

The Voices of Doctor Who – How Stylometry Can be Useful in Revealing New Information About TV Series

Joanna Byszuk <joanna_dot_byszuk_at_ijp_dot_pan_dot_pl>, Institute of Polish Language, Polish Academy of Sciences

Abstract

This article presents possibilities of effectively using stylometric methods popular in the analysis of literary texts in the study of texts written for television, on the example of Doctor Who. The article examines the changes driving the development of the show moving from character-oriented in the so-called "Classic Who" (1963-1989) to showrunner-oriented in its revival started in 2005. It also seeks to describe stability of the characterization of the protagonist as evidenced in the dialogues, and to discuss authorial relations between showrunners and their teams.

Introduction^[1]

Distant reading, or more generally: exploration of large collections of cultural objects with quantitative methods, has long been successfully applied not only in its original literary (prose) scope as proposed by Moretti (2000) but also in texts from other domains, e.g. journalistic reviews (e.g. Benatti and Tonra 2015), poetry (e.g. Plecháč and Flaišman 2017), blog posts (e.g. Schler et al. 2006), tweets (e.g. Juola et al. 2011) etc. This paper seeks to explore the possible application of methods established in stylometry, such as hierarchical clustering, network analysis (both discussed e.g. in Eder 2017) or contrastive analysis with Zeta (Craig and Kinney 2009, Burrows 2007), to audiovisual works, or, more precisely, to their textual layer alone. Furthermore, the paper aims to establish some conditions that need to be fulfilled to conduct such studies with respect to the existing traditions in both film studies and linguistics research.

The main substantive interest of this paper lies in presenting creative relations within television series production teams, with specific focus on the shift from character- to author-oriented script-writing. The first section of the paper sketches relations between distant reading and television studies, and summarizes existing research on quantitative studies of television discourse. The second section briefly sketches the outline of the show discussed in the case study in section three, a British science-fiction series Doctor Who. The third section proposes possible applications of quantitative methodology, namely stylometry, to exploring inter- and intra-textual relations in television shows, and the fourth presents the results of the case study, verifying the usefulness of the approach proposed here.

It is true that a focus on the textual layer alone may seem as anathema to both traditional and digital film scholars. Not for a moment can one forget that there is more to moving (as well as talking) pictures than whatever their characters have to say. This does not mean, however, that quantitative analysis of that one aspect of film is insufficient or frivolous. At worst, the text of film dialogue may be examined quantitatively as any other text and thus constitutes a valid area of study in this limited scope. More hopefully, however, holistic approaches to film text – approaches that could combine the study of its textual and its visual aspects – may be devised in time. Even the most ambitious specialists in quantitative analysis of text do not pretend that literary novels are just about word counts. This is even more true of theatre, and yet theatre is a major object of textual studies: to cite the most obvious example, quantitative analysis of word frequencies is now an accepted method – and indeed the most reliable – in Shakespearean authorship attribution, and its importance is further highlighted by the oftentimes heated disputes between its practitioners.^[2]

Still, the text of film dialogue alone is of interest to quantitative literary studies. It is exactly this doubtful state of

1

2

3

4

authorship concept and execution in television series that seems provocation enough to address some of the inescapable questions. For instance, is it at all possible to observe and to measure authorial impact in the textual layer of film alone? Can computational methods help us shed some light on the validity of the concept of authorship as argued by Martin (2013) and Jensen (2017) and as summarized below? While we know that authorship attribution methods can help us distinguish between authors in the difficult contexts of short or multi-authored texts, are they also effective with texts that undergo many editions and possibly have writers imitate the style of a showrunner; with texts that are but one of the many aspects – some may say, a fairly insignificant one – of the multifaceted medium that film undoubtedly is?

1. Distant reading and television studies

1.1 Quantitative approaches to television studies

Development of efficient image processing methods in recent years brought rising interest in application of quantitative methods to television and film works. As evidenced by papers presented during meetings of ADHO Special Interest Group Audio Visual Material in Digital Humanities, this concerns especially their visual layer. Hitherto-conducted studies range from cinematics (e.g. works of Mike Baxter and Yuri Tsivian) to applications of high-level computer vision methods to the study of moving images (e.g. the Distant Viewing TV project: Arnold and Tilton 2017) to whole systems dedicated to archiving and analyzing of videos (CLARIAH Media Suite). Unsurprisingly, the analysis of the audio layer is yet to be undertaken on a larger scale – state of the art solutions to problems such as speaker diarization are still far from useful, and, like computer vision, this type of analysis is greedy in terms of computational power. Comprehensive studies examining all multimodal aspects of films and TV series still belong to the future.

While the above-mentioned studies are of great importance, they align more with the term of “distant watching” or “distant viewing,” and in this paper I aim to draw attention to another aspect of audiovisual creations: to the textual layer of dialogue. Sarah Kozloff (2000) noted that film dialogue is often overlooked and considered of little importance and interest to both film scholars or linguists. She argued in favor of the usefulness of such studies, providing relevant examples and encouraging further ventures into linguistic aspects of audiovisual creations. Later years brought noteworthy projects on both film and television dialogue, most notably numerous works in corpus linguistics and discourse analysis, produced, among others, by Monika Bednarek, whose 2010 and 2018 books provide the most comprehensive review of the research that combines linguistics and television studies.

In the latter book, Bednarek noticed the rise of popularity of more formalized research on TV dialogue, including quantitative studies using corpora of transcribed dialogues, and pointed to the examples of TV series studied within various linguistic frameworks, including sociolinguistics (discussion in Richardson 2010), corpus linguistics (multiple examples for this and other linguistic branches in the extensive bibliography by Bednarek and Zago 2018), multimodal discourse analysis (Bateman and Schmidt 2012, Bateman and Wildfeuer 2014, and Wildfeuer 2014) and audiovisual translation (Hołobut et al. 2016, 2018a, 2018b, Bruti and Vignozzi 2016). TV series in their textual layer also become an object of research on narratives and themes, e.g. topic modeling-based plot explorations conducted by Benjamin M. Schmidt (2015) on a corpus of 80,000 film and television series scripts.

Significantly for this study, the research project conducted by Hołobut and her team is also one of the first to apply stylometry to the exploration of stylistic similarities in audiovisual data. In one of their first papers, Hołobut, Rybicki and Woźniak (2016) discovered that, in the case of television series, genre was likely to overshadow stylistic features specific to the author. In their further research, namely a stylometric analysis of *Pride and Prejudice* and its adaptations, Hołobut and Rybicki found that television series were “imitating Jane Austen’s verbal style more suggestively than the cinema productions” [Hołobut and Rybicki 2018, 8]. Another attempt at applying stylometric methods to television series dialogues was made by van Zyl and Botha (2016), whose study of *The Big Bang Theory* dialogue lists confirmed the existence of clear character idiolects in television discourse.

1.2 Authorship in television

Due to the multimodal nature of audiovisual works, in contrast to typical practice in literary and linguistic studies, the

concept of authorship is very difficult to attach to one person in television. In fact, the issue of authorship "has been rarely associated with television" at all, as authorship "was diffused among many individuals and may more properly have been claimed by a studio, company, or channel than by a person" [Hartley 2004, 401]. Collaborative storytelling and the number of people involved in the process of writing, rewriting and changes introduced during shooting and even in post-production, are often emphasized by both television creators and scholars.

While since the 1950s, film directors have been widely accepted as authors (or *auteurs*) of their works, and much attention is given to analyzing the works of such as directorial stars as Quentin Tarantino or David Lynch, the directors of television series have had a much smaller influence and typically only work on one up to a few episodes per season. Mittel also notes that while

the literary writer is ultimately celebrated for *authorship by origination*, assumed (however erroneously) to have written every word as an individual, a film's director is granted *authorship by responsibility*, ultimately making the necessary creative choices by supervising and guiding all of the film's many collaborators, even if much of the specific work was done by others. [Mittel 2017, 36]

This perspective on authorship by responsibility is also often linked to television, however, in relation not to directors but to showrunners (e.g. Steiner 2015, Jensen 2018, Lavik 2015). A showrunner is an informal term used to describe a person who holds most creative power in the show, usually combining the duties of main scriptwriter and executive producer. The concept is used mostly in relation to American culture (e.g. David Chase, Joss Whedon, Shonda Rhimes, David Simon), but there are also examples of showrunners in the UK (e.g. Russell T Davies, Charlie Brooker) and Scandinavia (Hans Rosenfeldt). Detailed origin and characteristics of this job have so far been subject to few studies, with the most comprehensive analyses by Brett Martin (2013) and Erland Lavik (2015). In his book, Martin looks into the phenomena associated with the popularity of series authored by some of the earliest openly-appointed showrunners and, based on conversations with them and their collaborators, he details existing forms of this mode of work. With all the information he collected, Martin does not hesitate to refer to TV showrunners (such as David Chase) as authors, comparing the conditions of their work to serial Victorian writers, such as Charles Dickens, Anthony Trollope or George Eliot. He also points out that given their decision-making over "story direction to casting to the color of seemingly insignificant characters' shirts" [Martin 2013, 8] they seem "all-knowing deit[ies]." Martin also observes that showrunners are typically writers (with few examples of showrunners holding only producer credits) and that they tend to emphasize their power over words and world-shaping. While typically there are many writers involved in the writing of a series, there are various ways their work can be organized: collaboration can include a "writers' room," where a showrunner meets with writers to pass/develop ideas together in detail before designating a team member to transform such detailed vision into actual script and again to discuss it in the room; in another manner, writers contribute episodes based on a general season narrative arch developed by a showrunner, who supervises execution and introduces corrections and changes to maintain consistency. In turn, in his book Lavik compares American, Dutch and Norwegian styles of showrunning and argues that outside of the American perspective detailed by Martin, showrunners in Europe give their writers much more authorial freedom and have themselves more creative than administrative responsibilities. Interestingly, Lavik also distinguishes between focused (one showrunner for the whole production of the show) or distributed (showrunners changing every (few) seasons) overarching authorship of showrunners, and actual authorship of writers.

The exact scope of influence of a showrunner is unclear and varies from series to series. A script is usually rewritten many times at various points before, during and after production, and, as noted by Jensen:

other writers work to support the vision founded by the showrunner, or creator. That, however, does not mean that these writers (...) do not make a difference in the final outcome of a series (they most certainly do), but they do so within a paradigm laid down by the creator(s) of the show. [Jensen 2017, 37]

Jensen also notes that despite this complex authorship of television productions, the works of particular showrunners

bear similarities across various series, both in style and in specific interests of the authors [Jensen 2017, 33]. He emphasizes interest in “seeing how the various works across an *oeuvre* depict a topic in different ways, maybe even in contradictory ways” [Jensen 2017, 34] – this is to be expected from an author able to create well-developed characters that differ in their ideas and nature as expressed in their dialogue lines, and can be used to evaluate character-building skills of the author. One method to do so may consist in contrastive analysis discussed later in this paper.

The concept of authorship of audiovisual works as adapted from literature poses even more problems, as discussed in detail by Steiner (2015) on the example of *Game of Thrones* and its “showrunner-auteur troika:” of author of the novels adapted into the series and series’ two showrunners, all participating in writing of the stories. Determining the scope of influence of any of them is a task posing a great challenge to quantitative and qualitative approaches alike. Again, I will argue that one of stylometric methods, sequential analysis, can facilitate this task, at least when it comes to the authorship of the textual layer.

2. Doctor Who – the background of the series

2.1 About the show

The British television series Doctor Who is a phenomenon for a number of reasons, and a perfect subject for a study applying a variety of quantitative methods. Broadcast since 1963 (with a break between 1989 and 2005), it is one of the most popular science fiction series, critically acclaimed, with generally a mainstream status (at least in Britain), and with a cult following.

The basic assumptions of the show are quite unique:

Its main character, the Doctor, travels with his companions in the TARDIS (Time and Relative Dimension in Space), a ship capable of traveling through space and time that takes the exterior form of a 1930s British police booth, but is bigger on the inside. (...) the Doctor is not consistently portrayed by the same actor; periodically the Doctor “dies” and regenerates in a new humanlike form with a new personality [Edwards 2014, 375].

This form certainly contributes to the popularity of the show. Frequent changes allow to keep it fresh and likely to attract both long-term fans as well as people interested in only following one specific regeneration of the Doctor. Rebuilding the show around new incarnation of the main character enables viewers to start watching the show from various episodes and seasons without normal difficulty related to following TV series from a random point, that is having to figure out previous plot twists relevant for its understanding.

2.1 Internal classification of the show

The show can be, and is, divided in numerous ways both by producers, fans, and scholars. The most basic division line is that between the “Classic” (1963-89) – and the “New,” sometimes spelled “Nu” (produced from 2005 onwards). The 1996 feature film is considered a part of the Classic Who.

Another division line is defined by the respective regenerations of the Doctor, thirteen as of 2018; this does not include the so-called Doctor 8.5, also known as the War Doctor, appearing in the 50th anniversary special. Since the Doctor is the main axis of the show and usually the one character who delivers most of the dialogue (in the New series, an average of 30-40% of all lines belong to the Doctor, i.e. twice as much as any of his companions), their^[3] change strongly influences the overall shape of the show.

The 2005 revival introduced many innovations, primarily in narrative style and in the introduction of the position of a showrunner. This has been discussed as both a factor ensuring the success of the refreshed series and a reason behind a major change in narrative style [Hills 2010] [Hills 2013]. As of 2018, the show has recently adopted its third showrunner, Chris Chibnall, with the first, Russell T Davies, serving in the years 2005-2010, and the second, Steven Moffat, between 2010 and 2017. Each of them had an established successful series writing career at the moment of taking over and was well-known by the British public. The showrunners were generally seen as bringing their individual

12

13

14

15

16

17

18

19

style and vision to the show, and the years of their work on Doctor Who were marked respectively as the Davies and the Moffat eras (see e.g. in Chapman 2014) – a quick search reveals that a “Chibnall era” is now also used in the Internet, though with as of yet small frequency.

This perspective of various strong influences offers an opportunity to examine not only the authorship of the episodes, consistency of the show and character stability, but also the relevance of particular divisions.

20

3. Empirical study

3.1 Problems

3.1.1 Obtaining the data

The classic problem of any quantitative study is obtaining the data, especially data of the right quality. The situation is no different if we decide to conduct an analysis of television dialogue. If anything, it is even more complex: we must first decide how we define the dialogue that is to be analyzed: the scripts or the spoken text as it was actually recorded? While some creators or broadcasters (e.g. BBC Writersroom) share scripts of their show after the series is released, such an approach is rare and usually concerns only a few sample episodes rather than a full series. Moreover, it is not clear how much such released scripts vary from the final televised versions of episodes; and since it is well known that dialogue is reworked continuously during production and post-production to meet the changing conditions of realizations, full compliance can never be assumed.

21

One tempting source of data are large amounts of subtitles available freely on the Internet. There are many arguments in favor of using subtitles: they are widely available and can be easily scraped, converted and cleaned without investing much effort and time. However, there are just as many arguments against using them. Subtitles found in popular aggregator websites, such as www.yifysubtitles.com or www.tvsubtitles.org, are usually of fairly good quality, but there is always a level of uncleanliness of the data. They are usually created by ripping original subtitles from DVD/BlueRay editions or through transcription conducted by amateur fans, both factors increasing the risk of typos and arbitrary alterations, or of missing some information.

22

It is also important to remember that subtitling is a highly conventionalized art. Professional standards (and distributors) place limitations as to the length of text displayed on the screen at the time. Common practices include rephrasing some original sentences, turning them into shorter equivalents when the pace of speech is too fast, adjusting punctuation to respond not to the actual manner of speaking but rather to the comfort of reading, or omitting content spoken in foreign languages or as part of the music soundtrack (e.g. song lyrics, even if sang by characters). Subtitles created by amateurs are also likely to present various spelling conventions requiring normalization.^[4]

23

Yet another approach proposes using transcriptions of the dialogue. When the transcription is performed per commission, this method allows for greater control over the text – including detailed information about who is speaking (and how), and agreeing on notation and spelling standards. Its main disadvantage is the cost in terms of time, obviously, but also money. Describing her SydTV Corpus, Bednarek explains her reasons for deciding on this solution: she was trading the size (choosing to have a 60-episode clean corpus of approximately 250k tokens rather than a larger, yet messier one) for quality – a factor crucial when conducting detailed analysis of language variation features [Bednarek 2018, 82–91].

24

In this study, I too built a corpus of transcriptions, but using fan transcripts selected from one of the many websites collecting them. Fan transcripts are never perfectly clean, however they have a great advantage over subtitles: they often offer information on who speaks the particular lines, and they are a transcription of actual spoken dialogue. In this particular case, the texts were transcribed by one person, with others only pointing out corrections via e-mails to the author, which definitely helped in maintaining consistency of notation. To ensure the quality of the dataset, apart from automatic preprocessing and cleaning I conducted sample manual verification of the quality by reading the texts and listening to New Who episodes, and by skimming through the rest of the texts in the corpus, correcting an almost negligible amount of typos.

25

3.1.2. Dataset

The dataset used for this study was prepared based on the web scraped fan transcripts^[5] and consisted of a total of 314 text files. The decision to use fan transcripts rather than subtitles was grounded in the corpus linguistic recommendations for research on TV series language [Bednarek 2018], as discussed in the previous section. While the corpus covers the whole show as of December 2018, the number of texts differs from the number of episodes and stories as listed e.g. on Wikipedia. Over the course of the development of the show, BBC offered various structures of a given season (or series, as this is commonly referred to in British television discourse). During Classic Who, each season was divided into “serials:” longer stories of continuous narrative, which were further divided into 25-minute episodes broadcast every week. Serials varied in length, going from two to twelve episodes. In contrast, New Who consists mostly of stand-alone 45-50-minute episodes, with occasional two- or three-part episodes, typical especially for season finales, and for the Davies era. I decided against splitting Classic “serials” into episodes and against concatenating New episodes into stories; instead, I applied sampling (drawing samples of equal length from the corpus) to even out the size of the examined texts.

Before the analysis, the data was cleaned of html tags and of any additional information which was not a part of the actual dialogue (comments in brackets, indications of speakers, etc.). Cleaning was done semi-automatically with a number of regular expressions matching observed patterns during the manual cleanup of several randomly selected episodes. The clean dataset was used to create three types of corpora: i) all dialogue lines, ii) lines spoken by the Doctor, iii) lines solely from the New Who seasons.

3.2 Methods

Stylometry is grounded in the observation that the distribution of certain features, most notably most frequent words, is a strong bearer of stylistic ‘signal’, which can be used to predict authorship, chronology, genre, or to lesser extent topical similarities. The earliest attempts to compare the use of words across texts date back to Lorenzo Valla in the 15th century and Augustus de Morgan and Wincenty Lutosławski in the 19th century, with the latter considered to be the first to compare top most frequent words, using this method to analyze works of Plato. The approach became more popular after Mosteller & Wallace’s famous study of *The Federalist Papers* (1964), and started to develop intensely with the availability of more powerful computers and Burrows’ study of Jane Austen’s works (1987), in which he introduced a more reliable method of comparing not just individual words but whole texts with his Delta measure (2002).

A stylometric study requires a few basic items: a corpus of texts which will be turned into a frequency table detailing use of all available words, a method for calculating a distance between their use across the texts (so called distance measure) and a classification algorithm. An optional, but convenient addition is a visualization method, allowing for easier interpretation of the final outcome. All analyses in this study were performed using stylo, an R package for computational text analysis [Eder et al. 2016], which includes algorithms for all these steps and offers a wide selection of types of distance measures and classification algorithms, not requiring more computational skills than preparing the corpus and making an informed choice between appropriate options.

One important step of preparing the experiment is choosing the right features to analyze. While Mosteller & Wallace opted for using most frequent words (MFWs); later studies experimented with the use of combinations (the so-called n-grams) of subsequent characters, subsequent words or even grammatical features. While character n-grams show nice performance for some types of texts, most notably historic and poetic texts [Kjell 1994] [Stamatatos 2009] [Kestemont 2014], n-grams of words typically perform worse [Eder 2011], as do grammatical features, e.g. POS tags [Górski et al. 2014] [Cafiero and Camps 2019], or lemmas. Given that the above studies seem to show that, in contemporary settings, most frequent words perform best, I decided to use them as features. Dismaying as such an approach may seem to the traditional humanities scholar, empirical evidence supporting it is overwhelming. The possibility of using such simple evidence for such large purposes rests upon the fact that words do not function as discrete entities. Since they gain their full meaning only through the different sorts of relationships they form with each other, they can be seen as markers of those relationships and, accordingly, of everything that those relationships entail [McKenna et al. 1999].

26

27

28

29

30

3.2.1 Conducted types of analysis

The first and mainly exploratory test was hierarchical clustering, one of the basic but most reliable algorithms of grouping similar texts, in its network variant of bootstrap consensus analysis. This variant is based on conducting a series of hierarchical clustering analyses with different counts of features and comparing their inter-agreement with a set threshold, so that only the texts classified as similar a set number of times are taken into account (e.g. for a threshold of 0.5, i.e. 50% and four iterations, two texts need to be classified as each other's nearest neighbors in at least two iterations for the connection to be considered valid). It presents a more objective and reliable view of the relations in the dataset than regular cluster analysis, decreasing the risk of cherry-picking by not relying on just one best outcome, and can be visualized in a graphical form of a network, which facilitates interpretation of the results in the case of large datasets, like the one in this study.

Network data produced by stylo was used in Gephi to create visualizations and to conduct further analyses, e.g. community detection with the Louvain modularity algorithm [Blondel et al. 2008]. Community detection is a method facilitating identification of naturally occurring groups within networks, often difficult to observe in detail with human eye, allowing for understanding their large-scale structures. The method divides a network into groups of nodes sharing strong connections between them. The Louvain algorithm is commonly used in many disciplines and was found to be the most reliably performing community detection in stylometric studies [Ochab et al. 2019] as it is well adjusted to the analysis of networks that may contain small structures, which can be expected also in the case of TV dialogue data.

In the contrastive analysis of Doctor regenerations I used the algorithm of Craig's Zeta [Burrows 2007] [Craig and Kinney 2009], which compares which features are used significantly and consequently more often across works in each of the two examined sets.

For a more detailed study of authorial influence of showrunner supervising scriptwriters I used Rolling Delta algorithm [Eder et al. 2016], a method that looks for "signals" of different authors within a single collaborative text. It conducts sequential classification of "slices" of such a text and allows for a fairly precise identification of authorial takeovers within one work (e.g. the case of Harper Lee in Choiński et al. 2019).

3.2.2. Identifying features to be used in the analysis

Dialogues in television series differ from both literary texts to unscripted spoken language. Awareness of this is crucial for proper performance and interpretation of a quantitative analysis. Functions of dialogue in television writing related to anchorage of the narrative, building continuity and helping viewers remember characters' names, locations etc. [Kozloff 2000, 33] make for a higher percentage of proper names both in the overall show and in particular episodes. As observed in the frequency tables of word usage in the corpus used in this study, at least in the case of Doctor Who, especially episodes introducing new characters tend to repeat names more often, and, as evidenced by the count of number of words uttered by companion characters in Doctor Who, introductory episodes are usually also the ones in which respective characters speak the most within the whole set of their appearances in the show.

This nature of TV language, combined with the thematic variety of episodes, calls for implementing some sort of limit to the influence of features that may score very high in individual texts (and this way also in the list of most frequent words) but be rather non-important for the general stylistic profile of the corpus. One such solution, commonly known as culling, has long been in use in stylometry, although as of yet there are no studies determining exact conditions of its use, thus enforcing special carefulness in its application. Culling is a procedure of removing such features from the list of analyzed features that only appear in a determined percentage of analyzed texts.

In this study I chose to first conduct a leave-one-out cross-validation classification test on the whole corpus (classifying each text using the rest of the dataset as the training model) and discovered that culling of 20% increased the performance of my chosen classifier (Support Vector Machine) performance for all examined numbers of MFWs (100-500). While such procedure might be useful, factors to be taken into consideration before adopting it to other studies are size of the corpus, thematic variation of the texts, and their length. Finally, one should remember that the use of culling heavily decreases the number of features that can be used in the study (e.g. in case of texts composed solely of lines

31

32

33

34

35

36

37

spoken by the Doctor, only 395 shared words remain).

As for other settings, I explored possibilities of using various distance measures and found Cosine Delta [Evert et al. 2017], which was proven to be the most reliable in a number of recent studies, to produce most stable results across settings, with Burrows's Classic Delta also providing fairly good results.

38

4 Results

4.1 Development of the show

The first experiment was set to examine the direction of the development of the show and see whether a division into Classic and New Who, or into the showrunner eras, result in the change of the stylometric "signals." For this test I used all the texts in the corpora and considered two scenarios: first, examining the whole texts; second, examining just the extracted lines of the Doctor part. For each of the scenarios I conducted two analyses: with full text, and with sampling (random sampling of four 1,000-word samples per text). The second solution was adopted as an additional safety measure due to my concerns about the uneven length of the texts: while New Who usually has 4,500-6,000 words per episode, some Classic Who stories go up to 20,000 words.

39

4.1.1 Episodes

Whole episode perspective:

The primary finding, confirming intuitive expectations, was a strong division between Classic and New Who series, and more: it is also clear at first glance that New Who also bears a strong division line, as the texts from the three respective eras form three distinct groups. As visible in Figure 1, not only does the layout of the nodes confirm the separation, it is also detected by the Louvain algorithm, with the communities it discovered marked in different colors. Interestingly, some of the episodes from Eleventh and Twelfth Doctor tenures (New Who, Moffat as showrunner) are closely connected to groups of texts from Classic Who, which can be explained by Moffat's commonly expressed intention to draw more parallels to the Classic series. Whereas Davies was hired to reinvent the show for the 21st century and to make it attractive for contemporary viewers, Moffat seemed to have been given more artistic freedom when it comes to exploring the roots of the show, as evidenced by his bringing back classic opponents such as Ice Warriors, Mondasian Cybermen or Zygons, as well as more frequent use of longer storylines developed over two episodes and stories overlapping whole seasons.

40

The newest additions, ten episodes of the latest (2018) season created by a new showrunner, Chris Chibnall, position themselves halfway between two other New Who showrunners, and application of the Louvain algorithm reveals them to be slightly more similar to the series supervised by Moffat. This result may be somewhat surprising, given that Chibnall was one of the writers employed at the time of Davis's tenure and also collaborated with him on the *Doctor Who* spin-off, *Torchwood*. However, as the amount of the data for the Chibnall era is still small, these particular results should not be taken as a definite marker of his preference towards the style of his colleague.

41

The network does not seem to show clear chronological progression. While the texts from particular Doctor regenerations exhibit strong internal similarity (visualized by thick lines in the figure), they are not determined to connect in the order of subsequent regenerations. While, as visualized in Figure 2, there is a clear split into three (and with increasing the resolution of Louvain algorithm, forcing greater generalization, only two, see Figure 3) New Who eras, the Classic series lack similar structure: the discovered communities are not tied to a particular regeneration or writer, but are more genre/topic-oriented.

42

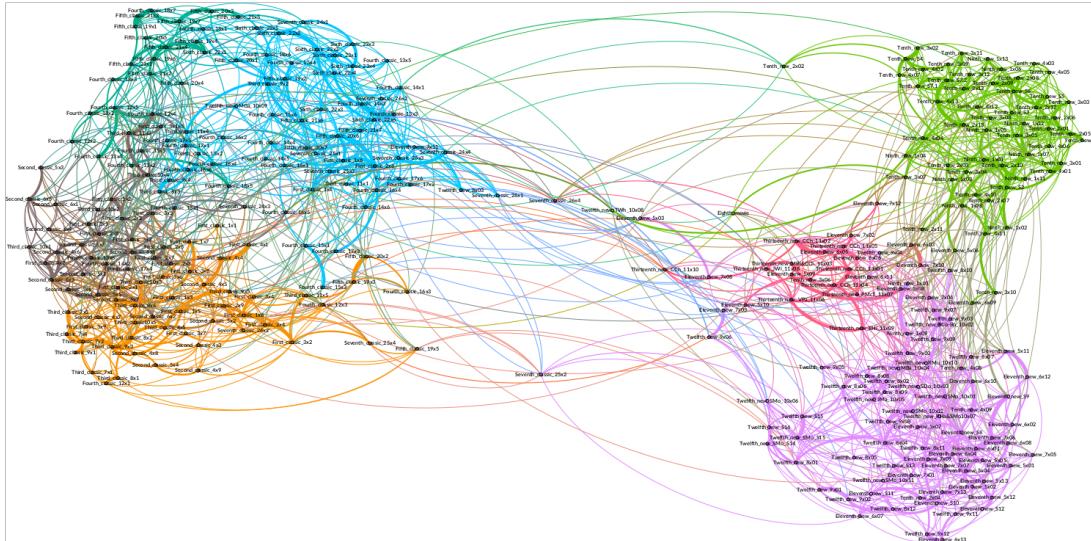


Figure 1. Network analysis (bootstrap consensus tree with Cosine Delta 100-500 MFWs on all episodes). Colors indicate communities discovered by the Louvain algorithm with modularity resolution of 1. On the left, Classic Who series include four distinct communities, corresponding to genre and topics rather than the particular regenerations or writers. On the right, three communities, mostly grouping the texts from each of the three showrunners.

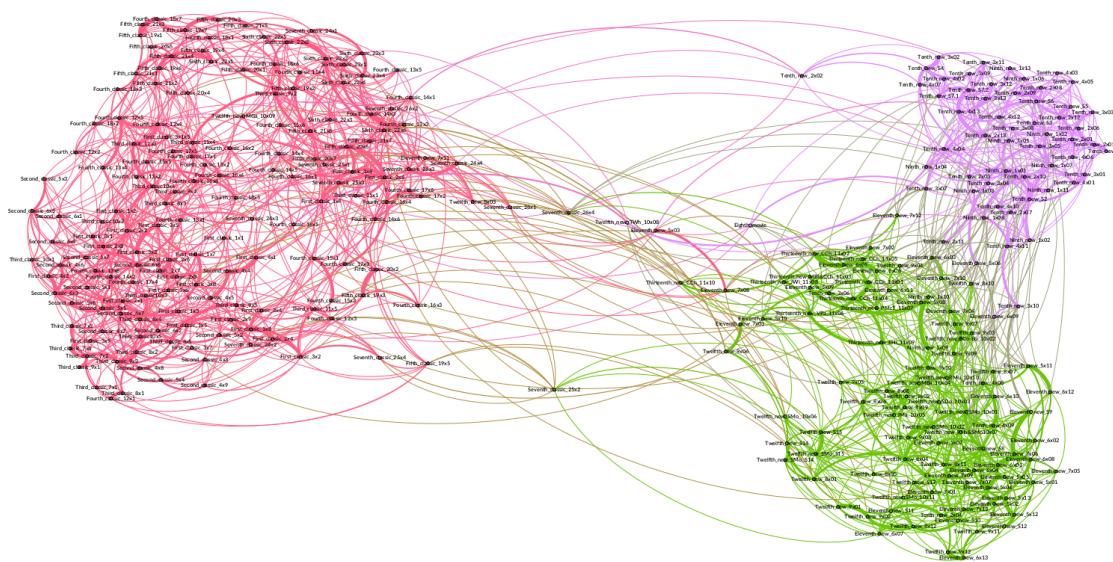


Figure 2. Network analysis (bootstrap consensus tree with Cosine Delta 100-500 MFWs on all episodes). Colors indicate communities discovered by the Louvain algorithm with modularity resolution of 3. On the left, with this level of generalization, Classic Who no longer has distinct inner groups; on the right, episodes from Chibnall era are now clustered with a larger Moffat community, Davies remains distinct.

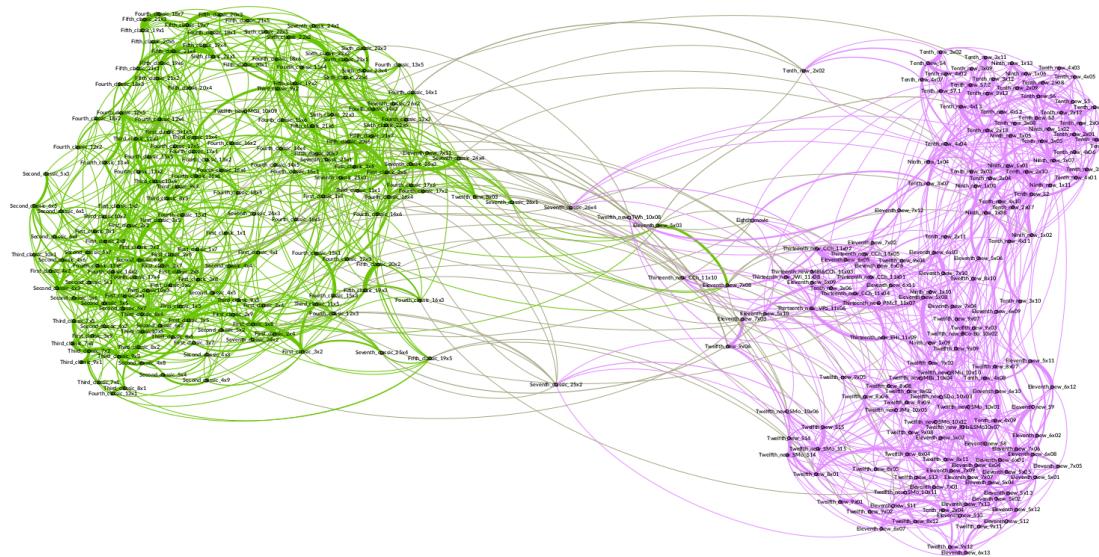


Figure 3. All episodes; colors indicate communities found by the Louvain algorithm with modularity resolution of 4. This shows the highest level of generalization, with Classic Who on the left, and New Who series on the right.

Sampled perspective:

To verify that the obtained results were not skewed by varying lengths of the texts, I conducted the same analysis using randomly selected samples of the texts. I automatically selected four random samples of 1,000-word length from each of the files in the corpus without replacement (ensuring that various samples would not overlap in text). In this particular case I also had to decrease the number of examined features to 300 in the last iteration, as this was the rough number of words common for all the texts in this setting. Applying the Louvain algorithm in this case showed that resolution had to be increased to much higher values (e.g. 15 on Figure 4) to reveal a smaller number of communities and a more generalized view. While New Who episodes form strongly interconnected clusters of individual episodes but have weak intra-connection edges, the Classic Who forms a uniform massive community showing strong relations between the texts; still, many episodes are mixed and do not recognize other samples from the same text.

43

Interestingly, some episodes are outliers from their own community. The New Who 11x06 episode, for instance, a debut by Vinay Patel describing the events around the Partition of India, appears among Classic Who episodes, forming significant relations with the Seventh Doctor's adventures that are seemingly unsimilar in language and topic. In turn, some episodes of the Seventh, and to a lesser extent of the Fourth and Sixth Doctors have stronger affinities with the New Who.

44

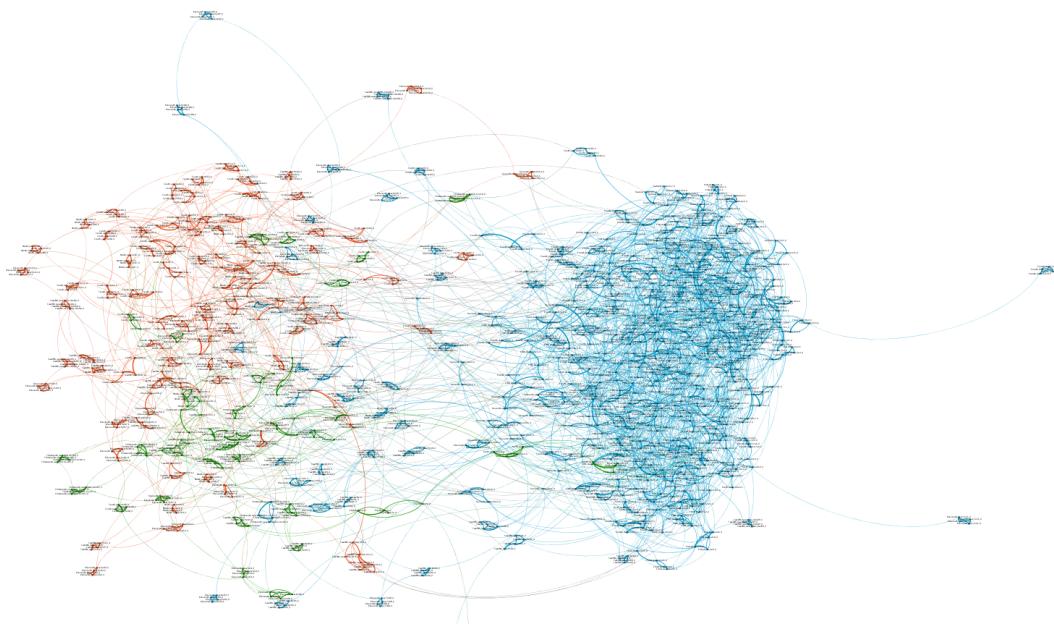


Figure 4. Network visualization of sampled episodes, 4 samples of 1,000 words per episode. Colors indicate communities discovered by the Louvain algorithm, with modularity resolution of 15. On the right, Classic series form a strong community with thick edges indicating strong relations. On the left, New Who series form small strongly connected groups of samples relating to a given episode, and while the Louvain algorithm distinguishes two groups, they do not indicate a particular showrunner era or a new regeneration of the Doctor.

4.1.2 Doctor

Even though the first set of tests already confirmed two basic divisions of the show into Classic and New Who, and the Davies and Moffat/Chibnall eras within the latter, no distinction was found regarding the influence of the Doctor regeneration on the development of episodes. Therefore, since their lines on average comprise 30 to 40% percent of the whole dialogue, I decided to look at them without the noise of lines spoken by other characters, seeing also if the writers succeeded – or, maybe, simply happened – to differentiate between regenerations. 45

The results here are much more interesting. As presented in Figures 5-7, Classic series show a clear distinction between Doctors and a trace of the chronological signal: while the First Doctor is equally strongly connected to his direct replacement and to later regenerations of the Fifth and Sixth Doctors, the communities of the Second, Third and Fourth Doctors are both distinct, and the fact that they have the strongest ties with their direct predecessors and successors indicate a visible chronology. Strikingly, the Fifth, Sixth and Seventh Doctors are all heavily intertwined together, which may coincide with the decline in the popularity and perceived quality of the show during their tenures; at this time, the series began to slowly but gradually lose its appeal to viewers, resulting in lower broadcast numbers and higher criticism of show development, leading to indefinite suspension in 1989. 46

Interestingly, while an earlier version of this study not including the last two series of Doctor Who alienated the First Doctor as the most distinct regeneration in the course of the history of the show (previous unpublished study of the author), the corpus used in this study distinguishes Second and Third (together) and Fourth (on his own) Doctors as most unlike the other Classic Who Doctors (see them form separate strong communities in the left bottom corner of Figure 6). 47

As for New Who, particular regenerations do not seem to have their own voice heard so much: the influence of the showrunners overshadows them, and this is visible even with the lowest resolution of Louvain algorithm in Figure 5 (pink for Moffat and green for Davies). One exception is the Thirteenth Doctor, who seems to mostly remain detached from other New Who Doctors over the course of increasing Modularity resolution. This could be the consequence of the new showrunner, except that Chris Chibnall wrote a number of episodes for both Davies and Moffat, the texts of which cluster with their respective eras, and the whole texts of the episodes from his era showed his similarity to Moffat's era 48

(Figures 1-3); another hypothesis might involve the gender switch.

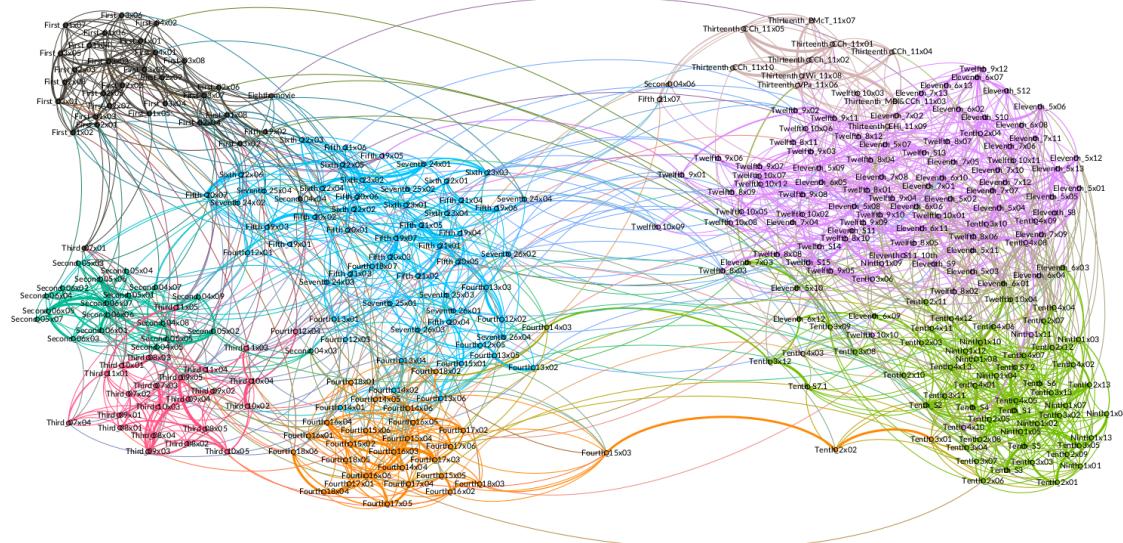


Figure 5. Doctor lines, bootstrap consensus network, Cosine Delta, 100-500 MFW. Colors indicate communities detected by the Louvain algorithm with modularity resolution of 1. On the left, the large community of Classic Who with five internal groups of First, Second, Third and Fourth Doctors separately and Fifth, Sixth, Seventh and Eighth together (blue). On the right, the community of New Who, divided further into three groups corresponding to showrunners.

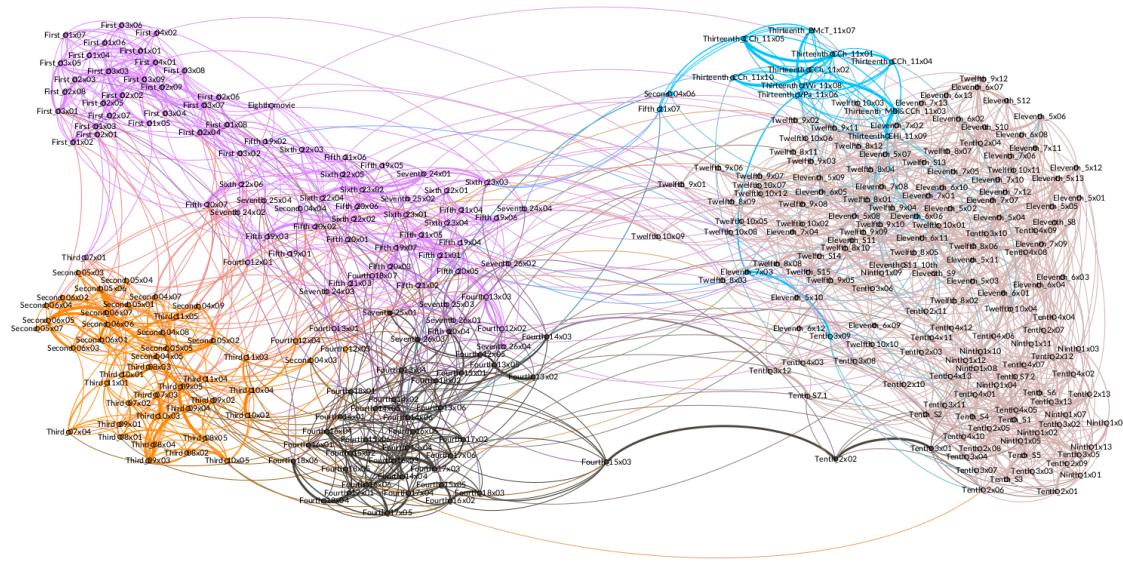


Figure 6. Doctor lines, bootstrap consensus network, Cosine Delta, 100-500 MFW. Colors indicate communities detected by the Louvain algorithm with modularity resolution of 3. Within the Classic Who community the First Doctor is now found to be more similar to the last four Doctors, and the Second and Third are clustered together – only the Fourth Doctor has his own group. Within the New Who community, all male Doctors, who also share being written in Davies's or Moffat's eras, are considered as one group, the Thirteenth Doctor forms her own group.

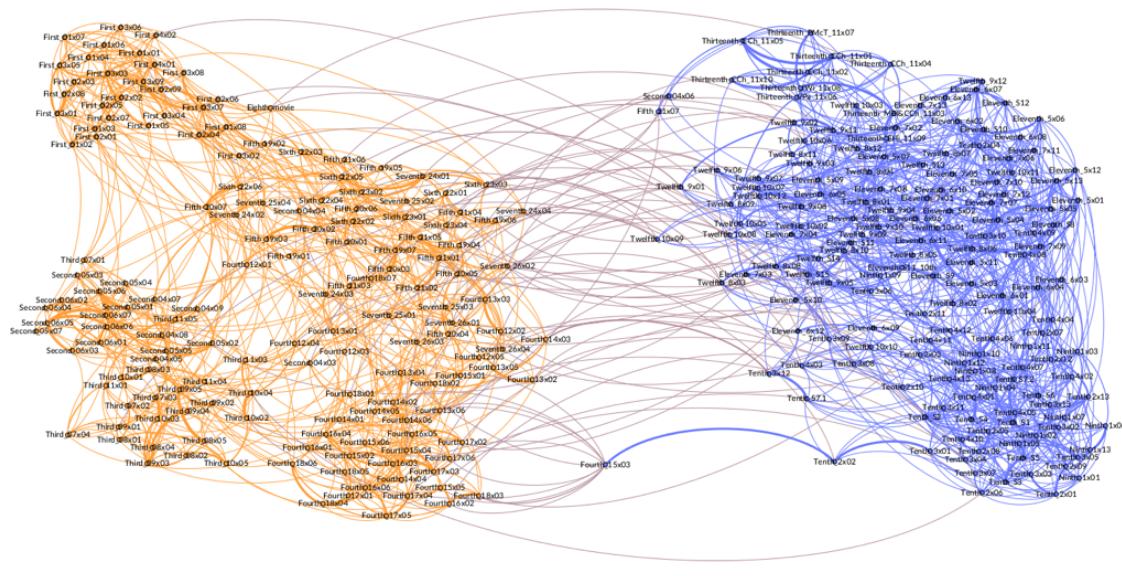


Figure 7. Doctor lines, bootstrap consensus network, Cosine Delta, 100-500 MFW. Colors indicate communities detected by the Louvain algorithm with modularity resolution of 4. The division line now falls between the Classic and the New series.

Overall, the break into Classic and New Who is the most important division, as revealed by the Louvain algorithm. However, while Finer and Pearlman state that “story is character” [Finer and Pearlman 2004, 41], this does not seem to hold for Doctor Who: the development of the main character is slightly different than that of the whole show.

4.2 Authorial influence

For this part of study I focused on the New Who exclusively. I started with a network analysis including only the whole texts from this time period, which was enough to form a number of observations. The first basic one is the high visibility of contributions of Steven Moffat. Most of his works cluster together, be it the ones written under supervision of Davies or during his own era, with a distinction of a few texts that cluster into other communities with a few other writers he supervised, which could be hypothesized as his influence over them (see Figure 8).

The situation seems quite different for the other two showrunners. Episodes written by Russell T Davies appear in all communities related to his tenure, mixing with other writers and showing strong narrative connections (in Figure 8, the group marked with light green connects mostly more dramatic two-part episodes and finales, whereas the brown one gathers more regular episodes from the middles of the seasons). Preliminary tests conducted for this study, not including the last two seasons, showed this even more sharply, with separate clusters for episodes opening and closing seasons: his writing and supervision seem to strictly control the development of the show, and while some influence of topics (e.g. episodes with traveling back in time) can be seen, the authorial control of a showrunner over writers, exhibited by strong ties between them, seems to be the main driving power behind the similarities shared by the texts.

Interestingly, as presented in Figure 9, the season created by Chris Chibnall clusters not only with his previous works on the show in the eras of his predecessors but also with the episodes written by “usual” scriptwriters earlier on, mostly from the Moffat era. This can be interpreted in a number of ways. Undoubtedly, Chibnall has a strong authorial presence, but the fact that other writers who largely contributed to the show and also built their own prominent careers such as Toby Whithouse (creator of *Being Human*) and Matthew Graham (co-creator of *Life on Mars*) cluster together with him could be interpreted as the community of writers who grew up in the Doctor Who writing room now finding their own voice (albeit somewhat similar when contrasted to Davies and Moffat).

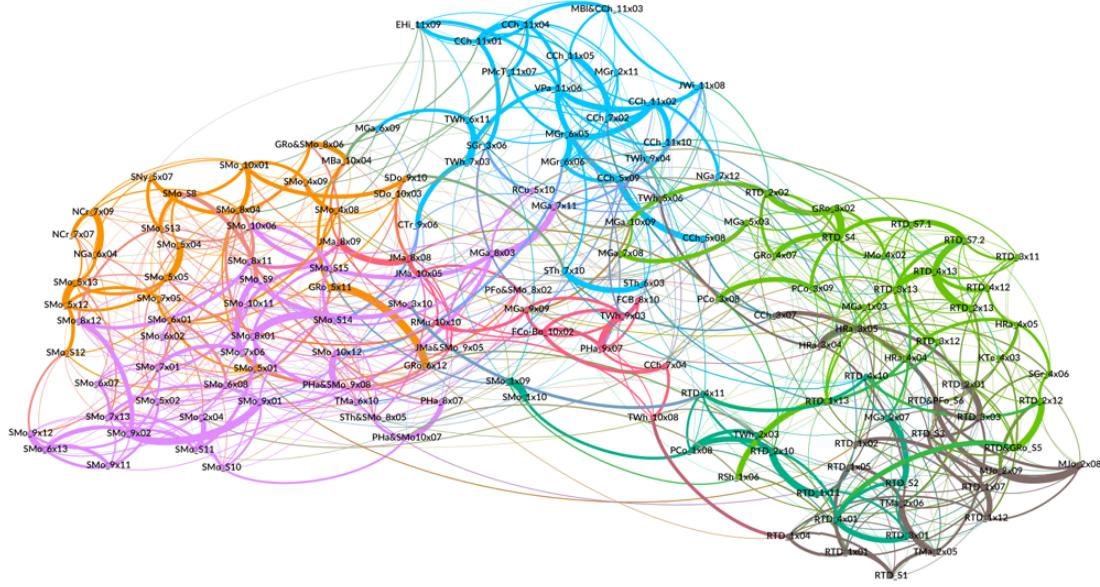


Figure 8. Network of New Who episodes, bootstrap consensus tree, Cosine Delta, 100-500 MFW. Colors indicate communities discovered by the Louvain algorithm with modularity resolution of 1. On the left and in the lower central part visible are three communities within the Moffat era. In the upper central part we find the Chibnall era, on the right, three communities form the Davies era.

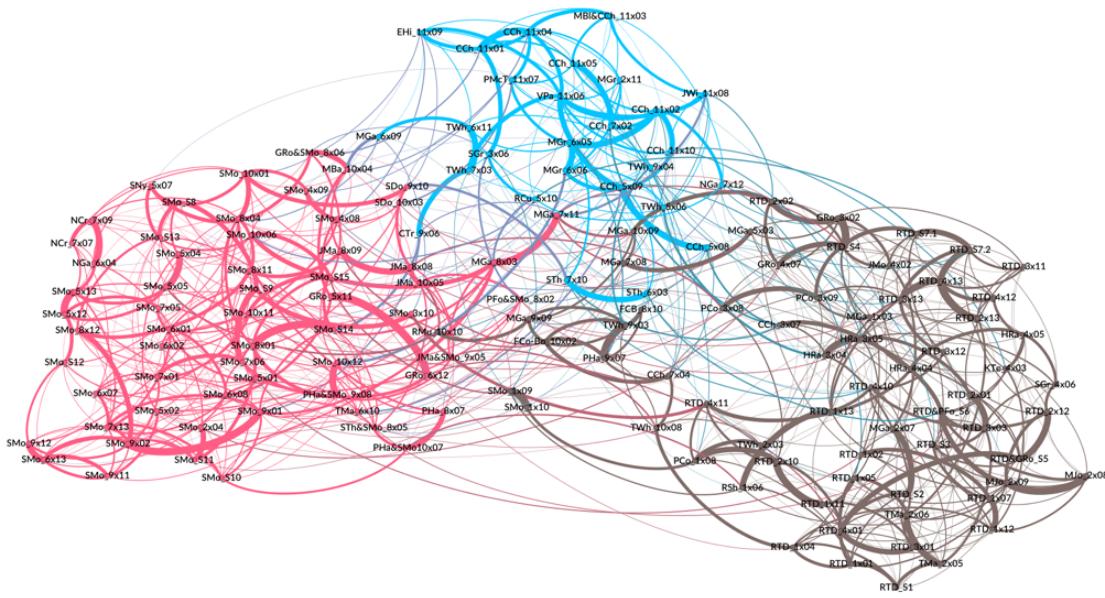


Figure 9. Network of New Who episodes, bootstrap consensus tree, Cosine Delta, 100-500 MFW. Colors indicate communities discovered by the Louvain algorithm, with modularity resolution of 3. Visible are three communities of respectively the Moffat (red) and Davies (brown) eras, and the Chibnall era (blue) extended with some of his previous works written earlier in the history of the show and episodes written by a few other prominent writers.

Detailed study of authorial influence:

For a more detailed study of authorial influence, I examined selected episodes of the show to compare Davies and Moffat in their roles as showrunners with Rolling Classify [Eder 2017], a method for sequential detection of stylistic idiosyncrasies within a single text. When choosing the episodes I followed some basic criteria: the credited author of the episode had to contribute at least a few other texts for the show, and to be considered as fairly independent according to available data on the working processes.

One methodological concern for this type of study is that the exact scope of the influence of the showrunner is unknown.

Using episodes from the show as training data carries the risk that all episodes written by a given person will be tainted by heavy showrunner impact, thus skewing the results in his favor. Doctor Who production does not report using Writers' Room in the American sense, although a version of it, based on meeting to read and discuss ready scripts was mentioned to exist at least in Moffat's time, as detailed in a special interview on their writing process.^[6] Instead, the show employs guest writers who create their own stories fitting some basic arch vision provided by the showrunner who also has the final say over the acceptance of their ideas and shape of the scripts. In fact, Davies claims that he writes "the final draft of almost all scripts – except Steven Moffat's, Matthew Graham's, Chris Chibnall's and Stephen Greenhorn's – and that draft becomes the Shooting Script" [Davies 2008, 180]. Despite these concerns, I decided to examine the situation, even if to offer doubtful findings to be further studied and verified. I present just a few cases below, with all available at the project website.

While in my study I also used Delta and Nearest Shrunken Centroid classifiers, here I present results obtained with Support Vector Machine (SVM) classifier, which is largely considered the most reliable classifier in stylometry [Stamatatos 2013], and which also in this case offered most stable results across various settings and for various datasets, as did using 100 MFW as features (although wider ranges of features were also tested). As Rolling Classify uses a sliding window to provide sequential classification of an examined text, the texts were divided into 1,000 word samples, with 750 words as overlap. The visualization graph produced by Rolling Classify shows two bars of authorial influence, reflecting the percentage of probable authorship of a given slice, with the lower bar showing the more likely impact.

55

No matter the method of classification, the outcomes revealed the major influence of Davies authorial style in all examined works of his era. However, whereas (e.g. Figure 11) Davies's influence dominates only a part of the episode written by Moffat, in case of the episodes Davies claimed in his memoir to have largely re-written (Figures 10 and 12) his style dominates the whole text, with the Rolling Delta algorithm recognizing him as their primary author. Assuming the honesty of the showrunner, this in itself is an important finding, showing that at least in the case of the texts clearly written by one scriptwriter, both his own authorial signal and that of the showrunner are easily detectable. The results for writers known to have been edited by the showrunner are less easily interpretable and reliable: while thanks to Davies's memoir we know he corrected them in this, fourth, season, a question remains whether the texts written by regular scriptwriters for earlier seasons were just as heavily corrected. Since I used them as training data to teach the model their style, positive answer to this question would mean that in the examined season Davies's impact was stronger than the combination of writers' style and his editorial impact as evidenced in earlier seasons. In the event that the earlier texts present pure style of particular authors, results such as obtained would mean that in this season showrunner's stylistic impact completely overshadows most of the credited authors. The case of Moffat's episode, unedited by Davies but still showing his stylistic impact is particularly interesting, giving grounds to raise the question how much of showrunner's stylistic fingerprint can be carried by others through imitation or following set tone of the show? While we know imitation is still detectable in literature despite best efforts (e.g. a case of translation by two people in Rybicki and Heydel 2013), perhaps the more controlled medium of TV series, or some of its features, like creating character idiolects, allows for easier speaking in someone else's voice?

56

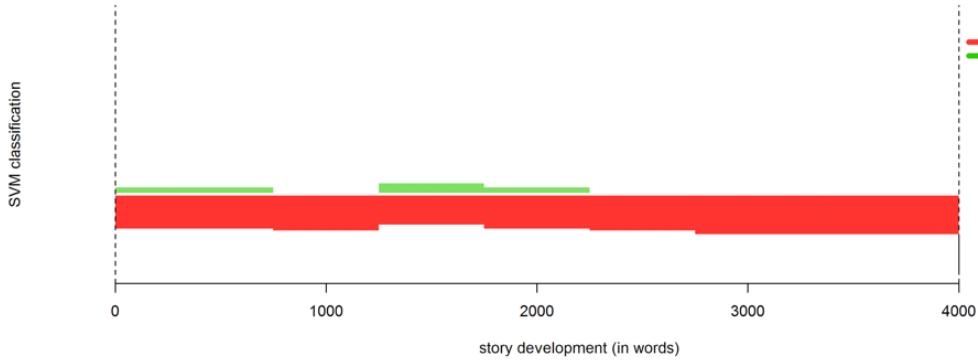


Figure 10. *The Doctor's Daughter (4x06)* contrasted sequentially against R.T Davies (red) and M. Gatiss (green); 100 most frequent words.

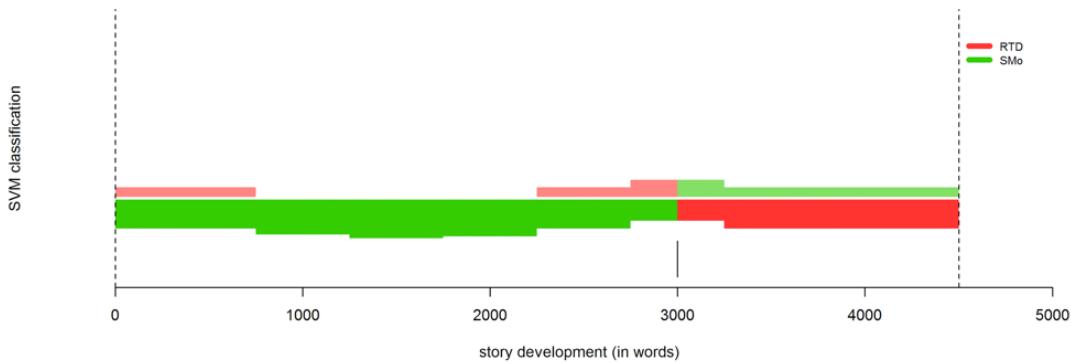


Figure 11. *Silence in the Library (4x08)* contrasted sequentially against R.T Davies (green) and S. Moffat (red); 100 most frequent words.

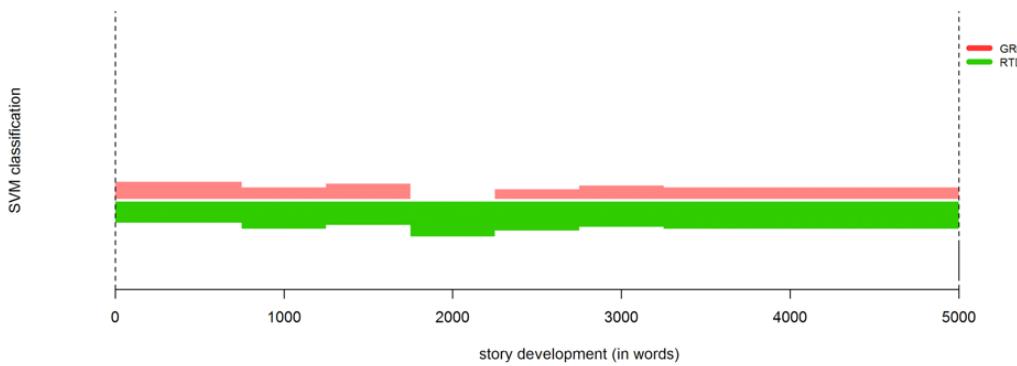


Figure 12. *The Unicorn and the Wasp (4x07)* contrasted sequentially against R.T Davies (green) and G. Roberts (red); 100 most frequent words.

Insofar as Moffat has published no memoir describing his writing and showrunning process and does not discuss the issue at length, with but occasional mentions that he just “sets the tone” of the show, I decided to examine episodes which co-credited him as an author, as well as some randomly selected works from his era. As visible in Figures 13-14,

his impact as a showrunner was determined to be less visible, although still significant. However, his impact when sharing authorial credits (Figure 15) seems to be quite dominating, even if still far from the influence executed by Davies.

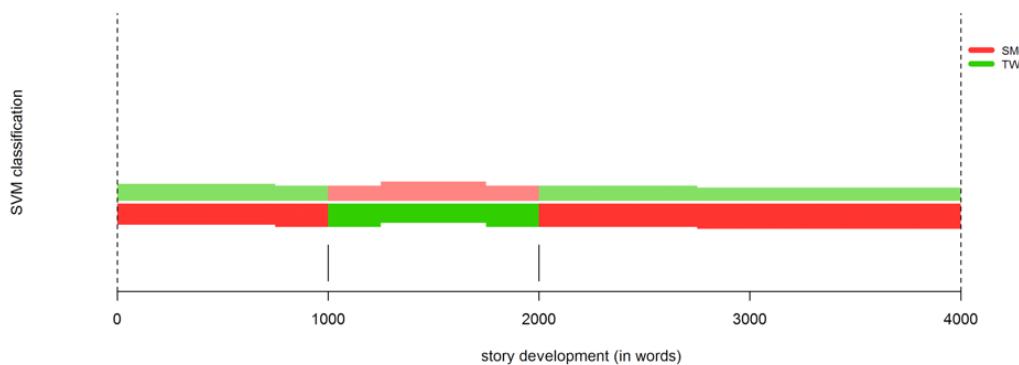


Figure 13. The God Complex (6x11) contrasted sequentially against T. Whithouse (green) and S. Moffat (red); 100 most frequent words.

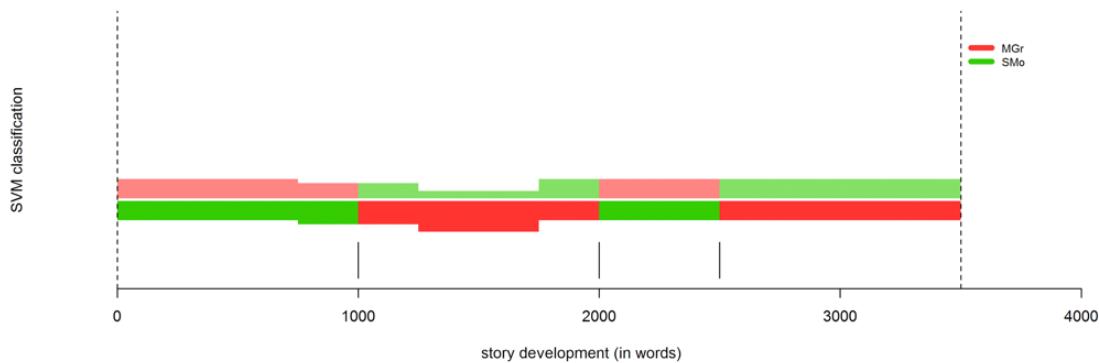


Figure 14. The Rebel Flesh (6x05) contrasted sequentially against S. Moffat (green) and M. Graham (red); 100 most frequent words.

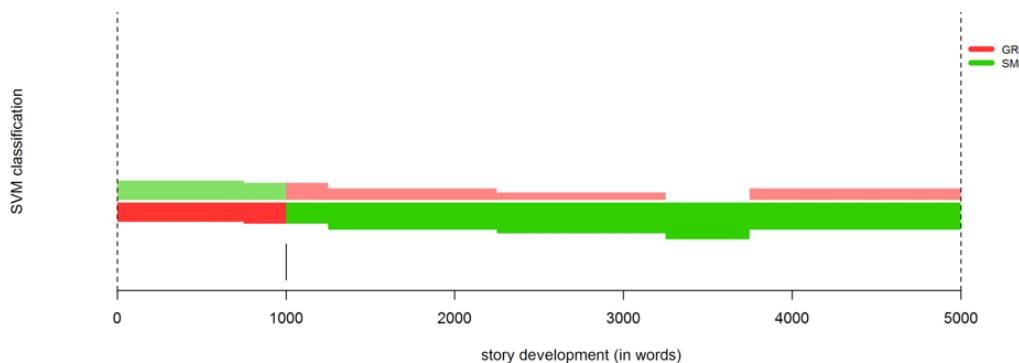


Figure 15. The Caretaker (8x06) contrasted sequentially against S. Moffat (green) and G. Roberts (red); 100 most frequent words. This episode credits them both as authors.

4.3 Character features

The analysis of sole lines of the Doctor revealed a level of distinction between the Thirteenth and other New Who Doctors. The easiest hypothesis would be the impact of gender – the new female Doctor versus all men – or distinct impact of the showrunner. However, at least the Davies era episodes written by Chibnall show the latter's stylistic "submission" rather than his own vision of the Doctor.

58

To verify this, I conducted contrastive analysis with Craig's Zeta [Craig and Kinney 2009] to compare the Thirteenth Doctor against her four direct predecessors, considered together and separately. Additionally, I also ran another test comparing Doctors created by Davies and by Moffat. Craig's Zeta allows for comparing two groups of texts to find words used consistently throughout texts in one of the groups, and consistently avoided by the other, and vice versa. Obviously, the most frequent words disappear, while content medium-frequency words come to the surface; this is also the reason why this version of quantitative analysis becomes (at least slightly) more similar to the traditional/qualitative approach.

59

As contrastive analysis was performed without culling, meaning that all words were included as they were, one immediate feature of lists of words characteristic to the particular regenerations is the high position of the proper names relating to companions – which is caused by frequent addressing of the given character or talking about them. This is true in most cases, but reveals various relations between the Doctor and their sidekicks: the names of all companions of the Ninth and Thirteenth Doctor are at the top of the list, but other Doctors show their preference towards a particular companion, e.g. Tenth to Rose over Martha and Donna, Eleventh to Amy over Rory and Clara.

60

On the first look, it would be easy to interpret the results as presenting a gendered portrayal of characters – the new female Doctor says "love," "woman," "together," and "please" more often than other Doctors. However, a deeper analysis of preferred words of her and other Doctors shows that it is not that simple. "Love" is more characteristic for the Eleventh Doctor, known for his expressive and emotional attitude to the world, and while "please" is high on the Thirteenth's most used list, its only seems high in comparison to Ninth and Twelfth Doctors, arguably more stereotypically male and harsher in behavior than the other two male Doctors. "Please" is also a word preferred by Moffat Doctors over Davies Doctors.

61

One characteristic feature of the language of the Thirteenth Doctor is a high ratio of elliptic adverbs such as "obviously," "presumably," "definitely," "possibly," as well as the adjective "sure." Also, regenerations written by Moffat use them more often than Davies, which may influence the pattern of relative frequencies in the corpus enough to partly explain the fact that, in the network studies presented above, Louvain Modularity recognizes Chibnall as closer to Moffat than to Davies. When analyzed in context, this use of adverbs reveals an interesting conclusion for the first female Doctor: compared to her predecessors who were largely explaining the world to their companions, the Thirteenth Doctor is more likely to engage in a discussion that may lead to her questioning hypotheses of her team or confirming their intuition and reasoning.

62

This is where knowledge of the audiovisual material and the whole corpus is helpful in the interpretation of the results: of the five New Who Doctors who struggle with posttraumatic stress disorder [Gibbs 2013] caused by their regeneration, loss of home planet and subsequent companions, the Thirteenth Doctor is the one who healed best and looks forward to a major change in her attitude. In the past, through travels with various companions, the Doctor often relied on their help, but at the end of the day the higher (intellectual) position of the Doctor was always clear. The last season shows the Doctor who chooses to engage into a team relationship rather than create a strong bond with an individual or duo of characters, and to give them more space for their own agenda. She takes on three companions (and only one of them is a young woman, which was a given for her predecessors), and while at various points of her introductory season some of them seem to be more important than others, overall they have about the same number of lines per episode and (which was not measured quantitatively) seem to have about the same amount of screen time. Altogether the companions speak about 30% of all lines per episode, almost matching the part of the Doctor (~35%), and this is the highest ratio of participation of the companions in the New Who, the most talkative of whom rarely surpassed half of the lines of the Doctor during the Davies era, and 25% under Moffat.

63

The keywords obtained in the analysis also show the Thirteenth Doctor to be quite tech-savvy, even compared to the Tenth Doctor who was famous for his technobabble such as "This is my timey-wimey detector. It goes ding when there's stuff" ("Blink", 3x10). She has "readings," "signals," etc. Finally, her list includes "kill," "killing," and "weapons," and while "fighting" and "killing" are characteristic for Davies when contrasted with Moffat, the Thirteenth Doctor is the one whom these words distinguish the most. All Doctors are pacifist and opposing violence, and the show generally avoids showing death of people, even though the Doctor has to die to regenerate. Already the Moffat tenure showed some relaxation of this rule (there is a recurring fan joke saying he was never able to permanently kill off a character), frequently using death by sacrifice as a plot development tool, only to later cancel it by introduction of a workaround. The Thirteenth Doctor deals with death much more often – it concerns both major characters and random people, and it is not justified by sacrifice; in fact, it may seem almost accidental, as when a Tzim-Sha warrior kills an elderly security guard peacefully doing his watch in 11x01. The most recent Doctor and her "gang" fight with cruel creatures and "people killing" on a regular basis.

Combined, observed differences between the most recent female and the earlier male regenerations seem to be related more to the change of the mode from "lone genius" to "team leader" rather than building the character according to traditional gender roles. Considering social-emotional keywords, Thirteenth Doctor seems to be most similar to Tenth and Eleventh. Interestingly, contrastive analysis of Davies and Moffat Doctors also revealed that the former has a preference towards social sphere of life (words such as "mother," "family," but also "alone," "human"), while the latter is distinguished by more verbs (e.g. "feel," "shut," "forget," "talking"), and expressive interjections.

5. Conclusions

This study allowed for automatic detection of creative forces that have a driving influence on the development of the Doctor Who TV show. It shows the basic distinction between Classic and New series and various modes of operation within them. The results of word frequency analyses show that the change the show underwent at the time of its reinvention in 2005 brought a major shift in its textual structure: while the Classic series seems to be more topic- or regeneration-oriented, the New series show a strong authorial signal of the showrunners, with the two longest-running ones being the most distinct.

The development of the show proceeds in a slightly different way when considering all dialogue lines rather than exclusively the lines of the protagonist. In the latter case, the Classic series exhibits well-defined communities for the first four regenerations, with this pattern breaking only for the last three Doctors, coinciding with a slow decline in the popularity of the show. In case of New Who, only Thirteenth Doctor clusters slightly apart, and the remaining four regenerations form two communities responding to showrunners creating respective regenerations. A closer analysis of the words distinguishing this first female Doctor from her male predecessors revealed that she is more of a team player who uses a higher ratio of words indicating dialogue involving exchange of opinions and cooperation. This strongly suggests that such a distinction has less to do with traditional gender perception of women and more with a more democratic attitude towards her companions.

While so far stylometry has rarely been used for more detailed studies of television discourse, the conducted contrastive analysis with Craig's Zeta produced results allowing for similar analysis as those obtained with Wordsmith by other scholars such as Bednarek, likely because they both also rely on algorithms of keyword detection. Furthermore, the obtained insight into authorial structures suggests that, at least in the case of some series, the television language may not be as heavily conventionalized as it is generally thought to be, and may bear quite significant stylistic markers of particular writers, which makes for an important point to consider while building a corpus for the purpose of examining general linguistic features of television language.

It may seem that a quantitative analysis resulting mostly in confirmation of observations made in qualitative analysis is pointless. However, reaching the same conclusions with different means supports the usefulness and validity of such methods, making an argument for their employment to problems exceeding possibilities of close reading, such as studies of much larger series of whole television genres, and granting that observations deducted via somewhat more intuition-driven and human perception-based methods stand. The latter is especially important for studies related to

popular culture, in the light of Tore Rye Andersen's (2012) hypothesis (as adapted to television discourse by Jensen 2017) that scholars relying solely on qualitative analysis are at greater risk of being unwillingly influenced by extensive authorial paratexts and marketing materials accompanying highly promoted works.

The use of stylometry for study of television dialogue requires carefulness and will certainly not be an attractive method in all kinds of studies. It should also be considered that, as explained, Doctor Who is fairly unique in many ways, which means that using similar methods with other data, and perhaps paying attention to other factors while doing so, will be needed to provide more final opinions. However, given that stylometry proved useful in various types of the analysis of Doctor Who, I dare express the hope that it can offer valid insight into stylistic relations inter and intra other works, and that hopefully with the further development of computer vision and speaker diarization methods it will soon be possible to compare the similarities between works on textual, visual and audio levels.

70

Acknowledgements

I would like to thank the editorial team and the reviewers for their valuable feedback. My gratitude also goes to Jan Rybicki for his inspiring this study a few years back, and to him and Maciej Eder for consulting the project and the text of this paper. The study was conducted as part of the Large-Scale Text Analysis and Methodological Foundations of Computational Stylistics project (SONATA-BIS 2017/26/E/HS2/01019) funded by the Polish National Science Center (NCN), for whose support I am very grateful.

71

Notes

[1] Supplementary materials mentioned in this article are available at <https://github.com/JoannaBy/The-Voices-of-Doctor-Who>

[2] For just one instance, see Craig and Kinney's *Shakespeare, Computers, and the Mystery of Authorship* (2009), scathingly reviewed by Vickers (2011) yet defended by Burrows (2012) and Hoover (2012).

[3] As of 2018, the Doctor new regeneration is female. This new Doctor, played by Jodie Whitaker, identifies as a woman while the past 12 regenerations were played by men, which is why I refer to the character using singular they/their/them pronouns unless I mean a specific regeneration.

[4] Subtitling standards are discussed in many papers and fora, e.g. a brief summarization here:

<https://translationjournal.net/journal/04stndrd.htm> As a side note, the use of subtitles being a translation from the original for an analysis of film/series content makes things even more complicated, highly criticised in both translation and film studies, with Sarah Kozloff explaining: "Subtitles only translate a portion of the spoken text, and only that portion that the subtitler has decided is most important. This filters out emphases that may be unique to the film or to that national cinema. Repetitions, interruptions, slang, curses, antiquated diction, regional accents, of course, are all lost in subtitles" [Kozloff 2000, 26].

[5] Collected from <http://www.chakoteya.net/DoctorWho/index.html> [last access: 10th Jan 2019]

[6] <https://thescriptlab.com/features/screenwriting-101/10415-television-writing-wisdom-from-doctor-who-showrunner-and-script-editor/> [access: 07.05.2020]

Works Cited

Andersen 2012 Andersen, T.R. "Judging by the Cover". *Critique: Studies in Contemporary Fiction* 53 (2012): 251–278.
<https://doi.org/10.1080/00111619.2010.484038>

Arnold and Tilton 2017 Arnold, T., Tilton, L. "Distant Viewing: Analyzing Large Visual Corpora". *Digital Scholarship in the Humanities* (2017).

Bateman and Schmidt 2011 Bateman, J.A., Schmidt, K.-H. *Multimodal Film Analysis: How Films Mean*, Routledge Studies in Multimodality. Routledge (2011).

Bateman and Wildfeuer 2014 Bateman, J., Wildfeuer, J. "A multimodal discourse theory of visual narrative". *Journal of Pragmatics*, Vol. 74 (2014): 180-208. <https://doi.org/10.1016/j.pragma.2014.10.001>

Bednarek 2010 Bednarek, M. *The Language of Fictional Television: Drama and Identity*. Continuum, London/New York

(2010).

Bednarek 2018 Bednarek, M. *Language and Television Series: A Linguistic Approach to TV Dialogue*. Cambridge University Press, Cambridge (2018).

Bednarek and Zago 2018 Bednarek, M., Zago, R.. Bibliography of linguistic research on fictional (narrative, scripted) television series and films/movies, version 2 (2018). Available at <http://unipv.academia.edu/RaffaeleZago> [date of last access: 07.05.2020]

Benatti and Tonra 2015 Benatti, F., Tonra, J. "English Bards and Unknown Reviewers: a Stylometric Analysis of Thomas Moore and the Christabel Review". *Breac: A Digital Journal of Irish Studies*, 4 (2015).

Bhargava et al. 2013 Bhargava, M., Mehndiratta, P., Asawa, K. "Stylometric Analysis for Authorship Attribution on Twitter". In: Bhatnagar, V., Srinivasa, S. (Eds.), *Big Data Analytics, Lecture Notes in Computer Science*. Springer International Publishing (2013): 37–47.

Blondel et al. 2008 Blondel, V.D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E. "Fast unfolding of communities in large networks". *Journal of Statistical Mechanics: Theory and Experiment Volume* (2008) <https://doi.org/10.1088/1742-5468/2008/10/P10008>

Burrows 2002 Burrows, J. "Delta': a Measure of Stylistic Difference and a Guide to Likely Authorship," *Literary and Linguistic Computing*, 17(3), (2002), 267–287.

Burrows 2007 Burrows, J. F. "All the way through: testing for authorship in different frequency strata". *Literary and Linguistic Computing*, 22(1) (2007), 27-48.

Burrows 2012 Burrows, J. "A Second Opinion on 'Shakespeare and Authorship Studies in the Twenty-first Century'". *Shakespeare Quarterly* 63 (2012), 355–392.

Busso and Vignozzi 2017 Busso, L., Vignozzi, G. "Gender Stereotypes in Film Language: A Corpus-Assisted Analysis". In: Basili, R., Nissim, M., Satta, G. (Eds.), *Proceedings of the Fourth Italian Conference on Computational Linguistics CLiC-It 2017*. Accademia University Press, (2017): 71–76. <https://doi.org/10.4000/books.aaccademia.2367>

Cafiero and Camps 2019 Cafiero, F., Camps, J.-B. "Why Molière most likely did write his plays" *Science Advances* 5, (2019) <https://doi.org/10.1126/sciadv.aax5489>

Chapman 2014 Chapman, J. "Fifty Years in the TARDIS: The Historical Moments of Doctor Who, Fifty Years in the TARDIS: The Historical Moments of Doctor Who". *Critical Studies in Television* 9, (2014): 43–61.

Choinski et al. 2019 Choiński, M., Eder, M. and Rybicki, J. "Harper Lee and other people: a stylometric diagnosis". *Mississippi Quarterly*, 70(3), (2019) forthcoming.

Craig and Kinney 2009 Craig, H. and Kinney, A. F. eds. *Shakespeare, Computers, and the Mystery of Authorship*. Cambridge: Cambridge University Press (2009).

Davies 2008 Davies, T R., Cook, B. *Doctor Who: The Writer's Tale*. BBC Books (2008).

Eder 2011 Eder, M. "Style-markers in authorship attribution: a cross-language study of authorial fingerprint". *Studies in Polish Linguistics*, 6 (2011): 99-114.

Eder 2016 Eder, M. "Rolling stylometry". *Digital Scholarship in the Humanities*, 31(3) (2016): 457-469.

Eder 2017 Eder, M. "Visualization in stylometry: Cluster analysis using networks". *Digital Scholarship Humanities* 32, (2017): 50–64. <https://doi.org/10.1093/lhc/fqv061>

Eder et al. 2016 Eder, M., Rybicki, J., Kestemont, M. "Stylometry with R: A Package for Computational Text Analysis". *The R Journal* 8, (2016): 107–121.

Edwards 2014 Edwards, V.L. "Fifty Years of Science Fiction Television". *Administrative Theory & Praxis* 36.3, (2014): 373-97.

Edwards 2015 Edwards, O.D. "As We See, So We Learn': Doctor Who as Religious Education." *Implicit Religion* 18.4, (2015): 527-40.10.

Evert et al. 2017 Evert, S., Proisl, T., Jannidis, F., Reger, I., Pielström, S., Schöch, C., Vitt, T. "Understanding and explaining Delta measures for authorship attribution". *Digital Scholarship Humanities* 32, ii4–ii16 (2017) <https://doi.org/10.1093/lhc/fqx023>

Finer and Pearlman 2004 Finer, A., Pearlman, D. *Starting Your Television Writing Career: The Warner Bros. Television Writers Workshop Guide*. Syracuse University Press (2004).

Gibbs 2013 Gibbs, A. “Maybe that's what happens if you touch the Doctor, even for a second”: Trauma in Doctor Who”. *Journal of Popular Culture*, 46, (2013): 950–972.

Górski et al. 2014 Górska, R., Eder, M. and Rybicki, J. “Stylistic fingerprints, POS tags and inflected languages: a case study in Polish”. *Qualico 2014: Book of Abstracts* (2014): 51-53.

Hartley 2004 Hartley, J. “From Republic of Letters to Television Republic? Citizen readers in the era of broadcast television”. In: Spigel, L. & Olsson, J. (Eds.) *Television after TV: Essays on a Medium in Transition*. Duke University Press, (2004): 386-417.

Hills 2010 Hills, M. “Triumph of a Time Lord: Regenerating Doctor Who in the Twenty-First Century”. I.B. Tauris & Co. Ltd., (2010).

Hills 2013 Hills, M. *New Dimensions of Doctor Who: Adventures in Space, Time and Television: Exploring Space, Time and Television*. I.B. Tauris & Co. Ltd., (2013).

Hoover 2012 Hoover D.L. 2012. “The Rarer They Are, the More There Are, the Less They Matter”, Digital Humanities Conference Abstracts, Hamburg, 218–220.

Hołobut and Rybicki 2018 Hołobut, A., Rybicki, J. “Pride and Prejudice and Programming: A Stylometric Analysis”. In: Raw, L. (Eds.), *Adapted from the Original. Essays on the Value and Values of Works Remade for a New Medium*. McFarland & Company, Inc., Jefferson, North Carolina (2018).

Hołobut and v 2018 Hołobut, A., Woźniak, M. *Historia na ekranie: Gatunek filmowy a przekład audiowizualny*. Wydawnictwo Uniwersytetu Jagiellońskiego (2018).

Hołobut et al. 2016 Hołobut, A., Rybicki, J. Woźniak, M. “Stylometry on the Silver Screen: Authorial and Translatorial Signals in Film Dialogue”. Digital Humanities 2016: Conference Abstracts. Kraków: Jagiellonian University (2016).

Jensen 2017 Jensen, M. “From the Mind of David Simon’: A Case for the Showrunner Approach”. *Series - International Journal of TV Serial Narratives* 3, (2017): 31–42. <https://doi.org/10.6092/issn.2421-454X/7610>

Juola et al. 2011 Juola, P., Ryan, M., Mehok, M. “Geographically Localizing Tweets Using Stylometric Analysis”. In: *Proceedings of the American Association of Corpus Linguistics 2011* (2011).

Kestemont 2014 Kestemont, M. “Function Words in Authorship Attribution. From Black Magic to Theory?” *Proceedings of the 3rd Workshop on Computational Linguistics for Literature (CLFL)* (2014), pp. 59–66.

Kjell 1994 Kjell, B. “Discrimination of authorship using visualization”. *Information Processing and Management*, 30 (1) (1994): 141–50.

Kozloff 2000 Kozloff, S. *Overhearing Film Dialogue*. University of California Press (2000).

Lavik 2015 Lavik, E. *Forfatterskap i TV-drama*. Universitetsforlaget (2015).

Martin 2013 Martin, B. *Difficult Men*. Faber & Faber (2013).

McKenna et al. 1999 McKenna, Wayne, John Burrows, Alexis Antonia. “Beckett’s Trilogy: Computational Stylistics and the Nature of Translation”. *R/SSH*, 35 (1999), pp. 151-171.

Mittell 2015 Mittell, J. *Complex TV: The poetics of contemporary television storytelling*. NYU Press (2015).

Mittell 2017 Mittell, J. *Narrative Theory and Adaptation*. Bloomsbury (2017).

Moretti 2000 Moretti, F. “Conjectures on World Literature”. *New Left Review* 1(2000): 54–68.

Ochab et al. 2019 Ochab, J. K., Byszuk, J., Pielström, S. and Eder, M. “Identifying similarities in text analysis: hierarchical clustering (linkage) versus network clustering (community detection)”. *Digital Humanities 2019: Book of Abstracts* (2019) <https://dev.clariah.nl/files/dh2019/boa/0981.html>.

Plecháč and Flaišman 2017 Plecháč, P., Flaišman, J. “Problém Barák–Neruda z pohledu současné stylometrie”. *Ceska Literatura* 65, (2017): 743–769.

Richardson 2010 Richardson, K. *Television Dramatic Dialogue: A Sociolinguistic Study*. Oxford (2010).

- Rybicki and Heydel 2013** Rybicki, J. and Heydel, M. (2013). "The stylistics and stylometry of collaborative translation: Woolf's 'Night and Day' in Polish". *Literary and Linguistic Computing*, 28(4): 708-17.
- Schler et al. 2006** Schler, J., Koppel, M., Argamon, S., Pennebaker, J. "Effects of Age and Gender on Blogging". In: AAA 1 Spring Symposium on Computational Approaches for Analyzing Weblogs (2006): 6.
- Schmidt 2015** Schmidt, B.M. "Plot arceology: A vector-space model of narrative structure". In: 2015 IEEE International Conference on Big Data (Big Data). Presented at the 2015 IEEE International Conference on Big Data (Big Data), (2015): 1667–1672. <https://doi.org/10.1109/BigData.2015.7363937>.
- Stamatatos 2009** Stamatatos, E., "A survey of modern authorship attribution methods". *Journal of the Association for Information Science and Technology* 60, (2009): 538–556.
- Stamatatos 2013** Stamatatos, E. "On the Robustness of Authorship Attribution Based on Character N-gram Features", 21 J. L. & Pol'y (2013).
- Steiner 2015** Steiner, T. "Steering the Author Discourse: The Construction of Authorship in Quality TV, and the Case of Game of Thrones". *Series - International Journal of TV Serial Narratives* 1, 181 (2015). <https://doi.org/10.6092/issn.2421-454X/5903>
- Tulloch and Alvarado 1983** Tulloch, J., Alvarado, M. *Doctor Who: The Unfolding Text*. MacMillan (1983).
- Vickers 2011** Vickers, B. "Shakespeare and Authorship Studies in the Twenty-first Century". *Shakespeare Quarterly* 62 (2011): 106–142.
- Wildfeuer 2014** Wildfeuer, J. "Coherence in Film: Analysing the Logical Form of Multimodal Narrative Discourse". In: *Multimodal Epistemologies: Towards an Integrated Framework*, Routledge Studies in Multimodality (2014): 260–274.
- Zyl and Botha 2016** Zyl, M. van, Botha, Y. "Stylometry and characterisation in The Big Bang Theory". *Literator* 37, 11 (2016). <https://doi.org/10.4102/lit.v37i2.1282>