Pandemics, epidemics, viruses, plagues, and disease: Comparative frequency analysis of a cultural pathology reflected in science fiction magazines from 1926 to 2015

Christopher B. Menadue [*]

The Cairns Institute, James Cook University, Cairns, QLD, 4870, Australia

ABSTRACT

Science fiction includes many dystopian narratives, often featuring epidemics, pandemics, plagues, viruses, and disease. As science fiction has grown in popularity and prevalence it appeals to an increasingly broad demographic, is employed in research communication and education, and as a genre it is frequently argued that it reflects contemporary cultural interests and concerns. To identify the relevance of science fiction as an indicator of popular trends relating to the pathologies of disease, a word frequency comparison of selected key words found in the Google Books 2012 English Corpus has been made to a representative corpus of science fiction magazines dating between 1926 and 2015. Selected issues were reviewed to identify concepts, situations, and outcomes that could readily be measured against real-world examples from current and recent pandemics. The findings indicate that science fiction does appear to mirror and magnify contemporary literary trends, and provides potentially revealing correlations to real-world historical events. In this regard, science fiction might be regarded as a form of 'cultural pathology' of popular interests related to the spread and impact of disease that may be valuable in gauging the degree to which society is engaged with these topics at any specific time.

1. Introduction

Science fiction has been described by the author Brian Aldiss as a form of 'cultural wallpaper' (Aldiss & Wingrove, 1986), and is increasingly employed in research, education, science advocacy, and communication as an easily recognised public product (Menadue & Cheer, 2017). The genre appears to have a special relevance for the public as a source of thought experiments and speculation about the way in which people think about the human capacity to bring about change through the use of tools and technology (Menadue, Giselsson, & Guez, 2019). In this regard, it may be regarded as a potent indicator of contemporary popular interests, and mapping of science fiction texts against historical changes in society and culture supports this approach (Menadue, 2017, 2018a, 2018b; Menadue, 2019a). It is a genre that has an increasingly broad popularity and influence, with a consumer demographic today that is highly educated, takes an interest in real-world science, and which has come to reflect a less specialised audience than might have been the case in the early 20th Century (Berger, 1977; Menadue & Jacups, 2018).

Science fiction is well known for significantly dystopian content, and the employment of dystopian scenarios across a wide range of cultural topics including gender, race, and health (Barrish, 2016; Bethlehem, 2014; Dziubinskyj, 2007; Gilarek, 2012; McFadden, 2015; Milerius, 2011; Milner, 2009; Penteado, 2015; Raphals, 2013; Rhines, 2003). As a literature of thought experiments that investigate dystopian scenarios and the implications of human impacts on our world, and our technological interactions with it, it might be expected that science fiction would feature diseases, plagues, epidemics, and pandemics, and the viral vectors of these. As a reflection of cultural interests, the frequency and prominence of these topics would focus on contemporary events and public concerns. The first notable historic work of science fiction that discusses pandemics and their potential to radically impact on human society is perhaps Mary Shelley's *The Last Man*, first published in 1826 (Shelley, 2015). The novel describes a far future (to Shelley) pandemic which eradicates all human life towards the end of the 21st Century and includes contemporary scientific discussion of vaccination, as well as sociological factors such as the conflict between religious and scientific representations of cause, effect, and cure.

Using digital tools to analyse text is a relatively modern approach to content analysis, although it draws upon earlier methods of exigesis and bibliographic study (Menadue, 2019b; Menadue, 2020). Increasing

---

* Academic Development Manager University of Suffolk, Ipswich, UK.
  E-mail address: ben.menadue@jcu.edu.au.

technological sophistication and computing power has led to the development of sophisticated methods and approaches to text analysis in the digital humanities (Bainbridge, 2004; Berry, 2012; Berry & Kogan, 2010; Clement, 2013; Flanders, 2013, p. 10; Halliday & Webster, 2005; Hammond et al., 2013; Hoover, 2013). As a sub-set of techniques, text mining has become a commonplace method for gathering data for analysis in biomedical and technological research, computing and social media, and the social and political sciences more generally (Abbe, Grouin, Zweigenbaum, & Falissard, 2016; Altman et al., 2008; Fogelman-Soulié, 2008; Irfan et al., 2015; Kaal et al., 2014; Obradović et al., 2013; Raychaudhuri, Chang, Imam, & Altman, 2003; Rodriguez-Esteban, 2009; Rodriguez-Esteban & Bundschus, 2016; Schummer, 2005; Thomas et al., 2011; Thompson et al., 2016; Wang et al., 2016; Zweigenbaum, Demner-Fushman, Yu, & Cohen, 2007).

An especially important resource for comparative corpus studies has emerged from the Google Books project, which has involved the digitisation of millions of books and the classification and tagging of the digitised text by parts of speech and grammatical use. Google Ngram data harvested from this long-running work on digitising literary corpora in several languages is freely available online (Michel et al., 2011). For this study, the Google corpus has been employed to provide benchmarking of proportionate frequency. It indicates the fluctuation of frequencies for specific terms over a time series for comparison with changes occurring in the selected science fiction corpus. Frequency graphs offer opportunities to identify correlation with real-world events that may have influenced the frequency of these topics arising in the literature. When examining the prevalence and degree of focus – gauged by word frequency – on dystopian topics relating to disease, infection, and death, it seems reasonable to label this feature of science fiction as an example of 'cultural pathology.' We find insights into the diffusion and emergence of popular interests that grow in currency among a specific cultural group, spread, and become endemic, or are perhaps even eradicated over time. Evidence of correlations between representations in science fiction of these topics and real-world historical trends may support the conjecture that science fiction focuses a lens on contemporary human concerns and interests.

2. Materials and methods

Two English literature corpora are contrasted in this study. The reference corpus is the Google Books 2012 English corpus, which contains word frequency data for scanned texts in the English language to 2008. The corpus for comparison is a digitised collection of 4490 science fiction magazines dated between 1926 and 2015. The rationale for selecting magazine fiction is that it has features that arguably make it more closely aligned to contemporary culture than longer fictional forms. These include a shorter cycle between writing and publishing, and the inclusion of advertising, editorials and readers letters, as well as factual articles. These features of magazines have been identified as making this particular ephemeral literary medium especially suitable for exploring the social and cultural environment from which the content is derived. It offers a more comprehensive sample of the influences on production than is found in novels while being less subject to stochastic changes that affect the content of daily newspapers (Scholes & Latham, 2006; Scholes & Wulfman, 2010). The date range for comparison commences in 1926 as *Amazing Stories* was first published in April of that year and marks the beginning of a rapid growth and expansion of magazine science fiction publication and readership, which is reflected in the density of the corpus of titles and issues available for study. This is a feature which enables the use of these sources as a coherent body for comparative analysis. Science fiction is an 'open' genre that accepts content that may be proscribed or limited in other forms such as historical novels or romantic fiction, making it more likely that we might find a wider range of interpretations and speculation in science fiction due to the lack of artificial restrictions on content.

The sample employed here represents approximately one quarter of

Table 1
Initial search and filtering results. Counts of terms appearing in the science fiction corpus and number of issues in which they are found.

Science Fiction	Initial search		Filtered		Post filtering	
Term	Mentions	issues	mentions	issues	mentions	issues
Epidemic	1464	806	29	12	1435	794
Pandemic	107	78	5	1	102	77
Plague	8503	2583	174	65	8329	2518
Virus	6592	1522	90	29	6502	1493
Disease	14817	3326	251	85	14566	3241
Total	31483	8315	549	192	30934	8123

the total professional and semi-professional publication of magazines that primarily featured science fiction content over this period including complete runs of the popular 20th Century titles *Astounding Science Fiction* (n = 361), *Galaxy Magazine* (n = 261), *Imagination* (n = 63), and *Worlds of If* (n = 176), and a significant proportion of issues of *Amazing Stories* (n = 437), *Analog* (n-208), and *Asimov's Science Fiction* (n = 323). The exhaustive *Science Fiction, Fantasy, & Weird Fiction Magazine Index* (Miller et al., 2019) lists 29,400 issues and is more broadly inclusive of content types, including earlier and more recent publications, as well as minor publications with low circulation numbers. The relative homogeneity of magazine publishing in terms of word counts and proportions of non-fiction content is also a valuable feature of the genre for text mining applications. A simple text search was used to identify issues containing the target words and the number of occurrences in each issue (Table 1.).

Issues were filtered to exclude duplicates and anthologies, for which the date of original publication of content would not match the issue publication date and might introduce skew in potential matches to historical events. The initial search was carried out on 4490 issues, each stored as a single text file. Filtering removed 108 issues during preliminary sorting by type and date of individual mentions, leaving 4382 issues for further analysis. Table 1 shows aggregate totals for mentions (the number of times a term occurs) and issues in which the terms appear. This demonstrates an overlap, with multiple terms appearing in the same issues. This is discussed in Section 3.5 (below) in the comparative frequency analysis for occurrences of the word 'disease.'

Concatenation of the corpus text files and use of the command line word count found a total of 296,347,954 words in the filtered science fiction texts. Dividing this figure by the total of 4490 texts established an average word count of 66,002 words per issue. This value was used to calculate frequencies of occurrence of individual terms. Issue names, dates of publication, and word frequencies for each word were tabulated and graphed in Microsoft Excel. Word frequency tables were calculated for the following terms: epidemic, pandemic, plague, virus, and disease in any format or grammatical use. Google Books Ngram Viewer (https://books.google.com/ngrams/) was searched to identify the frequencies of the same terms in all forms found between 1926 and 2008 in the English 2012 corpus. The resulting word frequencies for each term by year were exported into Excel for tabulation and comparison with the results of the search of science fiction texts. The supplemental file includes the list of science fiction sources and all data analysis undertaken for this study. The Google corpus currently does not extend beyond 2008, so direct comparison was not possible for the science fiction examples found between 2009 and 2015. These were included to indicate more recent historical trends in the literature. A combination of purposive and random sampling was used to identify individual issues of magazines for closer examination and discussion of content, as specified in the results.

3. Results and Discussion

Average word frequencies found in the science fiction corpus were higher than those in the general Google corpus across all terms except disease (Table 2, Fig. 1).

It was clear from the analysis that the word frequencies of 'pandemic,'

Table 2
Average frequencies of occurrence (%) of search terms.

Average Frequencies	epidemic (All)	pandemic (All)	plague (All)	virus (All)	disease (All)
Google (2012):	0.0000069	0.0000004	0.0000083	0.0000214	0.0001087
Science Fiction:	0.0000275	0.0000094	0.0000493	0.0000584	0.0000680
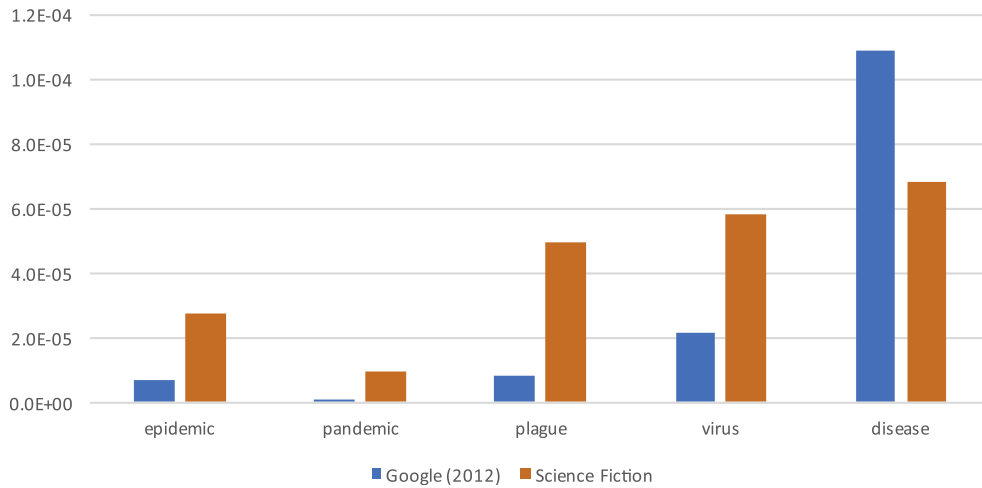
Fig. 1. Average word frequency (%) comparison between Google (2012) and science fiction corpus for the five search terms: 'epidemic,' 'pandemic,' 'plague,' 'virus,' and 'disease.'

'epidemic,' 'virus,' and 'plague' were significantly higher in science fiction literature than in the Google English corpus. The comparatively high frequency of these terms in science fiction may be an indicator of the dystopian focus of the genre, which is reflected in a focus on severe and devastating events. The lower frequency of the more general (and less sensational) word 'disease' may additionally reflect a tendency for science fiction to feature technical and specific language. As general scholarly and non-scholarly literature might be expected to discuss disease across a broader range of contexts including the inconsequential, incidental, and informative, it follows that it is unsurprising we would find epidemics, pandemics and plagues featuring more frequently in science fiction. Comparative analysis for the term 'pandemic' is less reliable than for the other terms due to the relatively low numbers of occurrences of the term over the period under study (See Supplementary Data File), with 107 mentions in 77 issues in the science fiction corpus, but there is an apparent upward trend in the appearance of this term towards the end of the period, which matches a similar increase in frequency in the Google English corpus, and closer analysis of content in Section 3.2 (below) indicates significant correlations with contemporary experiences and expectations of pandemics including the current COVID-
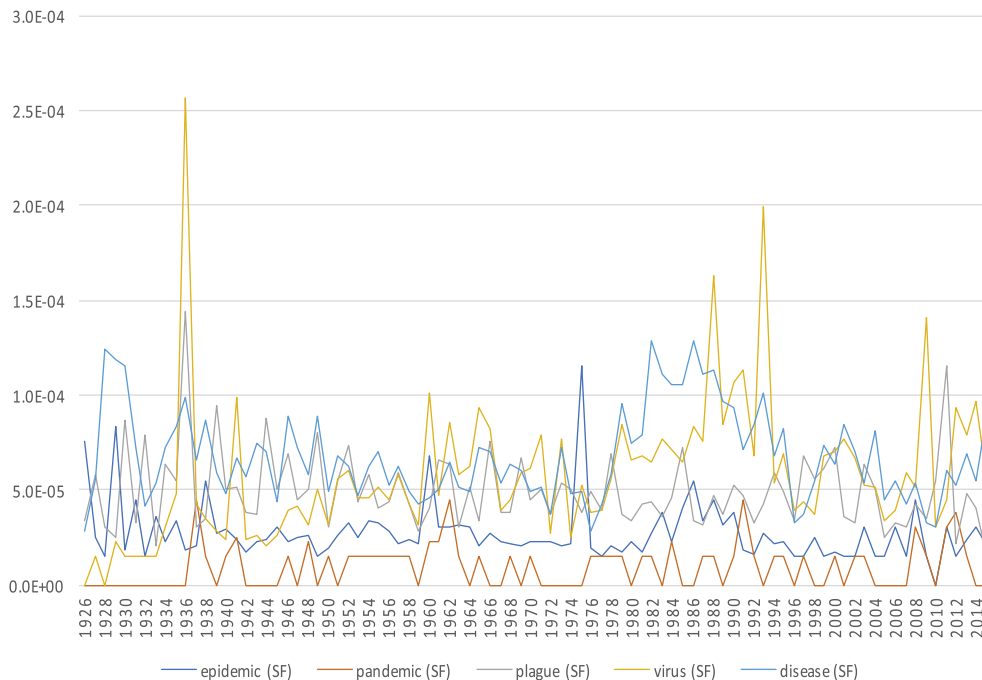
Fig. 2. Word frequencies (%) in the science fiction corpus 1926–2015 overlaid for all search terms: 'pandemic,' 'plague,' 'virus,' 'disease,' and 'epidemic.'

19 outbreak.

The word frequencies found in science fiction (Fig. 2) exhibit multiple peaks within longer-term trends, and periods in which specific terms increased or decreased in prevalence. There is also a tendency for frequencies of pathological terms to change in cohort, for example between 1977 and 1997, when 'epidemic,' 'disease,' and 'virus' all feature a gradual rise and fall in frequency.

The frequency graph for science fiction mentions show a marked similarity to identified historic trends in emerging infectious diseases (EIDs) between 1940 and 2000, which evidence a similar surge from the 1970s through the 1980s, and a decline in the following decade. It has been concluded in scientific literature that this trend in EIDs was significantly influenced by increasing vulnerabilities arising from the HIV/AIDS pandemic (Jones et al., 2008) and this appears to be consistent with the findings of this article for word frequencies in science fiction.

In the Google 2012 corpus, the prevalence of 'disease' was significantly higher than other terms, and a logarithmic scale has been used to show comparative frequencies of occurrence (Fig. 3).

Comparison of individual terms over the time series in both Google 2012 and science fiction show some similarities in the changes over time, which may indicate that science fiction follows real-world events and factual reporting that are included in non-fictional volumes in the Google corpus. The visible correlation between frequencies in the science fiction corpus and the EID data on the HIV/AIDS pandemic over the same historical period (Jones et al., 2008) is more marked than in the general corpus for the terms 'epidemic,' 'virus,' and 'disease.' The term 'pandemic' appears more consistently in science fiction, as well as in the general corpus, since the emergence of SARS, H1N1 Swine Flu and MERS in the first decade of the 21st Century.

3.1. Epidemic

There is a consistently higher word frequency of 'epidemic' in the science fiction corpus (Fig. 4. Note larger secondary scale used to plot Google frequencies). As noted previously (3. Results and Discussion), the surge in the 1980s coincides with the historical emergence of HIV/AIDS as a global issue, and may reflect the influence of historically significant cultural concerns on science fiction content.

The sources also reveal more generalised comprehension and interests. The localised spike between 1974 and 1975 arises from 49

occurrences in 8 publications and 21 individual issues in those two years, and a sample examination of all occurrences from one publication, *Galaxy Magazine* (n = 4), in this interval yields examples that are of current relevance to COVID-19, and illustrates the range of interests in the subject. In *Web of Everywhere*, by John Brunner, serialised in *Galaxy* in March and April 1974, we find a description of epidemics resulting from technology enabling rapid global transportation combined with vaccination hesitancy: 'Sickness after killing sickness had exploded like shrapnel from the few surviving reservoirs in less fortunate areas of the world into those whose inhabitants had neglected immunization procedures as needless' (Brunner, 1974). In July, we find Isaac Asimov writing a guest editorial about evolution driven by 'an endless epidemic of random death to make sure of an endless turnover of generations with their new gene combinations' that presages discussions of a self-regulating global ecology (Asimov, 1974). In August, we find a story featuring classic dystopian themes of post-apocalyptic devastation exacerbated by a lack of civilised infrastructure able to respond to infectious disease: 'fire scourged the cities and the short plagues followed, radiation deaths and epidemics of crushing virulence with no means for study or control' (Pangborn, 1974) – the progression of recent epidemics appears to confirm such 'study and control' is essential. In October we are told the threat of an epidemic 'waited on the wings of human experience like spectres of death' in a low technology society (Brennan, 1974). These samples illustrate that aside from mapping changing interests over time science fiction content mirrors more general reflections and understandings of epidemics.

3.2. Pandemic

The frequency of occurrence of 'pandemic' is the lowest for all terms, but we can still observe some similarities between science fiction and Google corpus frequencies (Fig. 5. Note larger secondary scale used to plot Google frequencies), with a surge in representations as we enter the 21st Century, and spikes in the mid-to-late-1930s. The granularity of data suggests caution in interpreting these results, however. We might expect the frequency of recent occurrences to be influenced by the emergence of SARS and MERS as pandemics causing significant cultural concern. A purposive selection of the most recent (2011–2013) occurrences of 'pandemic' from the science fiction corpus appears to support this conjecture. We find examples of behaviours, actions, and consequences that are very familiar to us in the current Coronavirus pandemic. In
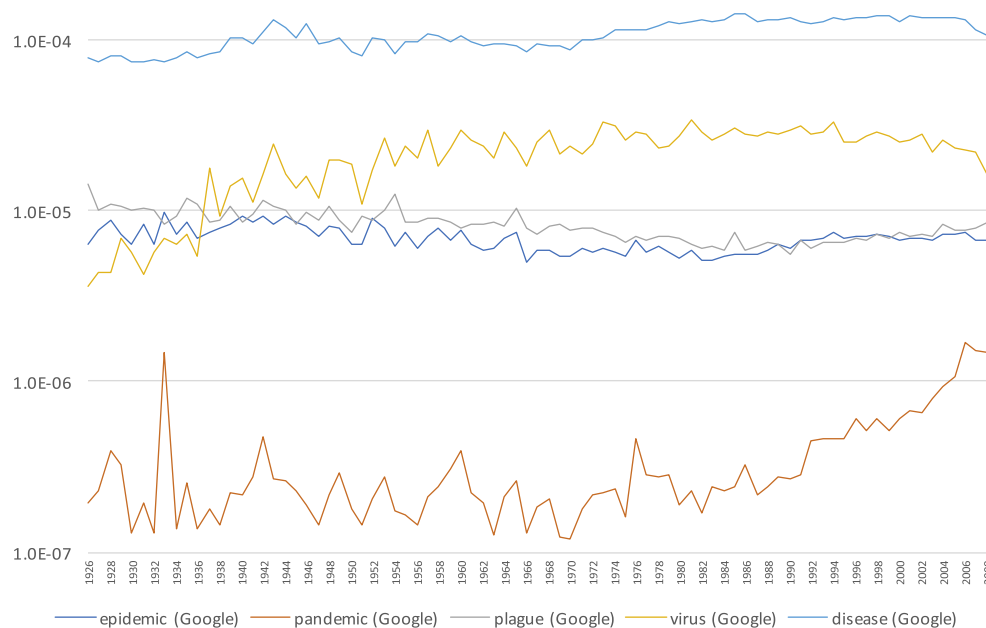

Fig. 3. Word frequencies (%) 1926–2008, Google (2012) corpus (logarithmic scale, base 10) for all search terms: 'epidemic,' 'pandemic,' 'plague,' 'virus,' and 'disease.'

'epidemic'

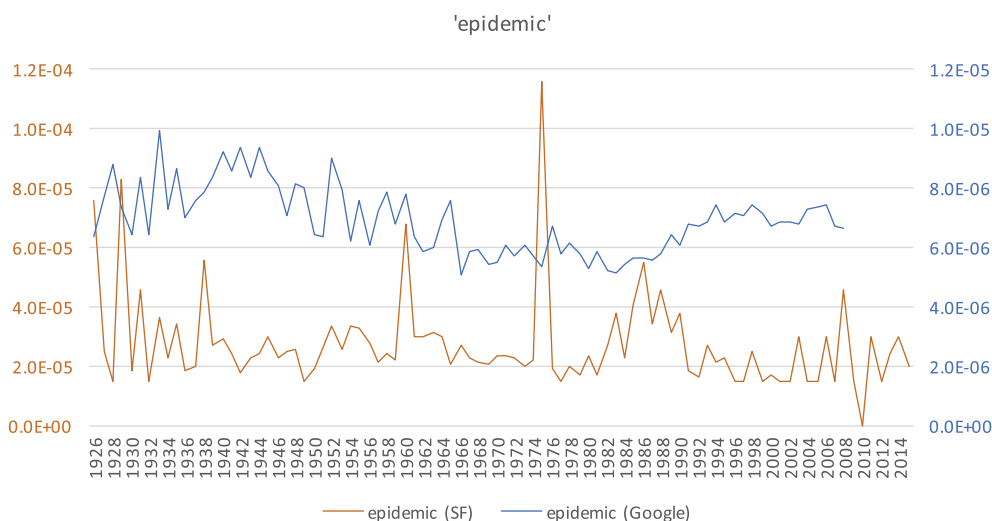**Fig. 4.** Comparison of word frequency (%) 'epidemic' in Google (2012) English corpus 1926–2008 and Science Fiction magazines 1926–2015.
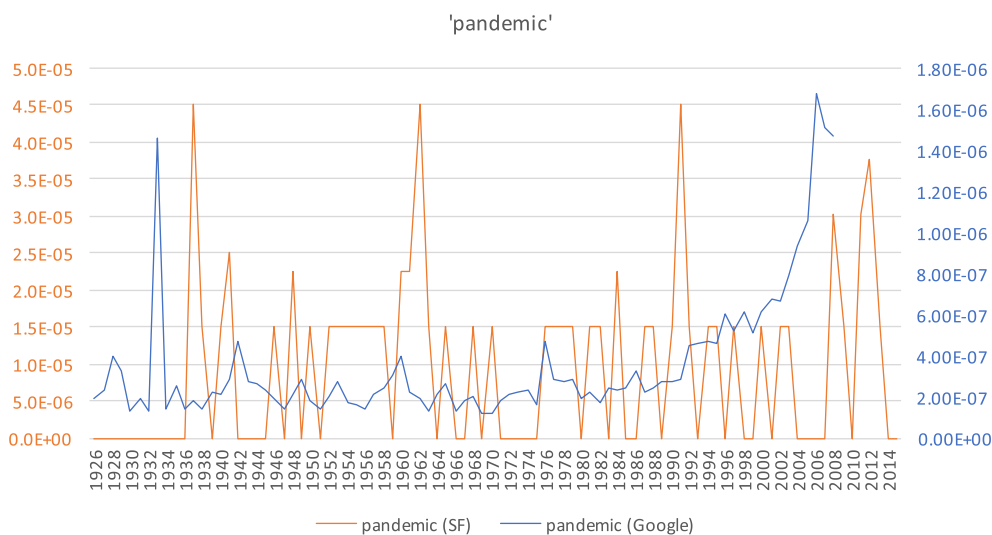
'pandemic'

Fig. 5. Comparison of word frequency (%) 'pandemic' in Google (2012) English corpus 1926–2008 and Science Fiction magazines 1926–2015.

Edward Lerner's 2013 short story *Time Out* a time-traveller infects a contemporary defenceless world with a bioengineered plague from the future in an act of pro-environmental terrorism: 'The virus had first appeared in 2013. Lethal. Incurable. Airborne and extremely contagious … four billion dead worldwide … Whole cities, even countries, depopulated. Civilization imploding … Radical Naturists … exulted. A pandemic beyond the reach of medical science, until a vaccine was developed in 2016' (Pitkin, 2012). The story combines contemporary understandings of the potential for devastating pandemics with notions of environmental consciousness, employing standard time-travel paradoxes as a vehicle for this message. In *A Murmuration of Starlings* by Joe Pitkin, published in *Analog* in June 2012, we find echoes of the current Coronavirus crisis in a story that is seven years older. A pandemic is described that arises in East Asia and spreads rapidly. It leads to closure of schools, 'social distancing' measures, voluntary lockdown and self-isolation, and the adoption of facial masks 'which everyone suspected with various degrees of certainty to be useless' (Pitkin, 2012). In *Lion Dance*, published in *Asimov's Science Fiction* in the same year, we also find references to lockdowns and mask wearing, and shortages of medications for other conditions – specifically antivirals for HIV due to 'selfish jackholes' getting 'prescriptions from quacks,' 'understaffed and overflowing' hospitals, and the distractions of 'streaming movies online'

(Kaftan, 2012). The family of the story's narrator has lost their retail and restaurant businesses in San Francisco because 'the city might have stopped the germs, but they trashed the economy.' Explicit reference is made to the H2N2 influenza outbreak of 1957 as a comparative example (Kaftan, 2012), and we also discover the high mortality rate in the medical profession is causing chronic staffing shortages (Kaftan, 2012). *Grandma Said* in the September 2011 issue of *Asimov's Science Fiction* follows the story of a 'cleanser' whose role is to dispose of the bodies and sanitise the possessions of the victims of a pandemic fungal plague. In a world where untimely death is a familiar part of daily life, mental health issues are presented as a commonplace feature of a fatalistic sub-culture (Neube, 2011). It is clear even from this small sample of the stories and other content that refers to pandemics that the accuracy and faithfulness of representations to real-world lived experiences is a significant feature of the genre.

3.3. Plague

The term 'plague' appears more frequently in the science fiction corpus than in the Google (2012) English corpus (Fig. 6. Note larger secondary scale used to plot Google frequencies), featuring 8503 times over 2583 individual issues. There are also similarities visible in a
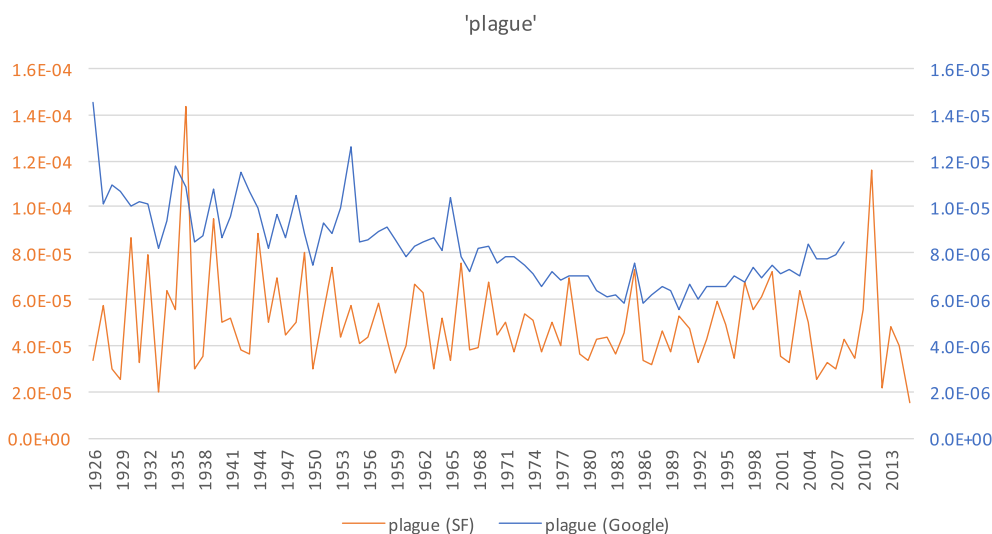
'plague'

Fig. 6. Comparison of word frequency (%) 'plague' in Google (2012) English corpus 1926–2008 and Science Fiction magazines 1926–2015.

general trend of gradual decline in frequencies as we approach the 1980s and recovery afterwards. Like 'disease,' however, 'plague' might be used freely to avoid word repetition for the purpose of introducing variety into a narrative. In literature that is often dystopian, we also expect 'plague' to be substituted for 'disease' when writers are looking for a generic term across a wide range of EIDs: 'plague' being associated with more serious, widespread and harmful historical events and infections, from bubonic plagues of early history through to HIV/AIDS. Supporting this, we find plague is a term applied to science fiction stories about EIDs with diverse causes as well as psychological and physical effects. A biochemical plague caused by a new element appears in the very first issue of *Amazing Stories* in April 1926. *The Man Who Saved The Earth* by Austin Hall describes the popular reaction of spreading panic as 'a moment of historical psychology' (Hall, 1926). In *The Flying Legion*, in the March 1930 issue of *Air Wonder Stories*, the issue of contact transmission between humans is noted, as well as an assumed geographic association of 'plague and other oriental diseases' when an archaeologist is inspecting a religious icon (England, 1926). This bears comparison to the misuse of 'Chinese flu' today as an appellation for COVID-19. In 1940, a plague 'some claimed … was the result of an experiment that had gone beyond control' destroys all life on Mars (Del Rey, 1940) mirroring contemporary rumours of COVID-19 as a bioengineered pathogen. The first issue of *Galaxy Magazine* in October 1950 describes the risk of new zoonotic plagues, even from alien forms of life: 'if the animals were like Earth animals, their diseases might be like Earth diseases, alike enough to be contagious, different enough to be impossible to treat' (MacLean, 1950). In *Nebula Science Fiction* in June 1958, a rise in automation leading to an age of leisure and plenty is described in cultural terms as a 'febrile mechanical plague' which has weakened Earth society and made it vulnerable to external invasion (Temple, 1958). An editorial on genetic utopias in *Analog* in March 1969 discusses the relative genetic immunity of different populations to measles, highlighting that for some it is a 'lethal plague' (Campbell, 1969). In 1979 Larry Niven describes a bacterial plague that has destroyed an advanced civilization by attacking the superconductor components of their electronic devices (Niven, 1979). In the January 1989 edition of *Interzone*, John Brosnan aggregates overpopulation, rampant inflation, plagues, and environmental damage in a summary of the state of a world that has driven people to live in artificial realities (Brosnan, 1989). The conflict between religious and scientific interpretations of the source of plagues is raised in *The Enclave* by Lois Tilton in 2000 (Tilton, 2000), and a fungal pandemic is described as a plague in 2011 (Neube, 2011). This sampling across the decades of mentions of plague in science fiction both illustrates the use of the term as a potentially more dramatic substitution for 'disease' and also the

versatility of the genre in highlighting a wide range of associations and effects that continue to be of popular interest.

3.4. Virus

A steady growth in mentions of 'virus' as the century progresses is found in both the general and science fiction corpus, but the frequency of occurrence in science fiction is considerably higher while following the same trend (Fig. 7. Note larger secondary scale used to plot Google frequencies). The lack of comparative Google corpus data post 2008 means that we cannot make corpus comparisons after that date, but the historical trends in the science fiction corpus show marked variations in frequencies that indicate more recent interest.

There are peaks in mentions of viruses that appear to correlate with significant historic epidemics including the Asian Flu pandemic of 1957–1958, the increasing public awareness of HIV/AIDS towards the end of the 1980s, and subsequent peaks contemporary to SARS 2002–2004, the West African Ebola epidemic of 2007, the 2009–2010 H1N1 'Swine Flu' pandemic, MERS in 2012, and the 2013–2016 West African Ebola epidemic. The visible peak occurring in 1936 does not appear to correlate with a specific historical epidemic, but may not be a statistically reliable indicator of genre word frequency as it derives from only two stories in two individual magazine issues in that year, and as noted in the limitations of this study, this is most likely to be a stochastic outlier (Section 3.6). Viruses are mentioned in 1522 out of 4385 individual magazine issues that comprise the science fiction magazine corpus. Considering the range of topics discussed in science fiction, the fact that viruses feature in 35% of the sample between 1926 and 2015 may seem surprising, but also reflects that magazines contain multiple stories as well as other articles, reviews and advertising. Some occurrences will be incidental, in lists of book titles for sale or reviews of films, for example.

Random sampling of ten of the issues in which viruses are mentioned finds a broad range of contexts. An advertisement for Listerine mouthwash in *Astounding Science Fiction* in March 1945 contains advice on the human use of disinfectant to fight viruses: 'THE TIME to fight a cold is at the very start … before potentially troublesome germs in the throat can stage a "mass invasion" of the tissue to aggravate the infection. In short, attack these germs before they attack you.' This is qualified when the advertisers add: 'Medical men say that though many colds may be started by a virus, these threatening "Secondary Invaders" can make it complicated and are responsible for much of its discomfort and misery' (Lambert Pharmacal, 1945). A feature editorial about censorship and scientific 'meddling' in *Authentic Science Fiction* in March 1953 cites a
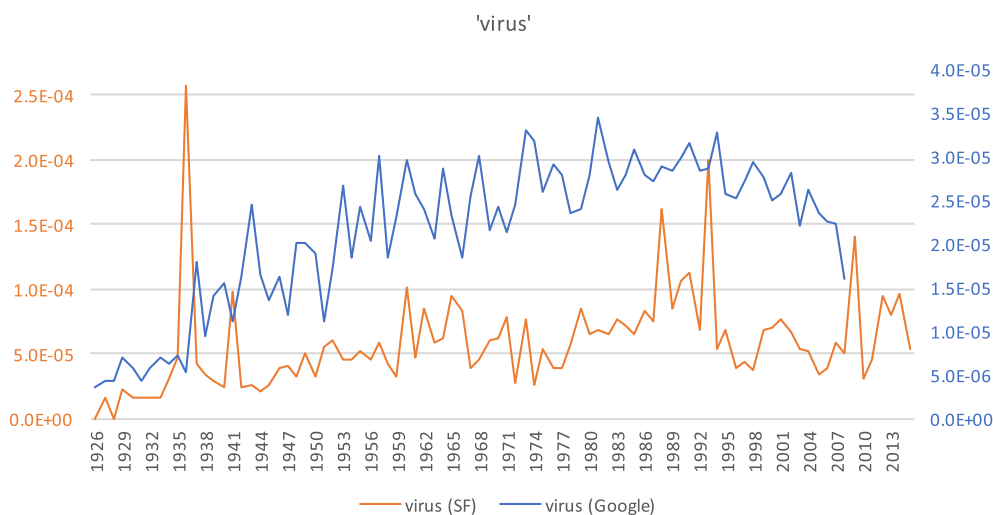
'virus'

Fig. 7. Comparison of word frequency (%) 'virus' in Google (2012) English corpus 1926–2008 and Science Fiction magazines 1926–2015.

journal article suggesting Russian scientists had found a way of converting viruses into bacteria, and consequently creating a new living organism (Temple, 1953). In 1956, James Blish describes a vault-dwelling society, deep underground, living in enforced isolation to escape from a long-running biological war. The society is afflicted with high levels of disaffection, crime, a mental health crisis, and is described as close to a collective 'psychotic break' from the effects of lockdown, with the authorities aware that trying to correct the social issues would have a devastating effect on the economy (Blish, 1956). The use of a retroviral technique to eliminate a post-viral skin colour change that has caused mass discrimination is a major plot feature of *Pariah Planet* by Murray Leinster (Leinster, 1961a, 1961b), and the story employs the viral theme to discuss issues around race and prejudice. Another story by Leinster from 1961 (Leinster, 1961a, 1961b) is set on a passenger ship, with health clearances and inspections imposed to guard against the spread of a deadly virus, which is defeated by employing a different viral species that preys on it. In a humorous story from 1974, we find British civil servants conspiring to destroy the French economy by employing a virus to destroy the grape harvest, forcing the French to re-enter the Common Market (which they have apparently left, with severe consequences to the British economy). The French plan to retaliate by releasing a virus which will eradicate tea (Peirce, 1974). In a 1976 science article in *The Magazine of Fantasy & Science Fiction* (Asimov, 1976), Isaac Asimov speculates on whether the destruction of the ozone layer will enable sufficient ultraviolet light to penetrate to kill viruses and other microorganisms, destroying the balance of life on Earth. Susceptibility to the same viral organism leads to the conclusion that an extinct Martian race evolved on Earth in Patricia Anthony's *Blood Brothers*, published in 1987 (Anthony, 1987). Viral RNA is encoded with the memories and personality of a cloned human in *Curtain Call* (Knowles, 1990), and the only appearance of a computer virus in this random sample of issues occurs in 2002 in a fictional story by the neuroscientist Tavis Allison (Allison, 2002). This indicative sample suggests science fiction examines mechanisms and impacts of viruses from biological, psychological, and socioeconomic perspectives.

3.5. Disease

Notably, 'disease' is the only word for which the frequency of occurrence in science fiction is consistently lower than the frequency with which it occurs in the more general corpus (Fig. 8). As noted above (3. Results and Discussion), this anomaly can be explained by the focus of SF on extreme and dystopian events as well as a preference for specific and technical language, with 'disease' likely to be found in general literature across a much broader range of contexts. The similarity in the

frequencies in SF between 'virus' and 'disease' from the late 1960s onward appears to be from the inclusion of 'disease' in the narrative of stories that feature viruses. Out of 1493 issues that include the word 'virus' over the period, 1269 (85%) also include the word 'disease.' The higher incidence of 'plague' compared to 'disease,' as discussed previously, may be the driver for the relatively lower frequency of word usage. 'Disease' is such a commonly occurring term in the corpus, with 14,566 examples being identified, that it is associated across all other terms and content types, consequently purposive contextual sampling for disease as a specific term has been omitted from this analysis.

3.6. Limitations

The distribution of digital magazine texts was not uniform over the period, with the majority of issues being published between 1950 and 1990. Consequently there are likely to be outlier effects that produce apparently significant results, such as the peak for mentions of viruses in 1936 that arises from only two issues. The close match between frequency changes and historic events for periods when more samples are available and the comparative lack of any apparently significant infectious disease outbreak around 1936 suggests this is an example of a stochastic outlier in this context, and might be clearly revealed as such if a larger sample of SF magazines were available for analysis for this year. The difference in scale between the Google corpus, which numbers in millions of books, and the science fiction magazine corpus (numbered in the thousands) is significant in that it increases the granularity of findings from the science fiction corpus, especially for terms which are inherently less frequently found than others. The granularity of results for the word 'pandemic' is due to the relative scarcity of the word in the science fiction texts, even though frequency is still significantly higher than is found in the Google English corpus as a whole. To carry out a close analysis of the context of the terms identified would require a significantly larger study and limited purposive sampling can be misleading. For the examples chosen here to illustrate the uses of terms in science fiction, purposive and random samples spread across multiple publications and decades was a pragmatic choice based on time and resources available. A future study might employ additional resources to investigate one or more topics at a different scale and level of detail.

4. Conclusions

Science fiction strongly features narratives on disease-related subjects and the prevalence of specific content appears to reflect real-world outbreaks both in timing and interest. There also appears to be a correlation between historical periods over which specific diseases have become
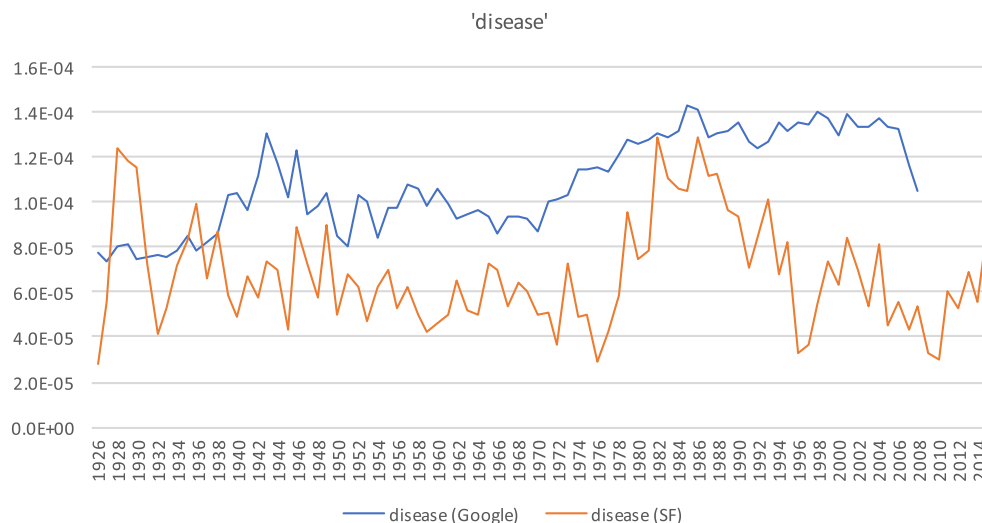
'disease'

Fig. 8. Comparison of word frequency (%) 'disease' in Google (2012) English corpus 1926–2008 and Science Fiction magazines 1926–2015.

more prominent in public awareness and an increase in these being featured in science fiction. Interestingly, the general term 'disease' was the only one which featured less frequently in science fiction than in the general English corpus, in significant contrast to the greater frequency of occurrences of virus, plague, epidemic and pandemic. The dystopian, technological and scientific content of science fiction may be responsible for this apparent anomaly, with stories perhaps tending to be more specific and specialised in their descriptions of diseases, and preferring terms that imply more severe and dangerous events (epidemic, pandemic, and plague).

The synchronisation of increasing occurrences of epidemic, pandemic, and viral themes during historical periods in which these have been real or emerging threats supports the hypothesis that science fiction is able to indicate the timing and intensity of a 'cultural pathology' of popular interests and concerns about disease, sickness, and death. This aligns with the frequently dystopian focus of the genre, but also reflects the lived experiences and expectations of the audience. Samples from issues feature specific contents supporting this finding, and also demonstrate how scientific and technical knowledge is included in the narrative. Many of these are immediately familiar in the current context of a Coronavirus pandemic. Narratives describe lockdowns, quarantine, wearing masks, zoonotic diseases, mis-use of medication, and socio-economic impacts. Science fiction may provide readily identifiable indicators of the intensity of public concerns and interests in outbreaks such as Ebola, MERS, SARS, and the current Coronavirus pandemic, as well as persistent threats such as HIV/AIDS. This includes concerns about the misuse of treatments as well as mental health impacts on people living in these high stress environments. The amplification of frequency of occurrence in science fiction of terms that are trending in the general corpus suggests that science fiction might be useful in providing early indicators for changes in public perspectives and focus that are more difficult to identify in a more general corpus. Useful insights into the awareness and understanding of disease could be expected in a population that has similar demographic characteristics to the science fiction audience. It may be we can look to the narratives of science fiction to provide us with the questions, if not the answers, that people will ask when faced with the threats to life, physical and mental health, and livelihoods that accompany emerging infectious diseases.

Funding

CRediT authorship contribution statement

Christopher B. Menadue: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Data curation, Writing - original draft, Writing - review & editing, Visualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Abbe, A., Grouin, C., Zweigenbaum, P., & Falissard, B. (2016). Text mining applications in psychiatry: A systematic literature review. *International Journal of Methods in Psychiatric Research, 25*(2), 86–100. https://doi.org/10.1002/mpr.1481

Aldiss, B. W., & Wingrove, D. (1986). *Trillion year spree: The history of science fiction.* London, England: Victor Gollancz.

Allison, T. (2002). In father christmas's court. *Asimov's Sci. Fiction, 26*(12). p. n.p. eBook.

Altman, R. B., Bergman, C. M., Blake, J., Blaschke, C., Cohen, A., Gannon, F., … Valencia, A. (2008). Text mining for biology - the way forward: Opinions from leading scientists. *Genome Biology, 9*(2). https://doi.org/10.1186/gb-2008-9-s2-s7. S7–S7.

Anthony, P. (1987). Blood Brothers. In *Aboriginal SF* (pp. 9–11). Woburg MA: The 2nd Renaissance Foundation.

Asimov, I. (1974). Forum: Is there Hope for the future? *Galaxy, 35*(7), 69–82.

Asimov, I. (1976). Silent victory. In *Fantasy & Science Fiction* (pp. 90–100). Conn. USA: Mercury Press.

Bainbridge, W. S. (2004). The evolution of semantic systems. *Annals of the New York Academy of Sciences*, 150–177. https://doi.org/10.1196/annals.1305.008

Barrish, P. (2016). Health policy in dystopia. *Literature and Medicine, 34*(1), 106–131. https://doi.org/10.1353/lm.2016.0006

Berger, A. I. (1977). *Science-fiction fans in socio-economic perspective: Factors in the social consciousness of a genre.* Science-fiction studies. *4*(3), 232–246.

Berry, D. M. (2012). *Understanding digital humanities.* , Basingstoke: Palgrave Macmillan.

Berry, M. W., & Kogan, J. (2010). *Text mining: Applications and theory* (Vol. 2). Chichester, U.K: Wiley. https://doi.org/10.1002/9780470689646. Aufl.;1.

Bethlehem, L. (2014). Lauren beukes's post-apartheid dystopia: Inhabiting moxyland. *Journal of Postcolonial Writing, 50*(5), 522–534. https://doi.org/10.1080/17449855.2013.813867

Blish, J. (1956). To pay the piper. In *Worlds of If.* (pp. 39–51). NY.

Brennan, H. (1974). The aerial machine. *Galaxy, 35*(11), 136–151.

Brosnan, J. (1989). An eye in paradise. In *Interzone* (pp. 39–43). Brighton: Interzone.

Brunner, J. (1974). Web of Everywhere. *Galaxy, 34*(6), 4–51.

Campbell, J. W. (1969). Genetics and utopia. In *Analog Science Fiction, Science Fact* (pp. 4–6). New York: Conde Nast.

Clement, T. (2013). *Text analysis, data mining, and visualizations in literary scholarship*. MLA Commons. https://doi.org/10.1632/lsda.2013.8

Del Rey, L. (1940). Dark mission. In *In Astounding Science Fiction* (pp. 117–129).

Dziubinskyj, A. (2007). Eduardo Urzaiz's 'Eugenia': Eugenics, gender, and dystopian society in twenty-third-century Mexico. *Science-Fiction Studies, 34*, 463–472.

England, G. A. (1926). The flying legion. In *Air Wonder Stories* (pp. 820–838). NY: Stellar Publishing.

Flanders, J. (2013). *The literary, the humanistic, the digital: Toward a research agenda for digital literary studies*. MLA Commons. https://doi.org/10.1632/lsda.2013.9

Fogelman-Soulié, F. (2008). *Treaty Organization. Public Diplomacy, Mining massive data sets for security: advances in data mining, search, social networks and text mining, and their applications to security* (Vol. 19). Oxford;Washington, D.C;Amsterdam, Netherlands: IOS Press.

Gilarek, A. (2012). *Marginalization of 'the other': Gender Discrimination in dystopian Visions by feminist science fiction authors*. Text matters: A journal of literature,. *Theor. Cult., 2*, 221–238. https://doi.org/10.2478/v10231-012-0066-3

Hall, A. (1926). The man Who saved the Earth. In *Amazing Stories* (pp. 72–91). NY: Experimenter Publishing Company.

Halliday, M. A. K., & Webster, J. (2005). *Computational and quantitative studies* (Vol. 6). New York;London: Continuum.

Hammond, A., Brooke, J., & Hirst, G. (2013). A tale of two cultures: Bringing literary analysis and computational linguistics together. In *Proceedings of the Second Workshop on Computational Linguistics for Literature*. Atlanta, GA: Association for Computational Linguistics.

Hoover, D. L. (2013). *Textual analysis*. MLA Commons. https://doi.org/10.1632/lsda.2013.3

Irfan, R., King, C. K., Grages, D., Ewen, S., Khan, S. U., Madani, S. A., … Li, H. (2015). *A survey on text mining in social networks*. (Vol. 30, pp. 157–170). The Knowledge Engineering Review, 2.

Jones, K. E., Patel, N. G., Levy, M. A., Storeygard, A., Balk, D., Gittleman, J. L., & Daszak, P. (2008). Global trends in emerging infectious diseases. *Nature, 451*(7181), 990–993. https://doi.org/10.1038/nature06536

Kaal, B., Maks, I., & Elfrinkhof, A.v. (2014). *From text to Political Positions. Text analysis across disciplines*. Amsterdam/Philadelphia: John Benjamins.

Kaftan, V. (2012). Lion dance. *Asimov's Sci. Fiction, 36*(10–11), 87–95.

Knowles, T. W., II (1990). Curtain Call. In *New Destinies* (pp. 73–92). NY: Baen Publishing Enterprises.

Lambert Pharmacal. (1945). Get after that cold in the 1st round. In *Astounding Science Fiction* (p. 2). New Jersey: Street & Smith.

Leinster, M. (1961a). Pariah Planet. In *Amazing Stories* (pp. 6–92). Chicago: Ziff-Davis.

Leinster, M. (1961b). Doctor. In *Galaxy Magazine* (pp. 168–192). NY: Galaxy Publishing Corporation.

MacLean, K. (1950). Contagion. *Galaxy Mag., 1*(1), 114–140.

McFadden, S. (2015). Dystopian stories used to reflect our anxieties. Now they reflect our reality. In *The Guardian*. Manchester: Guardian Media Group.

Menadue, C. B. (2017). Trysts tropiques: The torrid jungles of science fiction. *eTropic, 16*(1), 125–140. https://doi.org/10.25120/etropic.16.1.2017.3570

Menadue, C. B. (2018a). Cities in flight: A descriptive examination of the tropical city imagined in twentieth-century science fiction cover art. *eTropic, 17*(2), 62–82. https://doi.org/10.25120/etropic.17.2.2018.3658

Menadue, C. B. (2018b). Hubbard bubble, dianetics trouble: An evaluation of the representations of dianetics and scientology in science fiction magazines from 1949 to 1999. *SAGE Open, 8*(4). https://doi.org/10.1177/2158244018807572, 2158244018807572.

Menadue, C. B. (2019a). George Miller's mad max (1979-2015) and Ryan Griffen's cleverman (2016-2017) - Australian science fiction. In J. Fennell (Ed.), Sci-fi a companion. Oxford: Peter Lang.

Menadue, C. B. (2019b). @rts of d@rlcne55: A pilgrim's progress towards a humane digital methodology. *[Manuscript in preparation]*.

Menadue, C. B. (2020). *Science fictions, cultural facts: A digital humanities approach to a popular literature*. (PhD Thesis), James Cook University, Townsville, Australia. Retrieved from https://doi.org/10.25903/5ef01cb4754df.

Menadue, C. B., & Cheer, K. D. (2017). *Human culture and science fiction: A review of the literature, 1980-2016, 7*. SAGE Open. https://doi.org/10.1177/2158244017723690

Menadue, C. B., Giselsson, K., & Guez, D. (2019). An empirical revision of the definition of science fiction: It's all in the techne. *SageSubmissions*. *Preprint*. https://doi.org/10.31124/advance.9755561.v1

Menadue, C. B., & Jacups, S. (2018). Who reads science fiction and fantasy, and how do they feel about science? Preliminary findings from an online survey. *SAGE Open, 8*(2). https://doi.org/10.1177/2158244018780946, 2158244018780946.

Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., The Google Books Team, … Aiden, E. L. (2011). Quantitative analysis of culture using millions of digitized books. *Science, 331*(6014), 176–182.

Milerius, N. (2011). Visions of utopia and dystopia in cinema. the philosophical presuppositions of the banal genre. *Problemos, 79*, 81–89.

Miller, S. T., Contento, W. G., & Stephensen Payne, P. (2019). The science fiction, fantasy, & weird fiction magazine index, 24th August 2019 23 April 2020 http://www.philsp.com/homeville/sfi/0start.htm.

Milner, A. (2009). Changing the climate: The politics of dystopia. *Continuum: Journal of Media & Cultural Studies, 23*(6), 827–838. https://doi.org/10.1080/10304310903294754

Neube, R. (2011). Grandma said. *Asimov's Sci. Fiction, 35*, 70–80, 09.

Niven, L. (1979). The ringworld engineers. In *Galileo* (pp. 54–77). Boston MA: Vincent McCaffrey.

Obradović, D., Baumann, S., & Dengel, A. (2013). A social network analysis and mining methodology for the monitoring of specific domains in the blogosphere. *Soc. Network Analy. Mining, 3*(2), 221–232. https://doi.org/10.1007/s13278-012-0075-7

Pangborn, E. (1974). The company of glory. *Galaxy, 35*(8), 32–90.

Peirce, H. (1974). Unlimited warfare. In *Analog Science Fiction Science Fact* (pp. 150–159). NY: Davis Publications.

Penteado, M. P. (2015). From contagion to isolation: The insistent dystopian future in the rag doll plagues and sleep dealer. *Linguas & Letras, 16*(33), 149–164.

Pitkin, J. (2012). A Murmuration of Starlings. *Analog Science Fiction / Science Fact, 132*(6), 39–48.

Raphals, L. (2013). Alterity and alien contact in Lao she's martian dystopia, cat country. *Science-Fiction Studies, 40*, 73–85.

Raychaudhuri, S., Chang, J. T., Imam, F., & Altman, R. B. (2003). The computational analysis of scientific literature to define and recognize gene expression clusters. *Nucleic Acids Research, 31*(15), 4553–4560. https://doi.org/10.1093/nar/gkg636

Rhines, J. A. (2003). Blacks in dystopia: 1969-1971. *Futures, 35*(8), 869–881. https://doi.org/10.1016/s0016-3287(03)00041-7

Rodriguez-Esteban, R. (2009). Biomedical text mining and its applications. *PLoS Computational Biology, 5*(12).

Rodriguez-Esteban, R., & Bundschus, M. (2016). Text mining patents for biomedical knowledge. *Drug Discovery Today, 21*(6), 997–1002. https://doi.org/10.1016/j.drudis.2016.05.002

Scholes, R., & Latham, S. (2006). The changing profession: The rise of periodical studies. *PMLA: Publ. Modern Language Assoc. Am., 121*(2), 517.

Scholes, R., & Wulfman, C. (2010). *Modernism in the magazines: An introduction*. Yale University Press.

Schummer, J. (2005). Reading nano: The public interest in nanotechnology as reflected in purchase patterns of books. *Public Understanding of Science, 14*(2), 163–183. https://doi.org/10.1177/0963662505050111

Shelley, M. (2015). *The Last Man*. New York: Open Road Integrated Media, LLC.

Temple, W. F. (1953). Scientist and censor: Should they meddle?. In *Authentic Science Fiction Monthly* (pp. 30–31). London: Hamilton & Co.

Temple, W. F. (1958). War against darkness. In *Nebula* (pp. 43–96). Glasgow: Peter Hamilton.

Thomas, J., McNaught, J., & Ananiadou, S. (2011). Applications of text mining within systematic reviews. *Research Synthesis Methods, 2*(1), 1–14. https://doi.org/10.1002/jrsm.27

Thompson, P., Batista-Navarro, R. T., Kontonatsios, G., Carter, J., Toon, E., McNaught, J., … Ananladou, S. (2016). Text mining the history of medicine. *PloS One*. https://doi.org/10.1371/journal.pone.0144717

Tilton, L. (2000). The Enclave. *Asimov's Sci. Fiction, 24*(9), 100–136.

Wang, S. H., Ding, Y. J., Zhao, W. Z., Huang, Y. H., Perkins, R., Zou, W, & Chen, J. J. (2016). Text mining for identifying topics in the literatures about adolescent substance use and depression. *BMC Public Health, 16*(1), 279. https://doi.org/10.1186/s12889-016-2932-1

Zweigenbaum, P., Demner-Fushman, D., Yu, H., & Cohen, K. B. (2007). Frontiers of biomedical text mining: Current progress. *Briefings in Bioinformatics, 8*(5), 358–375.