to events detailed in other adaptations. This allowed the showrunners more freedom in their adaptation and Bryan Fuller has stated in interviews that "the show [for me] is very much fan fiction of these characters that I adore." (Scarano, 2014). It should be noted that the television series adapts the source material much more loosely than the previous film adaptations. While the broad strokes and characters remain largely the same, the television series expands Will Graham and Hannibal Lecter's relationship in ways that had not previously been explored.

The character has remained in the public consciousness to the point that when showrunner Bryan Fuller was creating a show around the character there was immediate interest in the project. The show premiered in 2013 and quickly gained a devoted following, both in the press and the wider public. It almost immediately gained a following on Tumblr, which garnered so much attention that it made news (Hall, 2013). The fan activity was so intense that Bryan Fuller credits the "Fannibals", as fans call themselves, with helping to save the show from cancellation after the first season's low ratings (Bryan Fuller, 2013). As discussed in Chapter 2, there is a symbiotic relationship between Tumblr and AO3 and the high level of fan engagement on Tumblr quickly turned into an active community of fan writers on AO3. It is this level of fan engagement, paired with the relatively small size of the fandom, that makes it a research object given the scope of this study.

Given the recency of the television series, it should come as no surprise that the most active segment of the Hannibal Lecter fandom surrounds this iteration of the fandom. This brings us to another relevant point, the multiple iterations of the character in books and film adaptations. The fan works concerning the character of Hannibal Lecter on AO3 are split into three main groups; the works listed under *Hannibal Lecter (Thomas Harris Tetralogy)*, the works listed under *Hannibal Lecter (Hopkins Movies)*, and the works listed under *Hannibal (TV)*. So, while all of these fan works can be considered to be about the same characters, they are split into categories based on their medium. This expands our analysis and can provide some insight as to how fans might create tags based on the media type of their fandom.

It should be noted that while I am no longer an active member of this particular fandom, I was an active participant in this fandom from 2014-2017. This has affected my approach in the

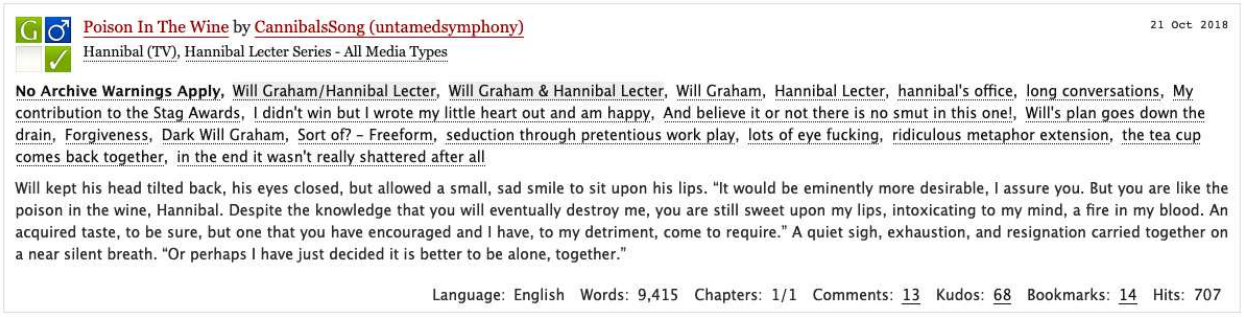
following ways. Firstly, my personal interest in this fandom preceded my academic interest. I would not have chosen to study this topic if I was not already embedded in the culture. Secondly, my familiarity with the community allowed me to see that there was something interesting occurring within the tags themselves. While this meant that I was primed to examine the more complex tags, it also meant that I had to be careful of my own bias.

Section 2. The Data

The methods used in this research project involve the collection and analysis of user-generated metadata from a popular fan fiction website, the Archive of Our Own (AO3). The data was collected on January 31, 2019 and consists of the user-generated tags from three fandoms on the website: *Hannibal (TV)*, *Hannibal Lecter (Thomas Harris Tetralogy)*, and *Hannibal Lecter (Hopkins Movies)*. These fandoms were chosen for comparison purposes, as they are all iterations of a single franchise. The three fandoms are of varying sizes, the *Hannibal (TV)* is the largest, with 157,788 tags before analysis was performed. The other two are smaller by comparison, with *Hannibal Lecter (Thomas Harris Tetralogy)* containing 3,627 tags before analysis and *Hannibal Lecter (Hopkins Movies)* containing 1,553 tags before analysis.

Section 2.1. Collection Process

The data was collected using three custom web scrapers that were created for this project. The data was collected on January 31, 2019. Since *Hannibal (TV)*, *Hannibal Lecter (Thomas Harris Tetralogy)*, and *Hannibal Lecter (Hopkins Movies)* are interconnected by virtue of being iterations of the same franchise, there was crossover in the three work lists. This meant that works tagged with all three fandoms were included in all three data sets. This was accounted for in the cleaning of the data after the data was collected.


Poison In The Wine by CannibalsSong (untamedsymphony) 21 Oct 2018
Hannibal (TV), Hannibal Lecter Series - All Media Types

No Archive Warnings Apply, Will Graham/Hannibal Lecter, Will Graham & Hannibal Lecter, Will Graham, Hannibal Lecter, hannibal's office, long conversations, My contribution to the Stag Awards, I didn't win but I wrote my little heart out and am happy, And believe it or not there is no smut in this one!, Will's plan goes down the drain, Forgiveness, Dark Will Graham, Sort of? - Freeform, seduction through pretentious work play, lots of eye fucking, ridiculous metaphor extension, the tea cup comes back together, in the end it wasn't really shattered after all

Will kept his head tilted back, his eyes closed, but allowed a small, sad smile to sit upon his lips. "It would be eminently more desirable, I assure you. But you are like the poison in the wine, Hannibal. Despite the knowledge that you will eventually destroy me, you are still sweet upon my lips, intoxicating to my mind, a fire in my blood. An acquired taste, to be sure, but one that you have encouraged and I have, to my detriment, come to require." A quiet sigh, exhaustion, and resignation carried together on a near silent breath. "Or perhaps I have just decided it is better to be alone, together."

Language: English Words: 9,415 Chapters: 1/1 Comments: 13 Kudos: 68 Bookmarks: 14 Hits: 707

Figure 6: AO3 work listing as it appears on the site

Fandom Tags:

All of the following tags are catalogued under the three fandoms by tag wranglers on the backend of AO3. While this fact is secondary to this research, it is interesting to note the variety of fandom tags that users assign to their work, especially in the case of *Hannibal (TV)*. This shows that there is a large amount of tag variety even in general fandom organization.

Fandom tags included under the *Hannibal Lecter (Hopkins Movies)* fandom tag	**Fandom tags included under the *Hannibal Lecter (Thomas Harris Tetralogy)* fandom tag**	**Fandom tags included under the *Hannibal (TV)* fandom tag**
Hannibal (2001), *Hannibal (NBC),* *Hannibal Series (Movies),* *Red Dragon (2002),* *Red Dragon - Fandom,* *silence of the lamb,* *Silence of the Lambs (1991),* *Silence of the Lambs (Movies),* *Silence of the Lambs - Fandom,* *The Silence of the Lambs (1991),*	*alludes to red dragon??,* *Hannibal - Harris,* *Hannibal - Thomas Harris,* *Hannibal Lecter Tetralogy,* *Hannibal Rising - Thomas Harris,* *Red Dragon - Thomas Harris,* *Silence of the Lambs - Thomas Harris,* *The silence of the lambs - Thomas Harris,*	*abbyrose: Hannibal,* *Alpha Hannibal - Fandom,* *Bedannibal - Fandom,* *Bedelia - Fandom,* *Bedelia and Hannibal,* *Bedelia Du Murier,* *Bedelia Du Murier/Hannibal,* *Lector - Fandom,* *brian zeller - Fandom,* *Brownham - Fandom,* *fandom: Hannibal,* *fandom:Hannibal,* *Fannibal,* *freddie lounds - Fandom,* *haniball XDDDD,* *Hannbial,*

		hanni - Fandom, *Hannibal (NBC),* *Hannibal (2013),* *Hannibal (TV Series),* *Hannibal - Fandom,* *Hannibal - TV series,* *Hannibal 2013,* *Hannibal and Bedelia - Fandom,* *Hannibal Lecter/ Will Graham - Fandom,* *hannibal lecter/bedelia du maurier - Fandom,* *Hannibal Lector - Fandom,* *Hannibal NBC,* *hannibal reclist,* *Hannibal TV,* *hannibal tv series,* *Hannibal's past - Fandom,* *Hannibal(TV),* *Hannibal(TV) - Fandom,* *Hannibal_TV,* *Hannibals,* *Hannibql,* *Hannigail - Fandom,* *Hannigraham - Fandom,* *HanniGrahm - Fandom,* *Hannigram - AU fandom,* *hannigram - Fandom,* *Hannigram AU - Fandom,* *Hannigram AU- Fandom,* *inspired by conversations between Hannibal and Will*, *lamb: hannibal*, *murder boyfriends - Fandom*, *Murder Family - Fandom*, *Murder Husbands - Fandom*, *NBC Hannibal*, *NBC's Hannibal*, *Nbchannibal - Fandom*, *Omega Will - Fandom*, *Relevés - Fandom*, *season 3 meeting in gallery-plot beginning*, *stag - Fandom*,

		team sassy science - Fandom, *Twittibal*, *We got some Hannibal, will - Fandom*, *Will gRA*, *will graham - Fandom*, *Will's past - Fandom*, *拔杯 – Fandom*, *汉尼拔 – Fandom*

Table 1: Fandoms catalogued under the three fandoms studied on AO3

Section 2.2. Building the Scrapers

In my previous research on tags, I used web scrapers to gather tag data on AO3. Using the *Thor (Movies)* fandom as a case study, the pilot study of this thesis discussed tag frequency within the *Thor (Movies)* fandom to explore how fandom members think about characters and relationships within their fandom. The data for the pilot study was gathered using a third party webscraper and as such a third party webscraper was the original choice for this project. However, most of them required monthly subscriptions that made them prohibitively expensive.

Dataminer.io (https://data-miner.io), a browser-based webscraper, was the main third-party scraper that I considered for this project. This scraper was used in the pilot study that inspired this thesis and as such it was my first choice for this research. Dataminer.io is useful for small projects, however it quickly becomes expensive for larger projects. The scraper limits the number of pages that the user can scrape for free to 500 pages per month. Unfortunately, as this research involves scraping over 1,000 pages of work lists, Dataminer quickly became far too expensive. I also needed to collect data from all three fandoms at the same time, so collecting half of the data in one month and half in the next was not feasible.

The choice then became a decision between two code libraries used for building data scrapers in Python, Beautiful Soup and Scrapy. Beautiful Soup is a comprehensive library for scraping data from HTML and XML files, but it requires a high level of comfort in the Python

programming language. This makes it highly customizable, but it requires a large amount of testing and learning time that, when compared to Scrapy, made it the less feasible choice.

This left Scrapy (https://scrapy.org). There were several factors involved in this decision. The first is that the documentation for the library is excellent. Basic tutorials are readily available, and the software documentation includes full articles on the library's more advanced functionality. The second factor is the active community of users willing to answer questions. The third is the tools provided by the library itself, the most important of which is the scrapy shell, an interface in which users can test their code before running a full scraper. This saved time and allowed me to test my code before implementing it on over 1,000 pages of data. The fourth factor is the built-in structure of the library. Creating a project in Scrapy creates a series of files that structures, cleans, and parses the data within the scraper itself. The result is a clean and structured output that can be easily exported in a variety of formats, which, for the purposes of this research, was the .csv format.

Three scrapy projects were created for this study: hannibalmovies, hannibalbooks, and hannibaltv. The three scrapers were limited by fandom; hannibalmovies scraped all the works in the *Hannibal Lecter (Hopkins Movies*) fandom, hannibalbooks scraped all the works in the *Hannibal Lecter (Thomas Harris Tetralogy*) fandom, and the hannibaltv scraper scraped all the works in the *Hannibal (TV series)* fandom. All three of the webscrapers scraped specific metadata for all of the works in all three fandoms up to January 31, 2019. The metadata categories scraped were 'author', 'title', 'date', 'fandoms', 'freeform tags', 'bookmarks', 'kudos', and 'hits'. Only the 'fandoms' category and the 'freeform tags' category were used for this study.

Section 2.3. Data Cleaning

The data was cleaned in OpenRefine and Excel. The data was initially cleaned in OpenRefine, a free and open-source data cleaning and organizing tool. I chose OpenRefine due to my familiarity with its ability to take .csv input files and its versatile data manipulation tools. In the raw data, all of the freeform tags associated with a work were contained in a single spreadsheet cell, which prevented individual tag analysis. OpenRefine allowed me to separate the tags into different cells while still associating the tags with the correct work.

Once all of the tags had been moved into one column and the rest of the raw data had been removed, the data was moved into Excel. Excel was used to remove capital letters, spaces between words in tags, punctuation, and non-letter characters using the find and replace tool. This was done to ensure that the analysis tools would read all of the tags correctly instead of separating tags on a punctuation mark. Excel was also used for removing non-English tags due to readability issues. This was done in the collected data instead of using AO3's search filters because for the most part tags on AO3 are limited to English. While there are some outliers, and AO3 does support tags in other languages, most users stick to English.

Section 2.4. Data Visualization

The data was visualized in Excel using the charting tools. The visualizations are simple, limited to tables, bar charts, and line charts. These tables and figures are used to compare the frequency values generated by Voyant, the mean, median, and mode of LIWC summary variable scores, and the averages of the other linguistic variables examined in this research.

Section 3. Data Analysis Tools and Methods

The first facet of the data analysis was performed with Voyant (https://voyant-tools.org/). Voyant is a text analysis tool developed by Stéfan Sinclair and Geoffrey Rockwell (Sinclair & Rockwell, 2019). The tool takes data in the word of the textfile, spreadsheet, or webpage submitted and performs various types of textual analysis. For the purposes of this research, the most important analysis tool in Voyant was the frequency counter. The tool counts the frequency of words (hence the removal of spaces within the tags in the data) and outputs the data in csv file format. Once the data was run through this tool it was downloaded, analysed, and visualized. Tag frequency was created by uploading a spreadsheet in which the spaces between words in tags were taken out (e.g., the tag 'alternate universe' became 'alternateuniverse'). This was done because, while Voyant can recognize repeating phrases, a significant number of tags consist of unique word combinations. It also does not recognize tag boundaries unless each tag is considered as one word.

The second facet of the data analysis was performed with Linguistic Inquiry and Word Count (LIWC). Linguistic Inquiry and Word Count is text analysis software created for use in sociological research. Created by Dr. James Pennebaker, the tool compares textual data to an internal dictionary annotated with various linguistic terms and psychological concepts (Boyd & Pennebaker, 2015). The inputted textual data is compared to this dictionary and analysed to determine the linguistic terms and psychological constructs present in any given dataset. This comparison generates output data in the form of scores (0.00 being the lowest score and 100.00 being the highest) in each piece of data, in this case each tag. The tool was initially released in 1993 and since then it has gone through three iterations, each iteration updating the default dictionary used by the tool. The latest dictionary, LIWC2015, accommodates text usage in the digital sphere. The dictionary now recognizes some emoticons and common "netspeak" language

found on Twitter and Facebook (Pennebaker et al., 2015). This inclusion of internet language usage makes LIWC2015 an ideal choice for this research, as it has already been optimized for the type of data analysed in this project.

The tool takes input in text files or spreadsheets and compares the inputted data to a dictionary coded with a variety of linguistic and sociological categories. The data is then outputted in the form of a spreadsheet that lists the incidence of the various categories in each tag and also lists the word count of each tag. The categories are listed in the *LIWC Language Manual*. They are separated into three categories that are relevant to this research: Summary Language Variables, Linguistic Dimensions, and Psychological Processes. Summary Language Variables consist of four variables; Analytical thinking (reflecting the level of formal, logical, and hierarchical thinking in a text), Clout (reflecting the level of expertise and confidence of the author, Authentic (reflecting the level of authenticity in the text), and Emotional tone (reflecting the emotional tone of the text, i.e., whether the text is positive or negative). Linguistic Dimensions encompasses the parts of speech category of this tool. It includes categories such as pronouns, articles, conjunctions, verbs, etc. The parts of speech category that this research is most interested in is function words, that is, words that do not contain much linguistic content (Malmkjaer, 2004). Examples of function words include pronouns (I, you, he/she), articles (a, an), and prepositions (with, above). Despite their lack of linguistic content, these function words can tell us how people think (Pennebaker et al., 2014; Chung & Pennebaker, 2007; Kacewicz et al., 2014; Newman et al., 2003) and when examined in a text they can provide insight (Burrows, 1987). It should also be noted that, according to the Corpus of Contemporary American English (COCA), function words are among the most frequently used words in American English (COCA, 2020).

The last category is Psychological Processes, which categorizes words by various processes that are coded into language. For the purposes of this research, I am most interested in the informal language category. This category includes swear words, netspeak, and fillers.

Chapter 4: Data Analysis and Findings

Section 1. Introduction

Using the methods laid out in the previous chapter, I examined the data from all three fandoms in an attempt to answer my three research questions; what linguistic features exist in user-generated tags across fandoms on AO3, what similarities and differences exist in user-generated tags across fandoms (television, books, and movies) on AO3, and what can we infer about discourse in the fan fiction communities from these user-generated tags? This chapter will lay out the data analysis and findings.

As previously stated, the data was run through two data analysis tools: Voyant and Linguistic Inquiry and Word Count (LIWC). The collection and cleaning process were covered in the last chapter. Once the tags were run through the Voyant analysis, they were taken and run through the Linguistic Inquiry and Word Count software to gain further understanding of their linguistic composition. There was a significant difference in size between the three fandoms. The *Hannibal (TV)* fandom was the largest dataset, followed by the *Hannibal Lecter (Thomas Harris Tetralogy)* fandom, and the smallest dataset was the *Hannibal Lecter (Hopkins Movies)* dataset. Table 2 lays out the total number of tags in each fandom before any analysis was done. This data includes duplicate tags and non-English tags, both of which were removed during the cleaning process.

Fandom	Total # of Tags (before frequency counts and cleaning)
Hannibal (TV)	157788
Hannibal Lecter (Thomas Harris Tetralogy)	3627
Hannibal Lecter (Hopkins Movies)	1553

Table 2: Total number of tags in each fandom

Section 2. Voyant Analysis and Tag Frequency

The three datasets were run through Voyant for frequency analysis. Table 3 lays out the results of this process, listing the number of tags in each data once the individual tokens (in this case, the duplicate tags) from each dataset had been removed, leaving only the types (a single instance of each different tag). As we can see in Table 2, despite this process the size ordering of the fandoms remains the same.

Fandom	# of Tags
Hannibal (TV)	44324
Hannibal Lecter (Thomas Harris Tetralogy)	2073
Hannibal Lecter (Hopkins Movies)	994

Table 3: Number of tags in each dataset once duplicates were removed

Once all of the datasets had been run through Voyant and downloaded into spreadsheets for ease of analysis, it became clear that the majority of the tags in each dataset were instances of *hapex legomenon* (hapaxes); that is, tags occurring only once in each dataset (OED Online). Despite the obvious size differences between the three fandoms, all three datasets were quite similar in content.

Fandom	# of Tags	# of Unique Tags	**Hapaxes in Dataset**
Hannibal (TV)	44324	35753	**80.6%**
Hannibal Lecter (Thomas Harris Tetralogy)	2073	1593	**76.8%**
Hannibal Lecter (Hopkins Movies)	994	788	**79.3%**

Table 4: Number of hapaxes compared to the total number of tags

Section 2.1. Overview of Tag Frequency

Some tags are used more frequently than others in user-generated cataloguing systems (Marlow et al., 2006). People often think in similar ways and this is heightened in communities devoted to the celebration of a specific piece of media. This is most definitely true on AO3, where the entire site is built around tags. This has given rise to a robust and varied tagging culture on the site, with some tags gaining popularity across fandom boundaries. These 'cross-fandom' tags are usually simple, descriptive nouns or noun phrases like 'fluff', 'angst', 'hurt/comfort', 'alternate universe', etc. On the other hand, some tags stay within a fandom or only spread across a few interconnected fandoms. These 'fandom-specific' tags are usually more specific, including character names, community shorthand, or references to specific moments in the media's narrative. For example, 'murder husbands' is a tag specific to the Hannibal Lecter fandoms explored in this thesis. As we will see in the following sections, while the most frequent tags in the three fandoms examined in this thesis are mostly cross-fandom tags, there are some fandom-specific tags that are quite popular. Across all three fandoms we see a significant difference between the frequency of the most frequent tag and the frequency of the next most frequent tag. This is predicted by Zipf's Law, which states that the frequency of a word is inversely proportional to its frequency rank (Zipf, 1935). Roughly 75% and 80% of the tags in each dataset are hapaxes. This indicates that there is a high amount of tag diversity in each fandom and that users tend towards creating their own tags over reusing established tags, although this bears more investigation and needs to be cross referenced with frequency counts. This led to a closer examination of the unique tag datasets to see if any patterns existed within that subsection of the data.

Section 2.2. Tag Frequency in Hannibal (TV) fandom

As we saw in Section 1, *Hannibal (TV)* is by far the largest fandom of the three. This means that the raw frequency numbers are much higher than the other two fandoms and it contains a larger number of tags that could be considered frequent. Despite this, there is a severe drop-off in frequency between the most frequent tag and the rest of the data (Figure #7). The distance between the most frequent tag and the next most frequent tag is the largest distance between frequency measurements in the entire dataset, as the Raw Frequency measurement drops by more than 700. While this is a preliminary indication that the trajectory of the data follows Zipf's law, the comparison is not exact, as the difference between data points becomes much smaller after the first data point. As we can see in Figure #7, the differences between frequency measurements are much smaller in the rest of the top 25 tags.

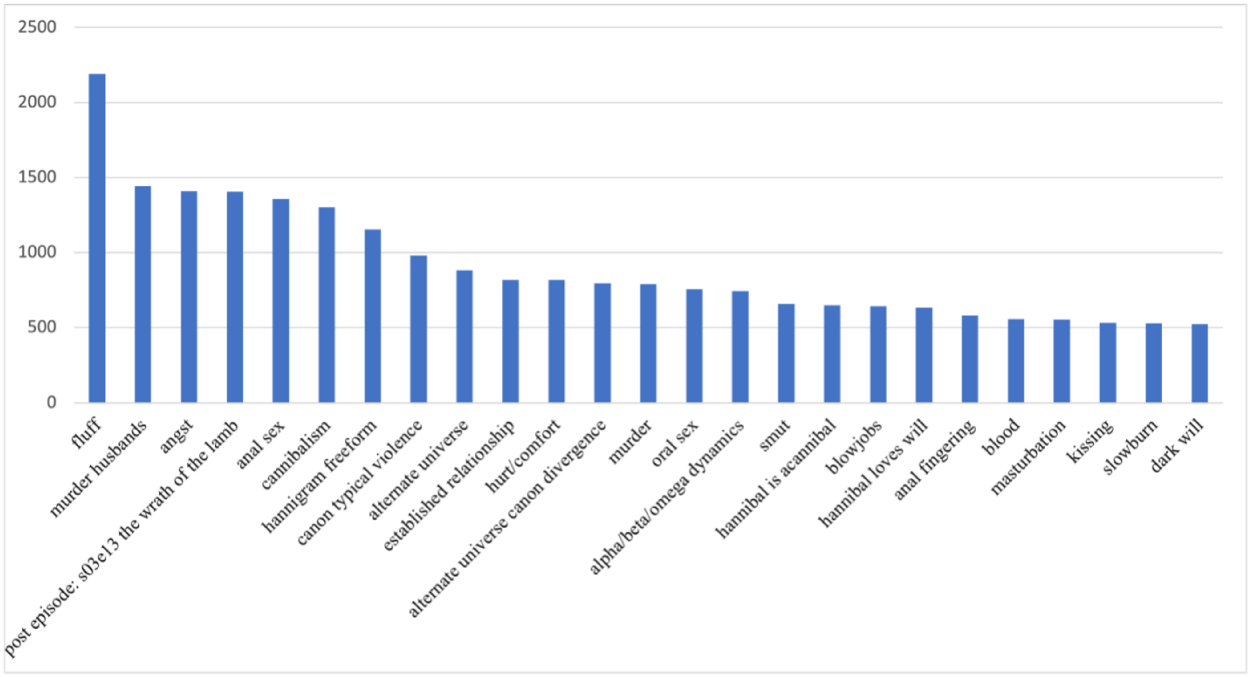


Figure 7: Hannibal (TV) Top 25 most frequent tags

Concerning the content of the tags themselves, we can examine Table 5 for more details. The first observation that we can make is that the most frequent tag is not fandom-specific. The most frequent tag that appears in the dataset, 'fluff', is a popular tag across AO3, over 800,000 works on AO3 use 'fluff' as a tag across a number of fandoms. Tags like 'angst', 'canon typical violence', and 'alternate universe' are all popular cross-fandom tags. Of the top twenty-five tags displayed in Table 4, six of the tags could be considered to be fandom-specific; 'murder husbands', 'post episode: s03ep13 the wrath of the lamb', 'hannigram freeform', 'hannibal is a cannibal', 'hannibal loves will', and 'dark will'. All of these tags reference the show in some way, referring to specific episodes ('post episode: s03ep13 the wrath of the lamb'), character relationship references ('hannigram freeform', 'murder husbands', 'hannibal loves will'), or character references ('dark will', hannibal is a cannibal'). It is interesting to note that the 'murder husbands' tag references a term that originated in the fandom rather than in the show itself (although the show referenced the term in season 3). Thus, we see that it is not only the canon material that users draw from to create tags, users on AO3 also draw from other fans.

Term	**Raw Frequency**
fluff	2191
murder husbands	1442
angst	1409
post episode: s03e13 the wrath of the lamb	1407
anal sex	1356
cannibalism	1301
hannigram freeform	1154
canon typical violence	981
alternate universe	882

established relationship	819
hurt/comfort	817
alternate universe canon divergence	797
murder	790
oral sex	757
alpha/beta/omega dynamics	744
smut	659
hannibal is a cannibal	649
blowjobs	642
hannibal loves will	635
anal fingering	583
blood	559
masturbation	555
kissing	534
slow burn	529
dark will	524

Table 5: Top 25 most frequent tags in Hannibal (TV)

Section 2.3. Tag Frequency in Hannibal Lecter (Thomas Harris Tetralogy) fandom

The next largest dataset is *Hannibal Lecter (Thomas Harris Tetralogy)*. Consisting of 1593 tags, it is smaller than the *Hannibal (TV)* dataset. Despite this size difference, the two datasets are similar in tag content. The raw frequency measurements in *Hannibal Lecter (Thomas Harris Tetralogy)* show smaller differences between the most frequent tag and the next most frequent much closer together, forming a much more consistent decline in Raw Frequency measurements than is seen in *Hannibal (TV)*.

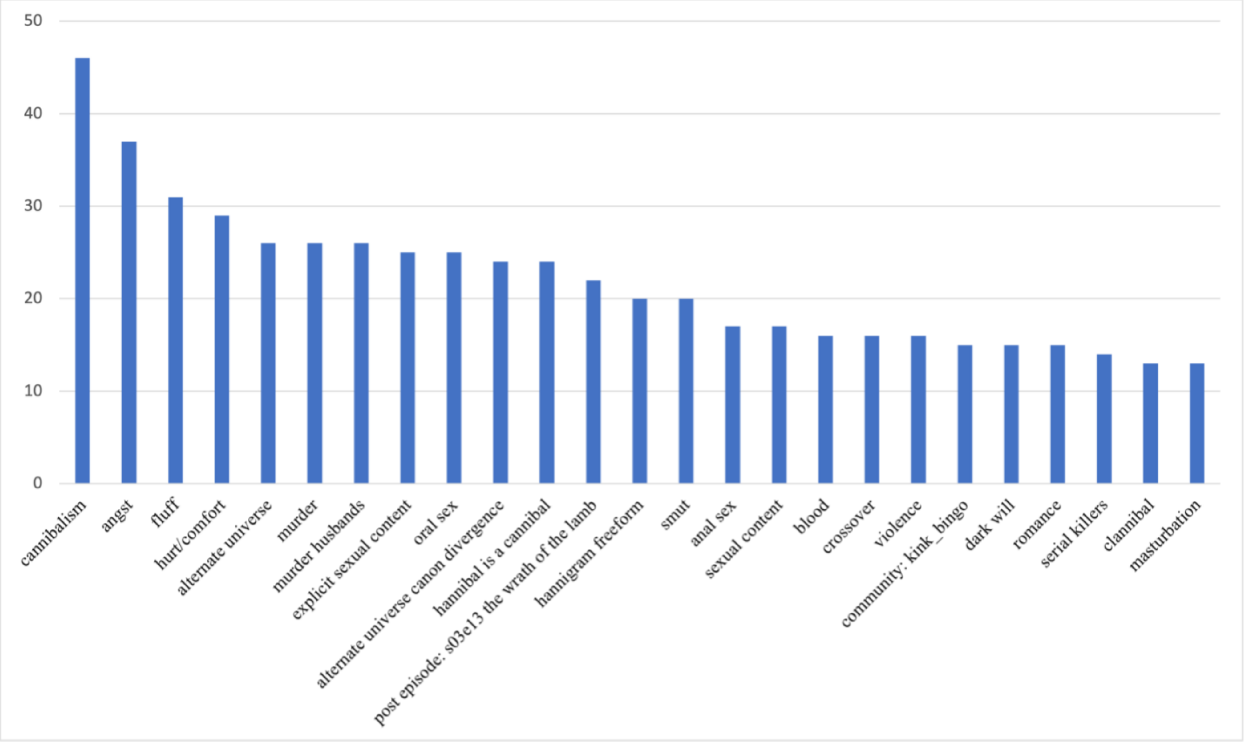


Figure 8: Hannibal Lecter (Thomas Harris Tetralogy) Top 25 most frequent tags

While the content of the dataset is similar to *Hannibal (TV)*, the top twenty-five most frequent tags are not the same. The most obvious difference is that the most frequent tag in the dataset is ‘cannibalism’, not ‘fluff’, although ‘fluff’ is still in the top three most frequent tags. This is an interesting change, as ‘cannibalism’ sits on the edge of being a fandom-specific tag. The tag is most often used within fandoms related to the Hannibal Lecter franchise, although it is used with other fandoms to a much lesser degree. Other than ‘cannibalism’, the fandom-specific tags that are most frequent in this dataset are ‘murder husbands’, ‘hannibal is a cannibal’, ‘post episode: s03ep13 the wrath of the lamb’, ‘hannigram freeform’, ‘dark will’, and ‘clannibal’. Thse tags reflect many of the same priorities seen in *Hannibal (TV)*, focusing on the same concepts and characters. The only tag seen here that is not present in the *Hannibal (TV)* dataset is ‘clannibal’, which is not surprising, as it is a portmanteau of ‘hannibal’ and ‘clarice’, a character

not present in the television show. As we see in Table 7 and above in Section 2.1., there is a high degree of overlap in the fandom-specific tags in these datasets.

Term	**Raw Frequency**
cannibalism	46
angst	37
fluff	31
hurt/comfort	29
alternate universe	26
murder	26
murder husbands	26
explicit sexual content	25
oral sex	25
alternate universe canon divergence	24
hannibal is a cannibal	24
post episode: s03e13 the wrath of the lamb	22
hannigram freeform	20
smut	20
anal sex	17
sexual content	17
blood	16
crossover	16
violence	16
community: kink_bingo	15
dark will	15
romance	15
serial killers	14

clannibal	13
masturbation	13

Table 6: Top 25 most frequent tags in Hannibal Lecter (Thomas Harris Tetralogy)

Section 2.4. Tag Frequency in *Hannibal Lecter (Hopkins Movies)* fandom

This dataset contains 1553 tags, making it the smallest dataset of the three. The content of this dataset is similar to the other two datasets, for example many of the same tags that we saw as being most frequently used in the other two datasets are also present in Table 7.

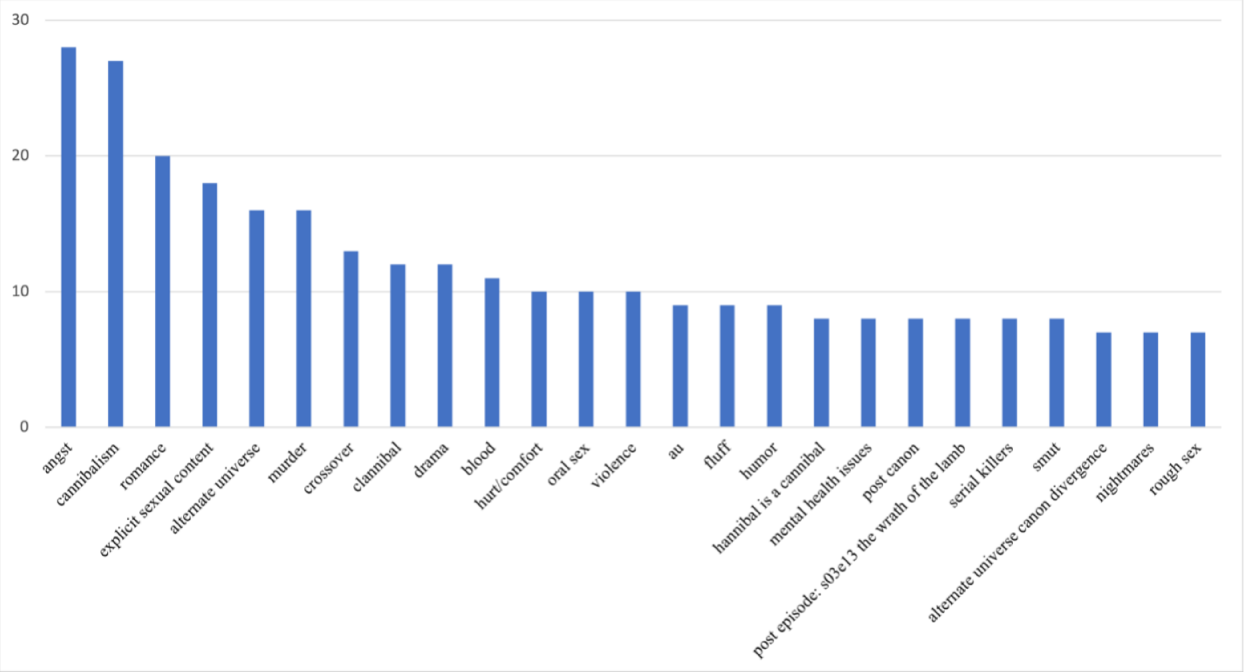


Figure 9: Hannibal Lecter (Hopkins Movies) Top 25 most frequent tags

The data contains a large number of hapaxes, comprising 79.3% of the dataset. This is in keeping with the other datasets. As we can see in Fig #9, the difference between the most frequent tag and the second most frequent tag is negligible compared to the previous two datasets. While in the other two datasets there is a noticeable gap between the most frequent tag and the next most frequent tag, in this dataset there is very little difference between the two most

frequent tags. The most frequent tag is ‘angst’, and the second most frequent tag is ‘cannibalism’, similar to the *Hannibal Lecter (Thomas Harris Tetralogy)* dataset. In contrast to the other datasets, the idea of Hannibal Lecter as a character isn’t even in the tag list until the 9th most frequent tag (‘clannibal’). This indicates a higher interest in narrative and character relationships over character names or descriptions.

Hannibal Lecter (Hopkins Movies)	**Raw Frequency**
angst	28
cannibalism	27
romance	20
explicit sexual content	18
alternate universe	16
murder	16
crossover	13
clannibal	12
drama	12
blood	11
hurt/comfort	10
oral sex	10
violence	10
au	9
fluff	9
humor	9
hannibal is a cannibal	8
mental health issues	8
post canon	8

post episode: s03e13 the wrath of the lamb	8
serial killers	8
smut	8
alternate universe canon divergence	7
nightmares	7
rough sex	7

Table 7: Top 25 most frequent tags in Hannibal Lecter (Hopkins Movies)

Section 2.5. Comparing the most frequent tags across fandom boundaries

As we can see in Figure #10 below, all three of the datasets contain similar content.

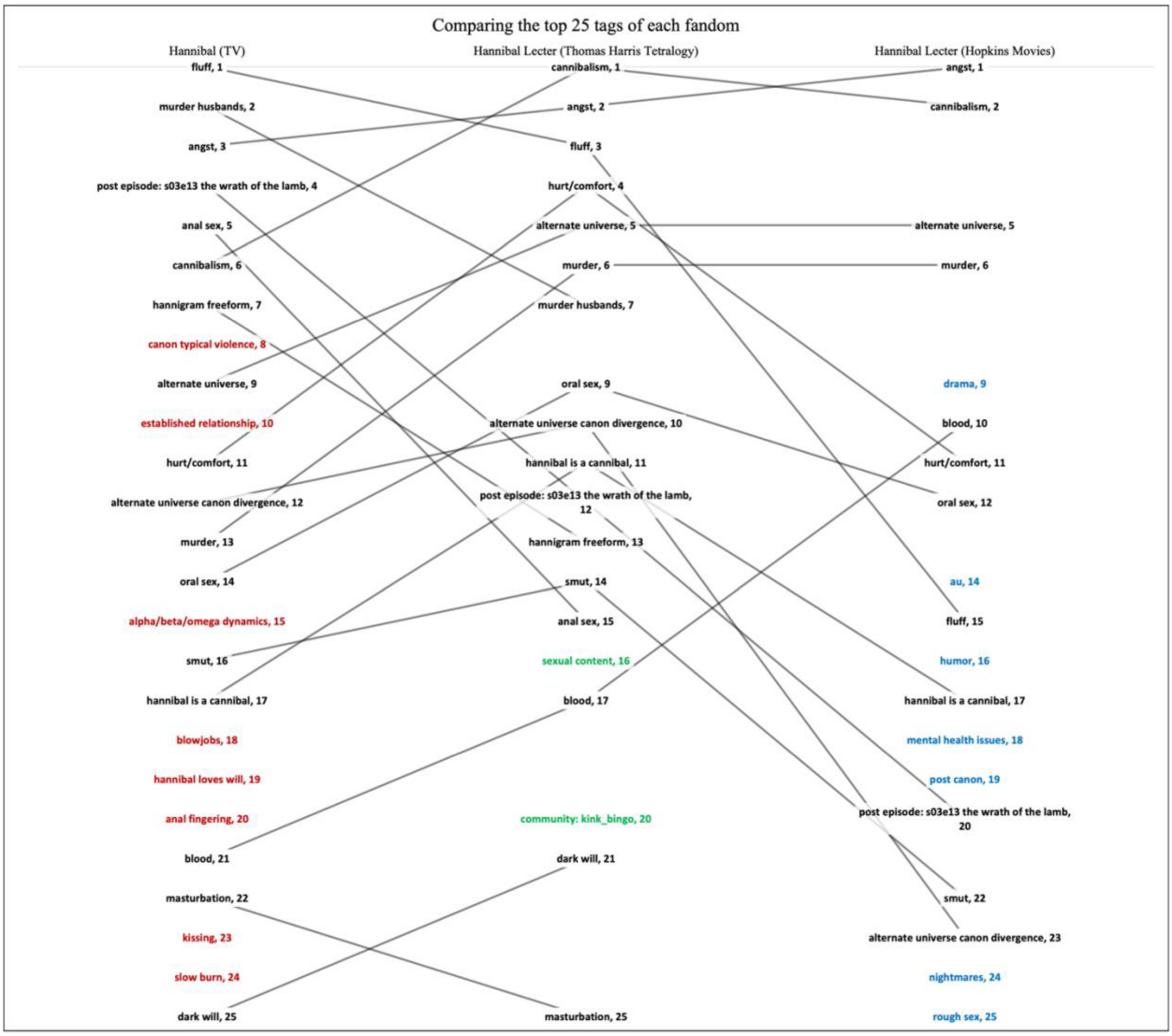


Figure 10: Comparing the top 25 tags of each fandom

Overall, the most frequent tags in all three fandoms reflect similar user tagging priorities, the top five tags in each fandom being various iterations of tags like ‘cannibalism’, ‘murder husbands’, ‘angst’, ‘fluff’, and ‘alternate universe’. Fittingly for media about a cannibalistic serial killer, the words ‘murder’ and ‘cannibalism’ occur in the top ten tags in all three fandoms, showing that despite the transformative nature of fan fiction, major themes from the original text

are still present. Another interesting content note is the presence of ‘post episode: s03e13 the wrath of the lamb’ in all three datasets. This tag refers to the last episode of the television series, it has no connection to the books or the movies. The presence of this tag shows that there is a large amount of overlap between the three fandoms and that the television series has a large influence over the other two fandoms.

The various references to ‘alternate universe’ in each fandom are another interesting note. Figure #10 shows that each fandom contains at least one reference to ‘alternate universe’, be it ‘alternate universe’, ‘alternate universe canon divergence’ or ‘au’. ‘Alternate universe’ as it is used in this context is a descriptive statement that refers to any changes that a fan fiction author might make to the source material, allowing them to explore placing characters in different situations. For example, a fan author might use ‘alternate universe cyberpunk’ to explore the characters or plot of the original media in a cyberpunk setting. The fact that ‘alternate universe’ is used so frequently as a tag implies that reimagining the original media is a popular activity in this community. As we see from the content of the tag (i.e., simply ‘alternate universe’ rather than ‘alternate universe genderswap’ for example), the tag itself is most likely used as a placeholder for whatever changes the author makes rather than a specific setting or overarching change. Thus, this tag would most likely be used for descriptive purposes rather than for findability purposes.

There is a large amount of overlap in the tag content between the three datasets, so much so that it is more interesting to examine where it is that the three datasets do not overlap rather than where they do. In Figure #10 we see that each of the three datasets has tags specific to that dataset (colour coded in the figure). The most obvious difference between the datasets is that while the *Hannibal (TV)* and *Hannibal Lecter (Hopkins Movies)* datasets each have multiple

specific tags, *Hannibal Lecter (Thomas Harris Tetralogy)* only has two. There are also several tags that only occur in *Hannibal Lecter (Thomas Harris Tetralogy)* and *Hannibal Lecter (Hopkins Movies)* and several tags that only occur in *Hannibal (TV)* and *Hannibal Lecter (Thomas Harris Tetralogy)*. There are no tags that occur in only the *Hannibal (TV)* and *Hannibal Lecter (Hopkins Movies)*. This is interesting, as we might expect the two visual media iterations of the franchise (television show and film) to have more in common. Rather, Figure #10 shows us that it is the book series *Hannibal Lecter (Thomas Harris Tetralogy)* that acts as a bridge between the three fandoms. Moving on to the content of the tags themselves, it is interesting to note that *Hannibal Lecter (Thomas Harris Tetralogy)* and *Hannibal Lecter (Hopkins Movies)* have the same two most frequent tags, which the *Hannibal (TV Series)* dataset does not share. This implies that the two datasets have more in common with each other than with the *Hannibal (TV Series)* fandom.

This similarity is also seen in the word count of tags in each fandom. The most common tag word count in the entire *Hannibal (TV)* dataset is two words per tag, while in both *Hannibal Lecter (Thomas Harris Tetralogy)* and *Hannibal Lecter (Hopkins Movies)* datasets the most common tag word count is one. This matches the word count distribution of the top twenty-five most frequent tag data. Even without an in-depth analysis of the word count of each tag, we can see just from a glance that most of the most frequent tags are quite short.

Fandom	**Most common word count**
Hannibal (TV Series)	2 word tag
Hannibal Lecter (Thomas Harris Tetralogy)	1word tag
Hannibal Lecter (Hopkins Movies)	1word tag

Table 8: Most common tag word count in top 25 tags

The one exception to this trend is ‘post episode: s03e13 the wrath of the lamb’. This is an interesting exception, as it appears in all three of the datasets, even in *Hannibal Lecter (Thomas Harris Tetralogy)* and *Hannibal Lecter (Hopkins Movies)* where it technically does not apply. ‘Post episode: s03e13 the wrath of the lamb’ is a descriptive tag that refers to the last episode of the television series. The use of this quite specific tag in all three fandoms is the clearest example of crossover between the fandoms, as the tag only has relevance to the television series yet is used in *Hannibal Lecter (Thomas Harris Tetralogy)* and *Hannibal Lecter (Hopkins Movies)*, fandoms dedicated to the book series and the film series respectively.

Section 2.7. The rest of the frequency data

While an examination of the most frequent tags is interesting, the majority of the data is not covered by an analysis of the most frequent tags in the data. As previously stated in the introduction (and laid out in Table 4), hapaxes comprise the majority of tags in all three of the datasets. While the majority of the analysis on the hapaxes was done using LIWC (see Section 3), there was a small amount of frequency analysis done to more clearly show the long tail of the data. As we can see in Figures 11 - 13, there is an inverse relationship between the frequency of tags and their rate of occurrence. The first item that leaps out is that the rate of change declines rapidly across all three datasets. The following graphs depict how often *a specific frequency* (not a specific tag) occurs. For example, the count of the most frequent tag in the *Hannibal (TV)*, 2191, only occurs once. On the other end of the spectrum, hapaxes occur 35,754 times. Interestingly, this significant difference appears to follow Zipf’s law, which in this case reveals that there is an inverse relationship between a tag’s rank in the dataset and its frequency in the dataset. This is most obviously shown in Table 5, in which we see a significant drop in frequency

between the most frequent tag and the second most frequent tag in the *Hannibal (TV)* dataset. This trend is also seen in Figure #11, which shows the occurrence of a tag to be inversely proportional with its frequency, mirroring natural language use to a certain extent. This indicates that there is a much greater number of hapaxes, meaning that tags that are only used once greatly outnumber tags that are used more than once.

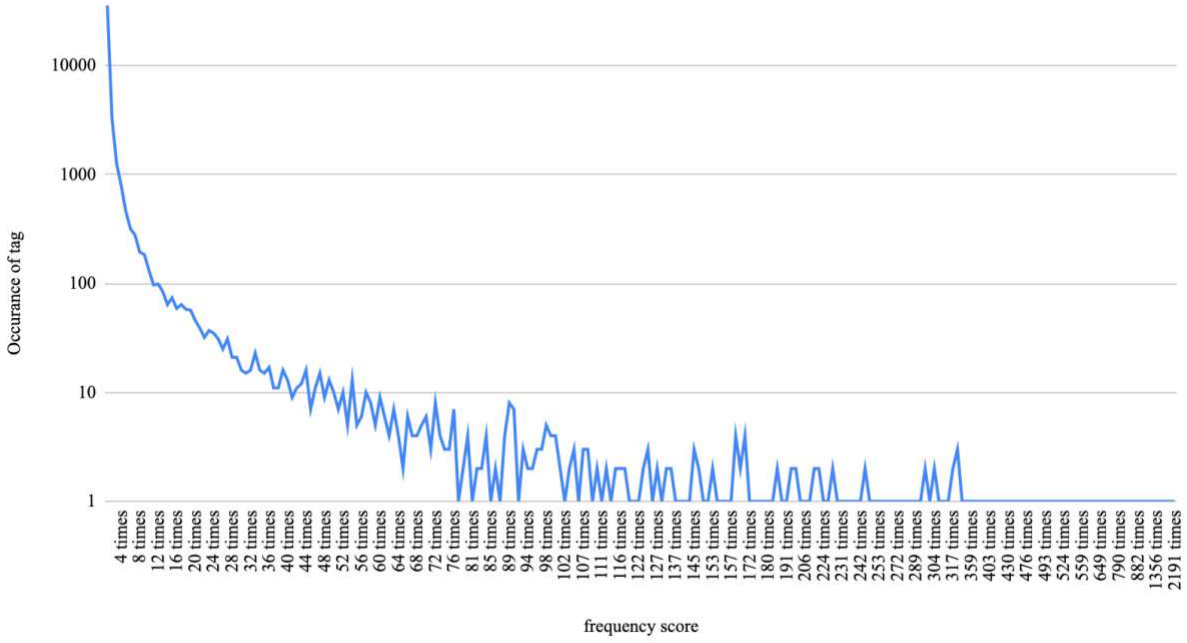


Figure 11: Tag frequency compared to the rate of tag occurrence in Hannibal (TV)

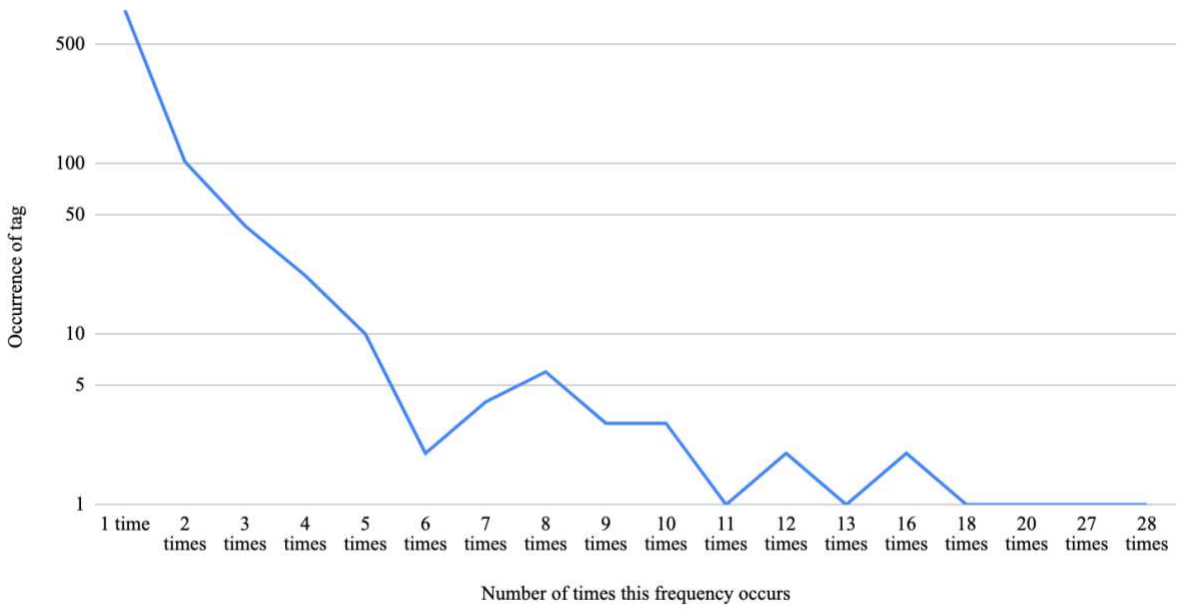


Figure 12: Tag frequency compared to the rate of tag occurrence in Hannibal Lecter (Hopkins Movies)

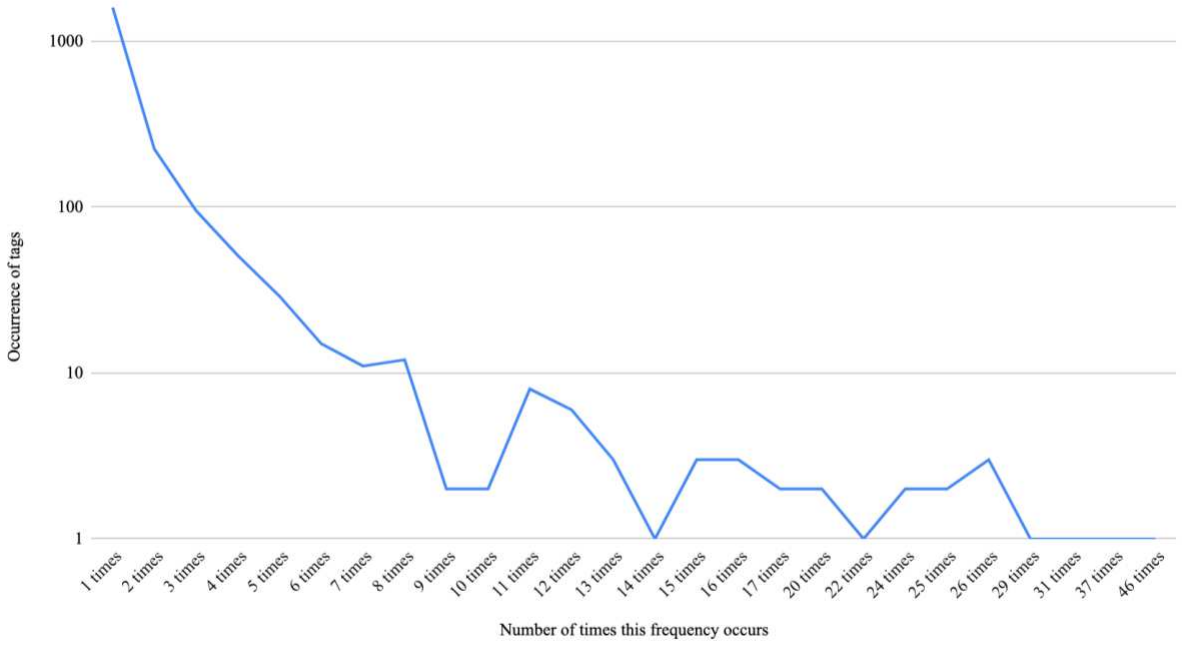


Figure 13: Tag frequency compared to the rate of tag occurrence in Hannibal Lecter (Thomas Harris Tetralogy)

Section 2.8. Conclusions

The most frequent tags in the dataset are similar across all three fandoms, the length of the tags is similar across fandoms, and the *Hannibal (TV)* fandom has a large amount of influence over the other two fandoms. The most frequent tag lists are mostly one- and two-word tags that do not relate directly to the fandoms, i.e., are not fandom-specific. The few tags that are fandom-specific represent characters and relationships that are specific to the media, i.e., 'clannibal' or 'hannibal loves will'. However, the most frequent tags do not represent the majority of the data in terms of frequency counts. The fact that the majority of tags in each dataset are unique tags is significant. This high degree of occurrence implies that the users engage in creative tagging practices very often, with a high degree of individuality.

Section 3. Linguistic Inquiry and Word Count Data Analysis

After performing frequency analysis, it was time to take a more in-depth look at the data using textual analysis. This analysis consists of examining the linguistic features of the tags as well as levels of logical thinking, candor, and authorial authority. Because the majority of the rich vocabulary exists in datasets consisting of hapaxes it was these that needed further analysis. Thus, it is important to note that this analysis was performed only on the hapaxes tags in the dataset, as they comprise over 75% of the data in each dataset. This data consists of 35,753 tags from the *Hannibal (TV Series)* dataset, 1,593 tags from the *Hannibal Lecter (Thomas Harris Tetralogy)* dataset, and 788 tags from the *Hannibal Lecter (Hopkins Movies)* dataset. Initial scoring comparison also showed no significant change in the data between the datasets only consisting of hapaxes and the full datasets. Lastly, because unique tags comprise over 75% of

each dataset, I determined that they contained the majority of the variety in the dataset and thus were the more interesting object of study.

LIWC was used to examine seven variables. Four of these variables are linguistic, consisting of the presence of personal pronouns in each tag, the word count of each tag, the use of character names, and the use of informal language in each tag. The other three variables were facets created by the software: analytical thinking, clout, and authentic/authenticity. These terms are defined in the *LIWC Operator's Manual* and outlined below:

- Analytical thinking is defined as the level of logical and hierarchical thinking in a text. A high analytical thinking score reflects formal, logical, and hierarchical thinking; lower scores reflect more informal thinking, and personal, here-and-now narrative thinking.
- Authentic/Authenticity is defined as the amount of candor within a text. Texts with higher authenticity scores display a more honest, personal, and disclosing attitude; lower scores suggest a more guarded, distanced form of discourse.
- Clout is defined as the amount of authorial authority in the text. A high score suggests that the author is speaking from the perspective of high expertise and is confident; a low clout score suggests a more tentative, humble, even anxious style.

Section 3.1. LIWC Summary Variables

As outlined above, there are three LIWC summary variables; analytical thinking (Analytic), authenticity (Authentic), and clout (Clout). These variables were formulated from research into language and speech patterns, most notably research into how function words can indicate levels of analytical thinking (Pennebaker et al., 2014), how authenticity can be predicted from linguistic style (Newman et al., 2003), and how pronoun use reflects social standing within

communities (Kacewiez et al., 2012). These three variables allow us to analyse the levels of emotion in a text and provide some insight into the authority of a text. In this case these variables were applied at the dataset level to provide insight into the trends of each fandom and how they were related. To this end, the mean, median, and mode of the Analytic, Authentic, and Clout variables were calculated for each fandom (see Table 9). All three of these variables are measured in a range of 1-100, where a score of 1 indicates a low level and 100 indicates a high level. Comparing the scores in this way gives us insight into the ways in which the datasets are both similar and different.

Fandom	Analytic			Authentic			Clout		
	Mean	Median	Mode	Mean	Median	Mode	Mean	Median	Mode
Hannibal (TV)	65.16	93.26	93.26	29.47	1.00	1.00	51.51	50.00	50.00
Hannibal Lecter (Thomas Harris Tetralogy)	77.53	93.26	93.26	25.24	1.00	1.00	53.76	50.00	50.00
Hannibal Lecter (Hopkins Movies)	79.26	93.26	93.26	28.84	1.00	1.00	52.82	50.00	50.00

Table 9: Mean, Median, and Mode of LIWC Summary Variables

The first variable we will examine is Analytic/Analytical thinking. The LIWC Language Manual defines a high analytic score as indicating a high level of logical thinking and formal language in the text analysed. The Analytic section of Table 9 describes the level of analytical thinking in all three datasets. Tags with high analytic scores do not usually contain personal pronouns or character names, they are short statements or noun phrases. In all three datasets there is a negative correlation between high analytic scores and personal pronoun use, i.e., personal pronouns are not often used in tags that have high analytic scores. (This will be covered

in more detail in section 3.2.1.) The majority of tags that have high analytic scores are also shorter on average than tags that have lower scores. (This will be discussed in more detail in section 3.2.3.) This makes sense, as shorter tags are often simple descriptive noun phrases or short statements that have little to no ambiguity, e.g., 'aftermath of violence' or 'character death'. As we can see from Table 9, the scores are quite high across all three datasets. There is also very little variability in the scores across the datasets. However, there is an interesting difference between the mean scores and the median and mode scores across all three datasets. While the median and the mode are consistently high, the mean is significantly lower. As we see in Table 9, the *Hannibal (TV Series)* dataset has the lowest average analytic score of the three datasets and the analytic score increases as the datasets get smaller. This seems to indicate that there are more tags with low analytic scores in the *Hannibal (TV Series)* dataset than in the *Hannibal Lecter (Thomas Harris Tetralogy)* dataset or the *Hannibal Lecter (Hopkins Movies)* dataset. This could be a function of the size of the fandom; however, the *Hannibal Lecter (Thomas Harris Tetralogy)* dataset is larger than the *Hannibal Lecter (Hopkins Movies)* dataset and it has a lower score (77.53 compared to 79.26).

The second variable is Authentic/Authenticity. This variable measures how candid or 'authentic' a text is. High authenticity scores indicate a more personal tone and a more open and disclosing attitude. On the other hand, low authenticity scores indicate greater detachment. All three datasets have a low average authenticity score. The median and mode for all three sets is 1.00 and the mean doesn't exceed 30.00. Detachment and distancing are often related to formal language so this tracks with the high levels of analytical thinking in the three datasets as well. There is also much less variation in the means of the three datasets as compared to the Analytic mean scores. While there is a significant amount of variation in the Analytic means, there is

much less in the Authority means. It should be noted that there are no personal pronouns (such as ‘I’, ‘you’, etc.) in tags that have an authenticity score of 1.00.

The third variable is Clout. This variable is defined as the level of authority in a text or how confident the author feels about the text. In this case, a high score indicates that the author of the text feels confident about their subject and has a high level of expertise. A low score on this variable suggests that the author is not confident and usually indicates a humbler tone and style. The scores for this variable are interesting in that the mean, median, and mode fall almost exactly in the middle of the range, indicating that overall, the tone of the datasets is neither excessively confident nor overly humble or tentative. This is interesting because based on the previous variables one would expect clout to skew one way or the other. For example, we might theorize that Authority and Clout might be directly correlated due to the relatedness of their definitions, however this is not the case in these datasets.

All of the fandoms have a middling average Clout, high averages in Analytics, and low Authentic averages. This implies that the datasets are largely more formal, logic driven, and analytical than informal and personal, with an amount of clout that is neither good nor bad.

Section 3.2. Linguistic Variables

After the summary variables were analysed the next step in the analysis was to examine the presence of various linguistic processes in the datasets. The four linguistic features that were examined were personal pronouns, the use of proper names (character names), the length of a tag as measured in words, and the use of informal language within the tags.

Two of the linguistic features examined in this section are parts of speech, that is, linguistic categories to which words are assigned related to the syntactic function they perform. The main parts of speech in English are as follows: nouns, pronouns, adjectives, determiners,

verbs, adverbs, prepositions, conjunctions, and interjections. The parts of speech that are examined in this analysis are personal pronouns and proper nouns in the form of character names. This information was gathered using LIWC. The following analyses were performed on the hapax datasets.

Section 3.2.1. Personal Pronouns

The first linguistic variable examined was the presence of personal pronouns in each dataset. For the purposes of this research, personal pronouns are pronouns that refer to grammatical persons, e.g. 'I', 'you', and 'he/she'. They indicate the gender of the person and number of people they are referring to. These pronouns, while they can indicate an inanimate object in wider language use, refer exclusively to persons in this context (see Section 3.2.2). The pronouns that were examined in this research are the first-person singular pronoun 'I', the first-person plural pronoun 'we', the second person singular pronoun 'you', the third person singular pronouns 'he/she', and the third person plural pronoun 'they'. The second person plural pronoun was excluded due to the ambiguity of the form of the pronoun, as English does not have only one form of the second person plural pronoun and most forms of the pronoun are informal. The singular form 'you' is also used as a plural form in some cases. (While Table 10 includes a second person plural pronoun category for balance, there are no second person plural pronouns in any of the datasets.)

Fandom	1st person pronoun		2nd person pronoun		3rd person pronoun	
	Sing. (I)	Plur. (we)	Sing. (you)	Plur.	Sing. (he/she)	Plur. (they)
Hannibal (TV)	7.1%	.87%	2.6%	n/a	3.7%	1.5%

Hannibal Lecter (Thomas Harris Tetralogy)	4.0%	.43%	1.2%	n/a	1.8%	1.2%
Hannibal Lecter (Hopkins Movies)	3.8%	.25%	1.1%	n/a	1.1%	.38%
Totals	14.9%	1.55%	4.9%	n/a	6.6%	3.08%

Table 10: Percentage of hapaxes that contain personal pronouns

As we can see in Table 10, the usage patterns of personal pronouns are correlated with dataset size to a certain extent; the larger the dataset, the more personal pronouns are used. Singular first-person pronouns are by far the most common. This indicates that authors include references to themselves in their tags (e.g., 'I'm not sorry' or 'why did i write this'). This creates a relationship between the audience and the author, the author actively inserting themselves into discourse. There is a positive relationship between the use of first-person pronouns and the length of tags; that is, first person pronouns are more likely to occur in longer tags. There is a significant difference between the usage of first-person pronouns and first-person plural pronouns, indicating that tags (and the authors that make them) don't often reference their work in the context of the wider community.

The second most common pronoun in the *Hannibal (TV)* dataset is the singular third person pronoun (e.g., 'he's jaded but he's stable' or 'why did hannibal do what he did'). This pronoun (either in the form 'he' or 'she') is largely used as a placeholder for character mentions in tags (see Section 3.2.2.). This indicates that there is some connection across user-generated tags. The use of a third person pronoun often implies the previous mention of a specific character in the tags, for example the tag 'hannibal lecter' might be followed by 'he's the worst'. This is an example of how multiple tags might be strung together in a sequence to create a narrative within

the metadata of a work. While this is an interesting observation of how third person pronouns are used in these datasets, ‘tag sequencing’ is not limited to use with pronouns. This will be discussed in further detail in chapter 5.

Following third person pronouns in popularity are second person pronouns. This category consists entirely of the word ‘you’, for example a tag containing a second person pronoun could be ‘you know you want it’ or ‘hannibal you made your bed’. These two tags exemplify the uses of the second person pronoun in this dataset: speaking to the audience and speaking to the characters. In the first example tag (‘you know you want it’) the author is speaking to the reader through the tag, creating a dialog with the reader.

Section 3.2.2. Character Names

The use of proper nouns like character names is below 50% in each dataset. Overall character names from the *Hannibal (TV)* occur in 10690 tags. This constitutes 29.89% of tags in the entire dataset, representing a relatively small portion of the dataset. Overall character names occur in 350 tags in the *Hannibal Lecter (Thomas Harris Tetralogy) dataset*. This constitutes 21.95% of the entire dataset. Character names occur in 134 tags in the *Hannibal Lecter (Hopkins Movies)* dataset, this constitutes 17.00% of the entire dataset.

The most popular character names in the datasets are that of the main characters, Hannibal Lecter and Will Graham. Other character names are used, but the names Hannibal Lecter and Will Graham (or variations thereof) are by far the most common. It is interesting to note that while it seems like there is little correlation between tag length and the use of character names, names are most often used within a noun phrase (for example ‘hannibal loves will’) and are rarely used by themselves in the additional tags category.

Section 3.2.3. Word Count of Tags

There is a great amount of variety in the length of the tags in each dataset. In all three datasets tags that have a length of two words are the most common, followed by tags that are three words long. This popularity of shorter tags is not surprising, as short noun phrases (such as ‘alternate universe’ or ‘established relationship’) are easier to standardize into usable tags than longer phrases. There is some linguistic theory behind this observation, as Zipf’s linguistic theory of abbreviation also states that languages tend towards shorter utterances over longer ones and there is research stating that Zipf’s law of abbreviation is a language universal (Bentz and Ferrer-i-Cancho, 2016).

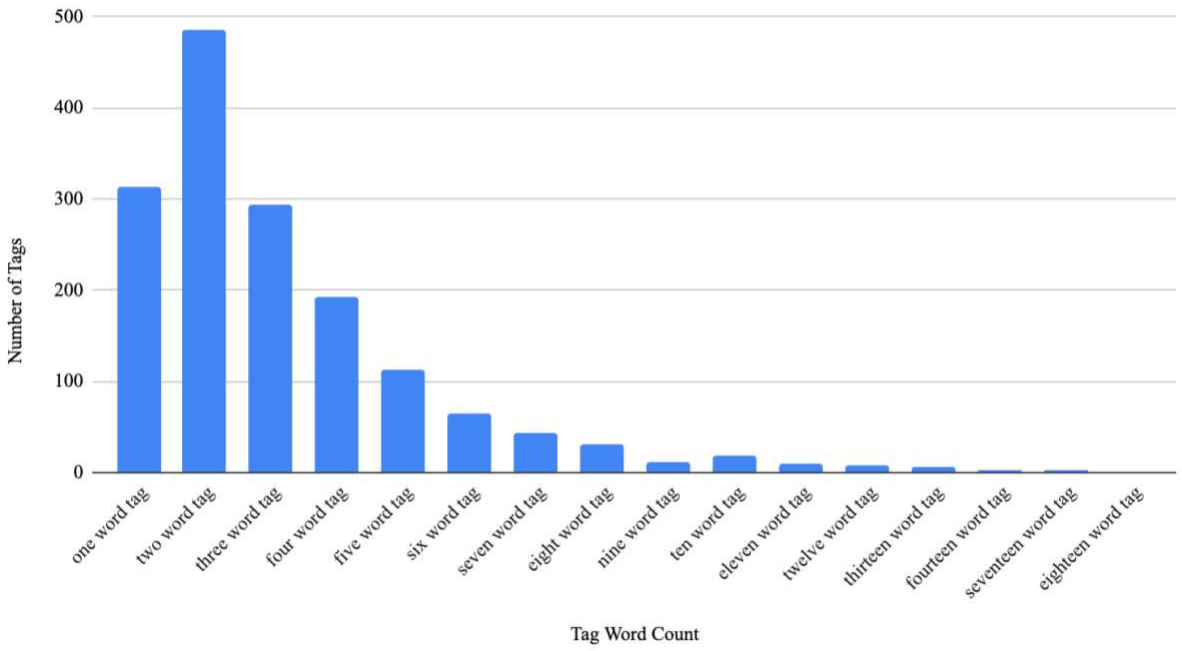


Figure 14: Hannibal Lecter (Thomas Harris Tetralogy) word count of tags in hapaxes dataset

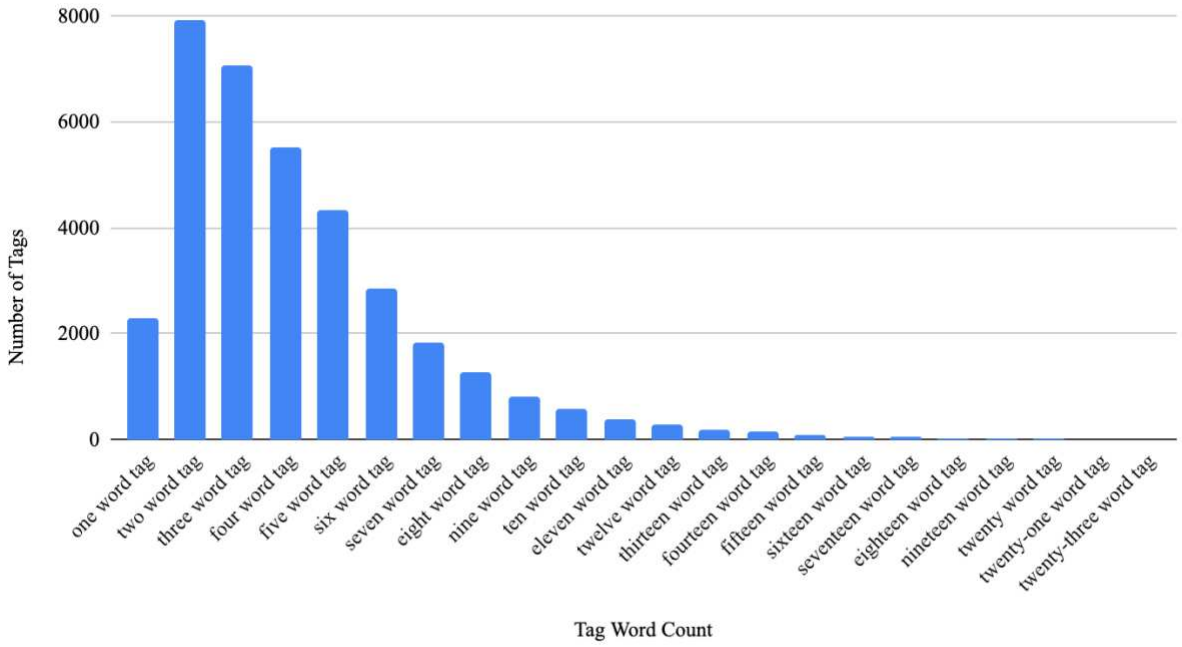


Figure 15: Hannibal (TV) word count of tags in hapaxes dataset

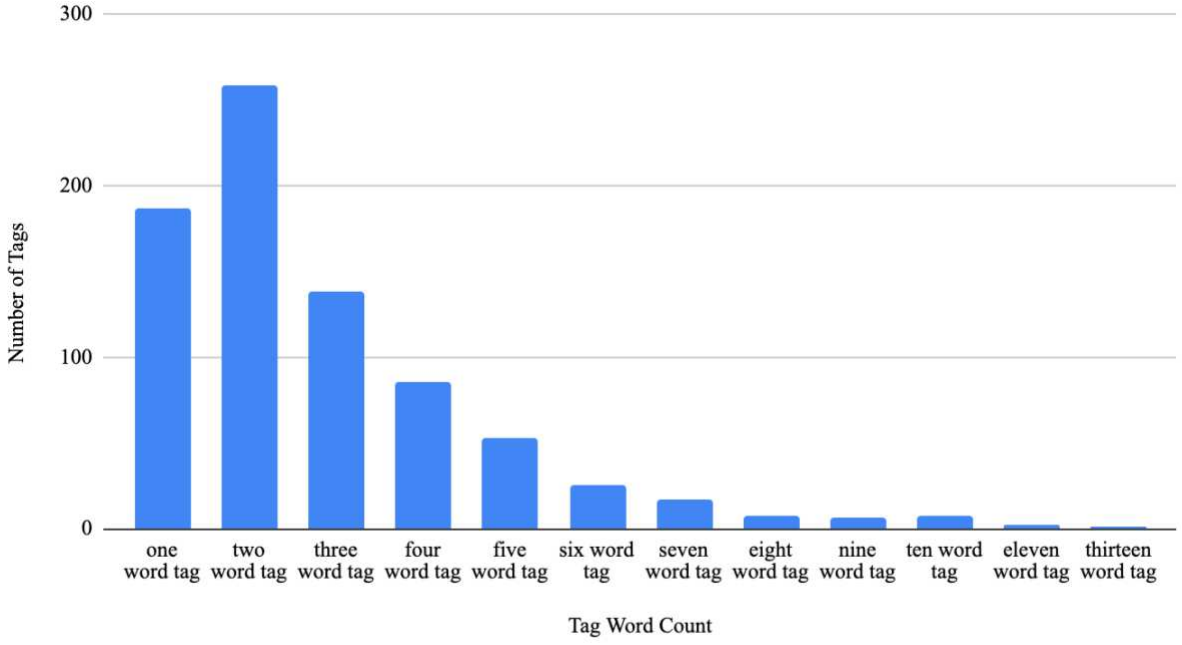


Figure 16: Hannibal Lecter (Hopkins Movies) word count of tags in hapaxes dataset

As we can see in Figures 14-16, the range of tag length varies in each dataset. The largest dataset, *Hannibal (TV)* has the largest range, with a tag length of twenty-three words. *Hannibal Lecter (Thomas Harris Tetralogy)* has the next largest range, with the longest tag length being eighteen words. The smallest dataset, *Hannibal Lecter (Hopkins Movies)* has the smallest range, with the longest tag containing thirteen words. This implies that there is some degree of correlation between the size of the dataset and the length of tags.

Section 3.2.4. Informal Language

The fourth linguistic variable that was explored was the use of informal language in unique tags. The types of language that fall under the informal language category in LIWC are assent words ('yeah', 'OK', etc.), filler words and phrases ('I mean', 'like', etc.), swear words ('fuck', 'shit', etc.), and 'netspeak' ('btw', 'lol', etc.). The amount of informal language presents a detailed picture of how formal or informal authors are in their tags.

	Tags containing Informal Language	Total Unique Tags	Percentage of dataset
Hannibal (TV)	2771	35753	7.75%
Hannibal Lecter (Thomas Harris Tetralogy)	74	1593	4.64%
Hannibal Lecter (Hopkins Movies)	29	788	3.68%

Table 11: Percentage of hapaxes that contain informal language

As we can see in Table 11, each dataset has a small amount of informal language. The amount of informal language seems to be slightly correlated with the size of the dataset, as the largest dataset has the highest percentage of informal language as compared to the two smaller datasets. This is interesting, as from this we could intuit that the larger the fandom, the more

informal language is used. However, this claim needs more analysis and would be an interesting path for further research.

Section 3.3. Conclusions

The data analysed in this chapter shows that there are both similarities and differences across the three fandoms. Short tags are the norm across all three datasets, with tags of one or two words being the most common length. A cross reference of the analytical thinking and word count reveals that one-word tags have a high degree of analytical thinking and a low degree of authenticity overall. This implies that there is some degree of correlation between the size of the dataset and the length of tags.

Pronouns are present in a significant portion of the data, although first person singular pronouns are by far the most common. Character names are included in a significant number of tags in each dataset. The most popular character names in the datasets are that of the main characters, Hannibal Lecter and Will Graham. While there is little correlation between tag length and the use of character names, names are most often used within a noun phrase (for example ‘hannibal loves will’) and are rarely used by themselves. Tag length data shows that shorter tags are more common than longer tags.

Chapter 5: Discussion and Conclusions

Archive of Our Own brings together a variety of fandoms and fans. These disparate communities are tied together not only by the site, but through the universal use of tags on the site. Tags on AO3 are versatile, allowing communication between an author and their audience, increasing findability, and enhancing a work's description. This research aimed to explore tags in a small selection of interrelated fandoms on AO3 with the purpose of examining patterns of tag use. This research answered the following questions:

What linguistic features exist in user-generated tags across fandoms on AO3?

What similarities and differences exist in user-generated tags across fandoms (television, books, and movies) on AO3?

What can we infer about discourse in the fan fiction communities from these user-generated tags?

The two major findings of this research are the use of personal pronouns and the presence of cross-fandom and fandom-specific tags. The presence of pronouns in the data indicates the formalization of a conversation between author and readers, along with interaction and characterization of characters in the work. The fact that tags can both cross fandom boundaries and be specific to an individual fandom shows AO3 as an environment that simultaneously allows different communities to interact while allowing them maintaining their tagging autonomy.

What linguistic features exist in user-generated tags across fandoms on AO3?

The consensus on user-generated tags in the wider discussion is that tags are short in length and focused on the content they are describing (Choi &Syn, 2016; Li et al., 2010; Tonkin, 2006). These conclusions hold true for the fandoms studied here. The majority of tags in all three

fandoms consist of short noun phrases consisting of 1-2 words. The largest syntactic category in each dataset is nouns. The main purpose of these short tags is description, as these short tags consistently describe the content of the work. The structure of these tags ranges from one-word genre descriptions such as ‘angst’ or ‘fluff’ to short phrases indicating the content of the work, e.g. ‘cannibalism’ or ‘explicit sexual content’. This reflects previous work into user-generated tags in various contexts (Choi & Syn, 2016). Similar to other online organization systems, the most popular tags in a user-generated tagging system tend to be short, consisting of a single word or a short phrase.

The most frequent tags in all three fandoms are short tags, indicating that shorter tags are used more, a trend that seems to be held across AO3. In a comparison of the top 25 most frequent tags in all three fandoms to the usage of those tags across the entire archive, the majority of the most frequent tags were popular across the entire archive.

Hannibal (TV Series)	**Hannibal Lecter (Thomas Harris Tetralogy)**	**Hannibal Lecter (Hopkins Movies)**	**AO3 most popular tags**
fluff	cannibalism	angst	fluff
murder husbands	angst	cannibalism	alternate universe
angst	fluff	romance	angst
post episode: s03e13 the wrath of the lamb	hurt/comfort	explicit sexual content	sexual content
anal sex	alternate universe	alternate universe	relationship(s)
cannibalism	murder	murder	sex
hannigram freeform	murder husbands	crossover	hurt/comfort
canon typical violence	explicit sexual content	clannibal	romance
alternate universe	oral sex	drama	friendship
established relationship	alternate universe canon divergence	blood	smut

hurt/comfort	hannibal is a cannibal	hurt/comfort	family
alternate universe canon divergence	post episode: s03e13 the wrath of the lamb	oral sex	humor
murder	hannigram freeform	violence	love
oral sex	smut	au	alternate canon
alpha/beta/omega dynamics	anal sex	fluff	oral sex
smut	sexual content	humor	alternate universe canon
hannibal is a cannibal	blood	hannibal is a cannibal	violence
blowjobs	crossover	mental health issues	established relationship
hannibal loves will	violence	post canon	deviates from canon
anal fingering	community: kink_bingo	post episode: s03e13 the wrath of the lamb	alternate universe canon divergence
blood	dark will	serial killers	firsts
masturbation	romance	smut	kissing
kissing	serial killers	alternate universe canon divergence	BDSM
slow burn	clannibal	nightmares	death
dark will	masturbation	rough sex	porn

Table 12: Comparison of Hannibal fandoms top 25 tags with AO3's top 25 tags

As we can see in Table 12, some overlap exists between the most popular tags in the three fandoms studied and the top 25 most popular tags on AO3 as a whole. The most popular tag in the *Hannibal (TV)* fandom, ‘fluff’, is the most popular tag on the site as a whole, something that is not true in *Hannibal Lecter (Thomas Harris Tetralogy)* and *Hannibal Lecter (Hopkins Movies)*. This is also true of the third most popular tag in the *Hannibal (TV)* fandom; ‘angst’. Both ‘fluff’ and ‘angst’ appear in the top three most popular tags in the two smaller datasets.

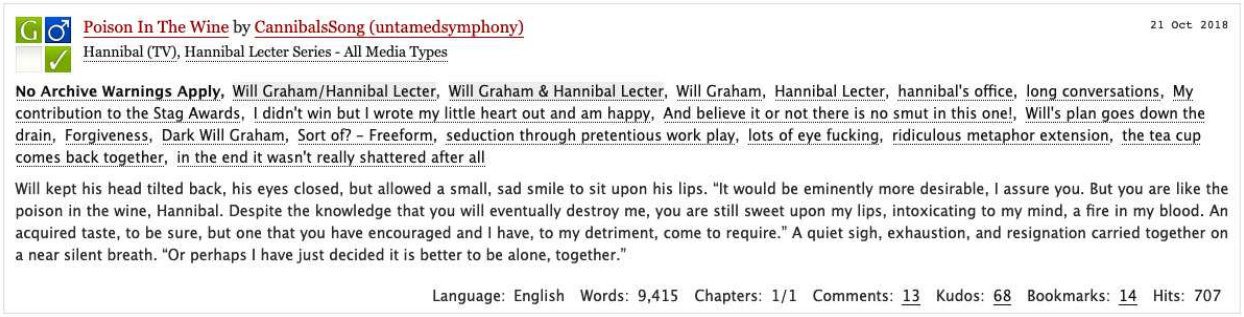
The data shows that while the majority of the tags are not fandom-specific, there are some tags that are only used within the three fandoms. The five most frequent tags in each fandom are for the most part not fandom specific. The only fandom-specific tags are ‘murder husbands’, ‘post episode: s03e13 the wrath of the lamb’, and ‘cannibalism’, the last of which is only adjacent to being fandom-specific as it is most used in Hannibal Lecter fandoms on AO3. The large presence of cross-fandom tags implies that bleed-through exists between fandoms on AO3. The ‘murder husbands’ tag references a term that originated in the fandom rather than in the show itself (although the show referenced the term in season 3). Thus, we see that it is not only the canon material that users draw from to create tags, users on AO3 also draw from other fans.

There is a surprisingly high use of personal pronouns in all three fandoms. This is a departure from tagging systems used on other websites, where tags are more traditionally used for findability and thus often do not include pronouns. Strong positive correlations are also present between the length of a tag and the presence of pronouns within that tag. Past research (Pennebaker and Chung, 2007) indicates that the presence of pronouns can provide insight into the psychological state of the author. In the case of this data, the presence of pronouns shows a high level of community and a sense of cohesion in the fandoms. The most common pronouns in each fandom are first person singular and first-person plural pronouns, followed by third person pronouns, and finally second person pronouns. From this we can infer that when pronouns are used, they are used by the tag’s author to connect with their audience and establish a discourse with them. The use of first-person pronouns in the data extends Jenkins’ participatory culture model into tags on AO3 by bringing the author’s own perspective into their contributions to their community and placing themselves within the community. This model of participatory culture is

also seen in a much smaller population of second person pronouns. These pronouns are almost exclusively used to draw in the reader, bringing the audience into conversation with the author. For example, in the tag ('you know you want it') the author is speaking to the reader through the tag, using the second person pronoun to create a dialog. Third person pronouns are the second most used type of pronoun in all three fandoms, and they are most often used to reference characters within a work. Personal pronouns are most common in the largest fandom, the *Hannibal (TV)* dataset. More investigation needs to be done into why this is the case, although it may indicate a relationship between the number of personal pronouns used in tags and the size of the fandom.

In their study of the AO3's metadata practices, Gursoy et al. (2018) identified three general tag categories: tags that identify elements in a work, tags that identify how those elements are used, and tags that provide emotional context (either of the author or characters). From the observation of these categories, they suggest that metadata systems can offer more than description and findability. They also observed that the sequencing of tags was far more common with expressive tags, that is, longer and more complex tags than declarative tags. This pattern of tagging has not been observed in more mainstream tagging systems such as those used on Twitter or Flickr. However, this has been observed on Tumblr, another hub of fan activity (Rose, 2016; Bourlai, 2018).

The importance of tags is also backed up by how fan fiction works are presented on AO3. If we look at how a fan fiction work is presented (see Figure 10), we see that tags are presented at the same level as the work summary, placing them at a high level of importance for the reader. This implies that tags are meant to be seen and that they are used as a descriptive tool and as a communication tool.


Poison In The Wine by CannibalsSong (untamedsymphony) 21 Oct 2018
Hannibal (TV), Hannibal Lecter Series - All Media Types

No Archive Warnings Apply, Will Graham/Hannibal Lecter, Will Graham & Hannibal Lecter, Will Graham, Hannibal Lecter, hannibal's office, long conversations, My contribution to the Stag Awards, I didn't win but I wrote my little heart out and am happy, And believe it or not there is no smut in this one!, Will's plan goes down the drain, Forgiveness, Dark Will Graham, Sort of? - Freeform, seduction through pretentious work play, lots of eye fucking, ridiculous metaphor extension, the tea cup comes back together, in the end it wasn't really shattered after all

Will kept his head tilted back, his eyes closed, but allowed a small, sad smile to sit upon his lips. "It would be eminently more desirable, I assure you. But you are like the poison in the wine, Hannibal. Despite the knowledge that you will eventually destroy me, you are still sweet upon my lips, intoxicating to my mind, a fire in my blood. An acquired taste, to be sure, but one that you have encouraged and I have, to my detriment, come to require." A quiet sigh, exhaustion, and resignation carried together on a near silent breath. "Or perhaps I have just decided it is better to be alone, together."

Language: English Words: 9,415 Chapters: 1/1 Comments: 13 Kudos: 68 Bookmarks: 14 Hits: 707

Figure 17: Appearance of a fan fiction work entry on AO3

What similarities and differences exist in user-generated tags across fandoms (television, books, and movies) on AO3?

Overall, there are more similarities across the three fandoms than differences. As seen in the analysis chapter, very little difference exists between the three datasets in LIWC variables and linguistic features. All three fandoms have similar scores in both LIWC variables and linguistic features, indicating similar levels and methods of engagement with the material across all three datasets.

The most obvious similarity we see in the three datasets is the large amount of overlap in content between them. At first glance there is little difference in the types of tags that are used in the television fandom versus the book fandom versus the film fandom. The clearest difference between the fandoms is in the characters that are referenced in the tags of each dataset. The character of Clarice Starling is a presence in the movie fandom and is absent elsewhere. The *Hannibal (TV*) fandom has a greater focus on the character of Will Graham compared to the *Hannibal Lecter (Thomas Harris Tetralogy)* and *Hannibal Lecter (Hopkins Movies)* fandoms, which have a greater focus on Hannibal Lecter as a character. This is likely due to the structure of the television show, which sets up Will Graham as the audience surrogate and point of view character of the narrative.

Another interesting content note is the presence of ‘post episode: s03e13 the wrath of the lamb’ in all three datasets. This tag refers to the last episode of the television series, it has no connection to the books or the movies. The presence of this tag shows that there is a large amount of overlap between the three fandoms and that the television series has a large influence over the other two fandoms.

There are no significant differences in tagging practices between media types on AO3 based on the data gathered in this study. More differences were observed between the larger *Hannibal (TV)* fandom and the two smaller fandoms, *Hannibal Lecter (Thomas Harris Tetralogy)* and *Hannibal Lecter (Hopkins Movies)* than there were differences between the two smaller fandoms. This implies that the differences in tagging behavior have more to do with fandom size rather than the type of media that users are writing about.

What can we infer about discourse in the fan fiction communities from these user-generated tags?

The use of personal pronouns in tags indicates the formalization of a conversation between author and readers, along with interaction and characterization of characters in the work. Generally, the popularity of short tags indicates an interest in description over community interaction. The significant presence of hapax legomena in each fandom suggests that there is a high amount of tag diversity in each fandom and that users tend towards creating their own tags over reusing established tags, although this bears more investigation and needs to be cross referenced with frequency counts.

Further Research

Further research needs to be done, widening the research scope to encompass more fandoms and more tags. A cross sectional study that investigates tagging patterns between unrelated fandoms is the next logical step, as well as expanding the research to larger fandoms. In a slightly different direction, applying the methodology showcased in Gursoy et. al would be another interesting avenue to explore, especially when combined with the analysis already done. It would definitely give more insight into fan discourse than is offered in this research. A serious flaw of this research is the lack of control variables that would help to determine whether the differences in tagging behaviour between fandoms is correlated with fandom size or the type of media for which the fandom is creating content.

Sequential tagging is another idea that could use more exploration in further work. The idea of complex tags being added onto simple tags to enhance their meaning is fascinating. This would also dovetail nicely into a comparison of Tumblr and AO3, as sequential tagging is common on Tumblr.

The most basic implications of this work is that while AO3 tags show similarities to other tagging environments, such as Tumblr, there are distinct differences. These differences are worth exploring for the insight they give us into online communities and how fans might potentially communicate with other fans online. It also indicates that, based on the similarities to Tumblr tags, that fan communities have developed their own style of tagging that is not seen elsewhere online.

The Archive of Our Own is a unique tagging environment that is worthy of study. This research has shown that the site's users create their own communities through the use of tagging and that the freeform style of the user-generated tags allow users to converse with each other through tagging, something that should be more common online.

Works Cited

About Home | Archive of Our Own. (n.d.). Retrieved March 13, 2018, from https://archiveofourown.org/about

AFI's 100 YEARS…100 HEROES & VILLAINS | American Film Institute. (n.d.). Retrieved February 3, 2020, from https://www.afi.com/afis-100-years-100-heroes-villians/

An Archive Of One's Own. (n.d.). Retrieved February 20, 2019, from https://astolat.livejournal.com/150556.html

AO3 News. "Tumblr Tagging and Why It's a Problem on Non-Tumblr Sites." Accessed November 19, 2018. http://ao3org.tumblr.com/post/58351045314/tumblr-tagging-and-why-its-a-problem-on.

Attu, R., & Terras, M. (2017). What people study when they study Tumblr: Classifying Tumblr-related academic research. *Journal of Documentation*, *73*(3), 528–554. https://doi.org/10.1108/JD-08-2016-0101

Beautiful Soup Documentation—Beautiful Soup 4.4.0 documentation. (n.d.). Retrieved February 12, 2020, from https://www.crummy.com/software/BeautifulSoup/bs4/doc/

Bentz, C., & Ferrer-i-Cancho, R. (2016). *Zipf's law of abbreviation as a language universal*. http://dx.doi.org/10.15496/publikation-10057

Bourlai, Elli E. "'Comments in Tags, Please!': Tagging Practices on Tumblr." *Discourse, Context & Media*, Discourse of Social Tagging, 22 (April 1, 2018): 46–56. https://doi.org/10.1016/j.dcm.2017.08.003.

Brydges, G., Carson, C., Chase, L., Dawson, A., Dionne, C., & O'Brien, M. (2017). Fanfiction. *YA Hotline*, *0*(104). https://ojs.library.dal.ca/YAHS/article/view/6938

Bryan Fuller on Twitter: "I RAISE A GLASS TO THE FAITHFUL & HUNGRY #HANNIBAL TWITTERERS FOR YOUR CONTINUED SUPPORT AND ENTHUSIASM #FEEDYOURFEAR http://t.co/kAXwwia4Xb" / Twitter. (n.d.). Twitter. Retrieved February 10, 2020, from https://twitter.com/bryanfuller/status/340265816379506689?ref_url=https%3a%2f%2fwww.buzzfeed.com%2fellievhall%2fmeet-the-fannibals-tvs-newest-and-most-intense-fandom

Burrows, J. F. *Computation into Criticism: A Study of Jane Austen's Novels and an Experiment in Method*. Oxford [Oxfordshire], 1987. http://hdl.handle.net/2027/mdp.39015013097244.

Chaney, K. & Liebler, R. (2007). Canon vs. Fanon: Folksonomies of Fan Culture. In Massachusetts Institute of Technology.–2007. URL: http://web. mit. edu/comm-forum/mit5/papers/Chaney_Liebler_MIT5.pdf.

Charmaz, Kathy. *Constructing Grounded Theory*. 2nd ed. London: Sage, 2014.

Coppa, F. (Ed.). (2017). *The fanfiction reader: Folk tales for the digital age*. University of Michigan Press.

Cattuto, Ciro, Vittorio Loreto, and Luciano Pietronero. "Semiotic Dynamics and Collaborative Tagging." *Proceedings of the National Academy of Sciences* 104, no. 5 (January 30, 2007): 1461. https://doi.org/10.1073/pnas.0610487104.

Chua, A., Hoe-Lian Goh, D., Razikin, K., & Sian Lee, C. (2009). Resource discovery through social tagging: a classification and content analytic approach. *Online Information Review*, *33*(3), 568–583. https://doi.org/10.1108/14684520910969961

Cho, Ji Young, and Eun-Hee Lee. "Reducing Confusion about Grounded Theory and Qualitative Content Analysis: Similarities and Differences." *The Qualitative Report*, August 11, 2014. https://www.questia.com/library/journal/1G1-381836298/reducing-confusion-about-grounded-theory-and-qualitative.

Choi, Y., & Syn, S. Y. (2016). Characteristics of tagging behavior in digitized humanities online collections. *Journal of the Association for Information Science and Technology*, *67*(5), 1089–1104. https://doi.org/10.1002/asi.23472

Collins, Sean T., and Sean T. Collins. "40 Greatest TV Villains of All Time." *Rolling Stone* (blog), September 4, 2019. https://www.rollingstone.com/tv/tv-lists/40-greatest-tv-villains-of-all-time-26500/.

Complex. "Bryan Fuller Knows You're Reading Into 'Hannibal's' Homoeroticism, and He Thinks It's Hilarious." Accessed February 12, 2020. https://www.complex.com/pop-culture/2014/09/bryan-fuller-hannibal-interview-slash-fiction.

Daglas, S., Kakali, C., Kakavoulis, D., Koumaki, M., & Papatheodorou, C. (2012). A Methodology for Folksonomy Evaluation. In P. Zaphiris, G. Buchanan, E. Rasmussen, & F. Loizides (Eds.), *Theory and Practice of Digital Libraries* (pp. 247–259). Springer Berlin Heidelberg.

Demme, Jonathan. *The Silence of the Lambs*. Crime, Drama, Thriller. Strong Heart/Demme Production, Orion Pictures, 1991.

Eco, U. (1989). *The Open Work*. Cambridge, Mass.: Harvard University Press.

Eynard, D., Mazzola, L., & Dattolo, A. (2013). Exploiting tag similarities to discover synonyms and homonyms in folksonomies. *Software: Practice and Experience*, *43*(12), 1437–1457. https://doi.org/10.1002/spe.2150

FanFiction.Net—Fanlore. (n.d.). Retrieved April 13, 2021, from https://fanlore.org/wiki/FanFiction.Net

FanFiction.Net's NC-17 Purges: 2002 and 2012—Fanlore. (n.d.). Retrieved February 20, 2019, from https://fanlore.org/wiki/FanFiction.Net%27s_NC-17_Purges:_2002_and_2012

FanLib—Fanlore. (n.d.). Retrieved February 20, 2019, from https://fanlore.org/wiki/FanLib

FanLib.com—Terms of Service. (2007, October 31). https://web.archive.org/web/20071031111520/http://www.fanlib.com/termsOfUse.do

Fiesler, C., Morrison, S., & Bruckman, A. (2016). An Archive of Their Own: A Case Study of Feminist HCI and Values in Design. *CHI*. https://doi.org/10.1145/2858036.2858409

Golder, S. A., & Huberman, B. A. (2006). Usage patterns of collaborative tagging systems. *Journal of Information Science*, *32*(2), 198–208. https://doi.org/10.1177/0165551506062337

Gursoy, A., Wickett, K., & Feinberg, M. (2018). Understanding tag functions in a moderated, user-generated metadata ecosystem. *Journal of Documentation*, *74*(3), 490–508. https://doi.org/10.1108/JD-09-2017-0134

Guy, M., & Tonkin, E. (2006). Folksonomies: Tidying up Tags? *D-Lib Magazine*, *12*(1). Retrieved from https://researchportal.bath.ac.uk/en/publications/folksonomies-tidying-up-tags

Hafner, Katie, and Lyon, Matthew. *Where Wizards Stay up Late: The Origins of the Internet*. New York, NY: Simon & Schuster, 1998.

Hall, Ellie. "Meet The 'Hannibal' Fannibals, TV's Newest And Most Intense Fandom." BuzzFeed. Accessed February 3, 2020. https://www.buzzfeed.com/ellievhall/meet-the-fannibals-tvs-newest-and-most-intense-fandom.

Hapax legomenon, n. (n.d.). In *OED Online*. Oxford University Press. Retrieved September 8, 2020, from http://www.oed.com/view/Entry/84040

Harris, Thomas. *Hannibal*. New York: Delacorte Press, 1999.

Harris, Thomas. *Hannibal Rising*. New York: Delacorte Press, 2006.

Harris, Thomas. *Red Dragon*. New York: Putnam, 1981.

Harris, Thomas. *The Silence of the Lambs*. New York: St. Martin's Press, 1988.

Hellekson, K. (2009). A Fannish Field of Value: Online Fan Gift Culture. *Cinema Journal*, *48*(4), 113–118.

Herzog, A. E. (2012). "But this is my story and this is how I wanted to write it": Author's notes as a fannish claim to power in fan fiction writing. *Transformative Works and Cultures*, *11*. https://doi.org/10.3983/twc.2012.0406

Hill, H. L., & Pecoskie, J. (2014). Iterations and Evolutions: Paratext and Intertext in Fanfiction. *Examining Paratextual Theory and Its Applications in Digital Culture*, 143–158. https://doi.org/10.4018/978-1-4666-6002-1.ch008

Neill Hoch, I. (2018). Content, conduct, and apologies in Tumblr fandom tags. *Transformative Works and Cultures*, *27*. https://doi.org/10.3983/twc.2018.1198

IGN. "Hannibal Lecter - #8 Top 100 Villains." Accessed February 3, 2020. https://ca.ign.com/top/top-100-villains/8

Jenkins, H. (2013). *Textual poachers: Television fans and participatory culture* (Updated 20th anniversary ed.). Routledge.

Jenkins, H., Ford, S., & Green, J. (2013). *Spreadable media: creating value and meaning in a networked culture*. New York: New York University Press.

Jenkins, H. (2006). *Convergence culture: where old and new media collide*. New York: New York University Press.

Jensen, T. (2014). Considering Collaborative Creativity. Akademisk kvarter/Academic Quarter, 9, 60-72.

Johnson, S. F. (2014). Fan fiction metadata creation and utilization within fan fiction archives: Three primary models. *Transformative Works and Cultures*, *17*. https://doi.org/10.3983/twc.2014.0578

Jorgensen, Corinne. "Image Access, the Semantic Gap, and Social Tagging as a Paradigm Shift," 2007. https://repository.arizona.edu/handle/10150/105339.

Kacewicz, E., Pennebaker, J. W., Davis, M., Jeon, M., & Graesser, A. C. (2014). Pronoun Use Reflects Standings in Social Hierarchies. *Journal of Language and Social Psychology*, *33*(2), 125–143. https://doi.org/10.1177/0261927X13502654

Kem, J. F. (2005). Cataloging the Whedonverse: Potential roles for librarians in online fan fiction.

Kipp, M. E. I. (2007). @toread and Cool : Tagging for Time, Task and Emotion. Retrieved from http://eprints.rclis.org/9868/

Körner, C., Benz, D., Hotho, A., Strohmaier, M., & Stumme, G. (2010). Stop Thinking, Start Tagging: Tag Semantics Emerge from Collaborative Verbosity. In *Proceedings of the 19th International Conference on World Wide Web* (pp. 521–530). New York, NY, USA: ACM. https://doi.org/10.1145/1772690.1772744

Lawrence, K. F., & Schraefel, M. C. (2006). Freedom and Restraint Tags, Vocabularies and Ontologies. In *2006 2nd International Conference on Information Communication Technologies* (Vol. 1, pp. 1745–1750). https://doi.org/10.1109/ICTTA.2006.1684649

Leavenworth, M. L. (2014). The Paratext of Fan Fiction. *Narrative*, *23*(1), 40–60. https://doi.org/10.1353/nar.2015.0004

Lin, N., Li, D., Ding, Y., He, B., Qin, Z., Tang, J., Li, J., & Dong, T. (2012). The dynamic features of Delicious, Flickr, and YouTube. *Journal of the American Society for Information Science and Technology*, *63*(1), 139–162. https://doi.org/10.1002/asi.21628

Literary Hub. "40 of the Best Villains in Literature," June 11, 2018. https://lithub.com/40-of-the-best-villains-in-literature/.

LiveJournal—Fanlore. (n.d.). Retrieved February 20, 2019, from https://fanlore.org/wiki/LiveJournal

Lu, C., Park, J., & Hu, X. (2010). User tags versus expert-assigned subject terms: A comparison of LibraryThing tags and Library of Congress Subject Headings. *Journal of Information Science*, *36*(6), 763–779. https://doi.org/10.1177/0165551510386173

Malmkjaer, Kirsten. *Linguistics Encyclopedia*. Routledge, 2004. https://doi.org/10.4324/9780203644645.

Mann, Michael. *Manhunter*. Crime, Horror, Mystery, Thriller. De Laurentiis Entertainment Group (DEG), Red Dragon Productions S.A., 1986.

Marlow, C., Naaman, M., Boyd, D., & Davis, M. (2006). HT06, Tagging Paper, Taxonomy, Flickr, Academic Article, to Read. In *Proceedings of the Seventeenth Conference on Hypertext and Hypermedia* (pp. 31–40). New York, NY, USA: ACM. https://doi.org/10.1145/1149941.1149949

Mathes, Adam. "Folksonomies - Cooperative Classification and Communication Through Shared Metadata," December 2004. https://adammathes.com/academic/computer-mediated-communication/folksonomies.html.

Mayring, Philipp. "Qualitative Content Analysis." *Forum: Qualitative Social Research* 1, no. 2 (June 2000): 105–14.

McCulloch, E., & Macgregor, G. (2006). Collaborative tagging as a knowledge organisation and resource discovery tool. *Library Review*, *55*(5), 291–300. https://doi.org/10.1108/00242530610667558

Jason Morrison, P. (2008). Tagging and searching: Search retrieval effectiveness of folksonomies on the World Wide Web. *Information Processing & Management*, *44*(4), 1562–1579. https://doi.org/10.1016/j.ipm.2007.12.010

Moulaison, H. L. (2008). Social Tagging in the Web 2.0 Environment: Author vs. User Tagging. *Journal of Library Metadata*, *8*(2), 101–111. https://doi.org/10.1080/10911360802087325

Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying Words: Predicting Deception from Linguistic Styles. *Personality and Social Psychology Bulletin*, *29*(5), 665–675. https://doi.org/10.1177/0146167203029005010

O'Reilly, T. (n.d.). What Is Web 2.0. Retrieved March 2, 2019, from https://oreilly.com{file}

Pearson, R. (2010). Fandom in the Digital Era. *Popular Communication*, *8*(1), 84–95. https://doi.org/10.1080/15405700903502346

Pennebaker, James W., Cindy K. Chung, Joey Frazee, Gary M. Lavergne, and David I. Beaver. "When Small Words Foretell Academic Success: The Case of College Admissions Essays." *PLOS ONE* 9, no. 12 (December 31, 2014): e115844. https://doi.org/10.1371/journal.pone.0115844.

Piantadosi, Steven T. "Zipf's Word Frequency Law in Natural Language: A Critical Review and Future Directions." *Psychonomic Bulletin & Review* 21, no. 5 (October 2014): 1112–30. https://doi.org/10.3758/s13423-014-0585-6.

Ratner, Brett. *Red Dragon*. Crime, Drama, Thriller. Universal Pictures, Dino De Laurentiis Company, Metro-Goldwyn-Mayer (MGM), 2002.

Scolari, Carlos. "Digital Eco_logy." *Information, Communication & Society* 12, no. 1 (February 1, 2009): 129–48. https://doi.org/10.1080/13691180802520798.

Scott, Ridley. *Hannibal*. Crime, Drama, Thriller. Metro-Goldwyn-Mayer (MGM), Universal Pictures, Dino De Laurentiis Company, 2001.

Scrapy 2.1 documentation—Scrapy 2.1.0 documentation. (n.d.). Retrieved June 13, 2020, from https://docs.scrapy.org/en/latest/index.html

Sinclair, Stéfan, and Geoffrey Rockwell. "Voyant Tools," 2020. https://voyant-tools.org/.

Tags FAQ | Archive of Our Own. (n.d.). Retrieved March 2, 2019, from https://archiveofourown.org/faq/tags?language_id=en

Tausczik, Y. R., & Pennebaker, J. W. (2010). The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology*, *29*(1), 24–54. https://doi.org/10.1177/0261927X09351676

Tonkin, Emma. "Searching the Long Tail: Hidden Structure in Social Tagging." *Advances in Classification Research Online* 17, no. 1 (October 7, 2006): 1–10. https://doi.org/10.7152/acro.v17i1.12494.
Usenet—Fanlore. (n.d.). Retrieved February 20, 2019, from https://fanlore.org/wiki/Usenet

Webber, Peter. *Hannibal Rising*. Adventure, Crime, Drama, Horror, Thriller, War. Young Hannibal Productions, Carthago Films S.a.r.l., Dino De Laurentiis Company, 2007.

Zipf, G.K. *The Psycho-Biology of Language: An Introduction to Dynamic Philology*. The Psycho-Biology of Language: An Introduction to Dynamic Philology. Oxford, England: Houghton Mifflin, 1935.