

A Computational Approach to Urban Space in Science Fiction

Federica Bologna

Federica Bologna, Università di Bologna

Peer-Reviewer: Matthew Erlin

Dataverse DOI: 10.7910/DVN/VXEK7A

Journal ISSN: 2371-4549

ABSTRACT

This study analyzes the presence of urban space in 20th century science fiction in English using computational methods. Three theoretical approaches are used to model urban space as a measurable feature. First, urban space is formalized as a topic. LDA topic modeling is used to retrieve the urban topic from the corpus and estimate its presence in each book. Secondly, urban space is formalized as the sum of the linguistic fragments that form a setting. A list of urban terms is created and their frequency is measured for each novel. Lastly, cityspace is formalized as the number of references to urban locations. Textual Geographies' geographic data was used to measure the presence of named urban locations in each book. The results of these approaches all point to similar conclusions. A low presence of urban space is found in science fiction compared to general fiction, alongside a historical trend. Urban presence in science fiction is greater at the beginning of the 20th century, declines in the 30s and 40s, and successively increases in the 50s. No such dip is present in other types of fiction across the twentieth century.

The goal of this study is to measure and analyze the presence of urban space in 20th century science fiction in English using computational methods.¹ Urban space is modeled using three different approaches. This study's results demonstrate science fiction's significant lack of attention to urban space in comparison to other works of fiction. They also provide strong evidence of a peculiar historical trend in urban attention. In science fiction, high attention to urban space at the beginning of the century is followed by a decline between the world wars and a rise in the 50s. In other types of fiction, by contrast, we observe a steady increase in urban attention over time.

Urban space and science fiction

Many have argued for space's relevance for the investigation of social life. Both Henri Lefebvre and Edward W. Soja affirm that it is in the production of space that

all social relations, “whether they are linked to class, family, community, market, [or] state power” are made concrete and material;² Michel Foucault claims that the spatial dimension enables one to unveil the workings of power.³ Moreover, Lefebvre believes that society’s development was possible only because of the creation of urban society.⁴ Thus, investigating urban life brings insights on the nature and transformations of a society.

On the other hand, as Fredric Jameson famously argues,⁵ science fiction provides readers with a multiplicity of mock futures in which our complex and unfathomable present time is crystallized. This genre enables the deconstruction and defamiliarization of our experience of our time. Its concern with alternative worlds sparks reflections on contemporary issues. Furthermore, many have used science fiction to investigate questions related to space and urbanism. Ursula K. Heise mentions well-known sci-fi novels to demonstrate the development of a global spatial consciousness in literature.⁶ Carl Abbott studies sci-fi cities to enrich the contemporary discussion on how cities should be built.⁷

Moreover, several scholars have pointed out the relevance and extensive presence of futuristic metropolises in science fiction. Under the entry “Cities” of *The Encyclopedia of Science Fiction*, Brian M. Stableford, David Langford and John Clute organize numerous sci-fi works in different categories depending on their vision of future cityscapes. They conclude:

In both sf writing and sf art, the city is one of the most important recurrent images, and carries with it one of the richest, densest clusters of associations to be found in the whole sf iconography.⁸

Similarly, before departing on his exploratory survey of the cities represented in science fiction literature from the mid of the 20th century onward in “The Science Fiction City”, John Dean not only states that cities bring dramatic force to sci-fi narrations, but also that urban spaces are numerous in the genre:

The presence of a city in a work of science fiction heightens the drama of living in the future. [...] The city plays an important part in a wide range of modern science fiction literature [...] To repeat: in sf cities are legion.⁹

For his part, Jameson argues that not only cities, but space in general plays an instrumental role in sci-fi stories. Nonetheless, while discussing the surprising narrative powers at play in science fiction, he confirms the presence of several urban figurations in the genre, especially futuristic ones:

Many SF cityscapes and utopias seem to participate in this curious paradox: that what signals the constructed, invented, artificial nature of SF as a genre - the palpable fact that an author has strained her or his invention to contrive some near or far future city [...] is here an unexpected source of strength.¹⁰

The great number of cities in science fiction and the genre's lasting concern with cityscapes seem to be generally accepted by the existing scholarship on the matter. In addition, more recent works now draw upon this belief to build specific analyses and arguments regarding urban space in sci-fi literature.¹¹

Conceptualizing space in text

In the context of the vast literature about the conceptualization of space, the present analysis locates itself in the realm of Edward Soja's theories. Specifically, it aims to investigate what Soja refers to as First space: the perceived, material, empirical and socially produced space constructed in the novels under study.¹² Being only an initial exploration of space, this study does not take into consideration the two other modes of spatiality identified by Soja - Second and Third space. Nonetheless, detecting First space in sci-fi novels provides us with insights on these two other modes of spatiality as well.

To model the concept of urban space as a measurable textual feature, we employed three approaches. First, urban space is formalized as a topic, a coherent semantic entity that crosses different genres. The best technique to retrieve topics from a large corpus of texts is LDA topic modeling. This technique provides estimates of each topic's presence in each book of the collection. Therefore, LDA is performed on the collection of sci-fi novels and a topic is chosen as the most concerned with cityscapes. The topic's presence in a novel is equivalent, for analytical purposes, to the presence of urban space in said novel.

Secondly, urban space is formalized as the sum of linguistic fragments that create a setting when combined in a text. The best technique to measure this type of textual feature is to create a list of urban terms and calculate their frequency for each novel in the collection. The number of occurrences of urban terms in each novel is then a measure of the presence of urban space in that novel.

Lastly, cityspace is formalized as the number of references to existing urban geographical locations. The best technique to measure the presence of named urban locations in each book is to use the data available through the Textual Geographies project.¹³ Thus, the number of names of urban locations in a book signals the level of presence of urban space in that book.

All these features provide different perspectives on urban space. When combined, they create a more comprehensive understanding of the presence of cityspace in science fiction. Since each of these approaches is put in practice using a different technique, the results too are heterogenous in form. However, all three types of data point to similar conclusions. They show a low presence of urban space in science fiction compared to general fiction, alongside a historical trend. Urban presence in science fiction is greater at the beginning of the 20th century, declines in the 30s and 40s, and successively increases in the 50s. No such dip is present in other types of fiction across the twentieth century.

The corpus

The corpus is built by combining the nominations for the Nebula, the Hugo, and the Retro-Hugo awards. The Nebula and the Hugo have been awarded annually for science-fiction and fantasy works of the previous year since 1966 and 1955 respectively. Whereas, the Retro-Hugo awards were awarded in 1996, 2001, 2004, 2014, 2016, 2018, 2019, and 2020 for works published in 1946, 1951, 1954, 1939, 1941, 1943, 1944, and 1945 respectively. To fill in the decades preceding the establishment of these prizes, I include the sci-fi novels used by Ted Underwood in his book *Distant Horizons*.¹⁴ These sources form an initial corpus of 726 novels, of which only 330 (45%) are found in the HathiTrust collection. Among these, 234 are prize-nominated books and 96 derive from Underwood's list.

Combining these sources ensures that the novels in the corpus constitute one broadly plausible and compelling representation (among many) of the genre. Indeed, the Nebula is given by science-fiction writers, whereas the Hugo and Retro-Hugo are awarded by fans. The genre designations of the individual works are not manually checked, in order to avoid imposing my own biased and fixed definition of science-fiction. The concept of genre has always evaded delineation and critics are yet to reach an agreement on which are the distinctive features of long-standing genres such as sci-fi.¹⁵ Consequently, I decide to trust the definition that spontaneously arises from the preferences expressed by the science-fiction community, here represented in the form of the nominations for the three aforementioned awards.

Urban space as a topic

Topic modeling is performed on individual pages using the scikit-learn Latent Dirichlet allocation (LDA) library.¹⁶ The model is fine-tuned by running the algorithm several times and evaluating each result depending on: how even is the topics' distribution in the topic space; the presence of a topic about urban space; and which novels show the highest percentage of this topic.

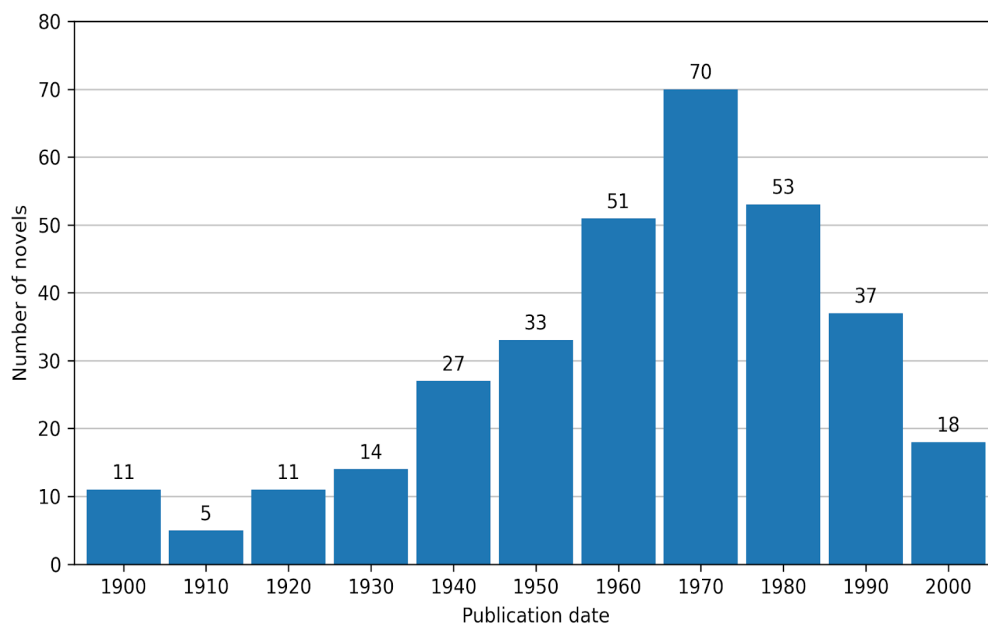


Figure 1: Distribution of the novels in the science fiction corpus over time

The best model, in terms of even distribution of topics in the topic space and inner semantic coherence of the topics themselves, is obtained by setting the number of topics at 80, the number of features (or unique words) at 10000, and at 30% maximum document frequency for each word. Looking at the model visualization, topic 19's top words are closely related to urban spaces, both ancient and modern: city, street, building, crowd, traffic, park, driver, police, alley, aisle, sidewalk, pavement, plaza, tar, pedestrian, metropolis, outskirt. Other topics in the model are not as strikingly urban; for example, number 22 seems linked to nature and explorations: mile, rock, mountain, range, road, flower, cliff, far, reach, valley, hill. Therefore, topic 19 was chosen as the most suited for this analysis.

The books that contain a high proportion of topic 19 form a coherent if somewhat unexpected bunch. Indeed, several of these novels depict ancient-looking cities, challenging the conventional assumption that science fiction is concerned with envisioning the city of the future. For example, Edgar Rice Burroughs' *The Chessmen of Mars* (1922), *A Princess of Mars* (1912), and *Carson of Venus* (1939) are set on fictional planets structured in medieval-looking city-states ruled by royal families, Williamson's *The Reign of Wizardry* (1940) is set in Crete's ancient capital Knossos, and *The Years of Rice and Salt* (2002) by Kim Stanley Robinson travels through many ancient and modern cities. Other books are set in cities, but do not portray urban life: Christopher Priest's *The Inverted World* (1974) is mostly set outside the city's perimeter; Ray Cummings' *The Shadow Girl* (1929) tells of time travelling through different epochs in New York City; Paolo Bacigalupi's *The Windup Girl* (2009) recounts futuristic Bangkok's fight against climate change and bioterrorism. However, there are also a few truly metropolitan stories, such as E. L. Doctorow's *Ragtime* (1975) and George S. Schuyler's *Black No More* (1931), both of which tell of social life and racial discrimination in New York City; or John Brunner's *The Squares of The City* (1965) that discusses issues like segregation, inequality and political power in a fictional metropolis. A peculiar case is Jack London's *The Scarlet Plague* (1912), which is set in San Francisco after an epidemic has depopulated the world. However, the flourishing metropolitan life preceding the catastrophe is re-lived through the protagonist's memories. Both danger and nostalgia are connected to cities in the book:

I've seen this beach alive with men, women, and children on a pleasant Sunday. And there weren't any bears to eat them up, either. And right up there on the cliff was a big restaurant where you could get anything you wanted to eat. Four million people lived in San Francisco then. And now, in the whole city and county there aren't forty all told¹⁷

As noted above, many of these novels depict ancient-looking cities. Although some of them stand on the thin and foggy line separating science fiction and fantasy, they all share an interest for scientific elements and have been recognized as worthy exemplars of the genre by the science-fiction community delivering the Nebula and Hugo awards. Therefore, these results challenge the shared belief¹⁸ that science fiction represents futuristic cities, and reveal how difficult it is to detect and distinguish different types of urban spaces linguistically, since they share the same keywords. They raise questions regarding which type of urban spaces are to be defined as unmistakably “urban” and which urbanism should be analyzed in literary criticism. Nonetheless, it is possible that topic modeling might not be successful at retrieving modern urban settings. Detecting urban space using two alternative techniques – token counts and geolocations – helps us further understand the presence, or absence, of cityspaces in sci-fi.

Urban space as token-counts

In “Towards a theory of space in narrative”, Zoran claims that, due to the linear and temporal quality of language, the latter cannot grasp the totality and simultaneity of space’s existence.¹⁹ In verbal narration, space is presented in segments that the reader has to combine in order to reconstruct the narrated world. Ruth Ronen seems to agree with Zoran’s theory, as she too argues that fictional spaces are “semantic constructs” built by integrating a series of linguistic elements.²⁰ Interestingly, she argues that places can be denoted using both direct and indirect identification. For example, a room can be identified by both its common noun and its objects.²¹ In “Toward a Computational Archaeology of Fictional Space”, Dennis Yi Tenen too argues that a “space can be understood through things”.²² All these theoretical approaches to space in text seem to agree that, when the analysis is not interested in the narrated space’s topography, space can be conceptualized as the sum of its linguistic components. As a result, for the purposes of this study, space in text is

modeled as the sum of terms that indicate units and objects found in urban landscapes.

To achieve this end, it is necessary to create a list of “urban” terms whose counts are to be retrieved from the novels. The terms are manually collected from three sources: Raymond Williams’ *The Country and The City*;²³ Will Eisner’s *New York: The Big City*;²⁴ and the Macmillan Dictionary of English Language.²⁵ This unconventionally diverse set of sources provides a variety of terms that many sources of one typology alone could have not forged: Williams’ work brings scholarly and refined idioms; Eisner’s comics add more popular and slang expressions; and lastly, Macmillan supplies standardized words. The terms are listed in their tokenized and lemmatized form.²⁶ The selected tokens depict the type of cityspace that is of most interest to my research: a densely populated and technologically advanced space. The full list of 77 lemmas can be found in Appendix 1. In order to evaluate the presence of these terms in science fiction, a secondary corpus is needed as a benchmark. For this reason, a corpus that is fifty times the size of the first one is sampled at random from HathiTrust fiction.^{27,28} This contrast corpus contains works from different genres and

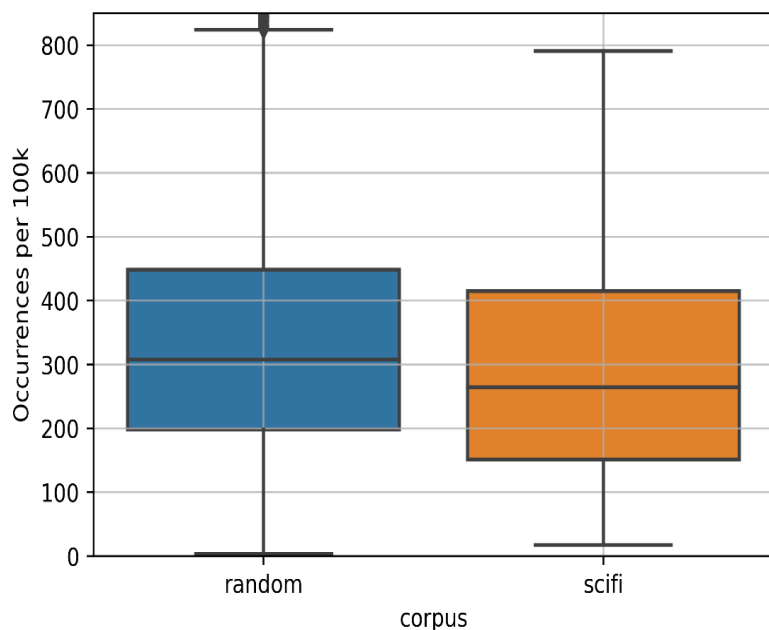


Figure 2: boxplots representing the distribution of occurrences of urban locations per 100,000 words in each corpus.

is constructed to follow the same distribution of novels by date of publication as the science fiction corpus.

The first feature to analyze is the number of occurrences of urban terms per 100,000 words in a novel. In the box-and-whisker plots (figure 2), science fiction novels show fewer urban words than general fiction books. To assess whether the corpora are statistically different, three tests are used: the Welch's t-test for independent samples; the Mann-Whitney U; and the chi-squared test. Since the p-value is significantly lower than 0.05 for all three tests, a statistical difference is found in the number of occurrences of urban terms per novel among the two corpora.

In addition, the number of occurrences per decade are studied. The occurrences at the book level are summed and grouped by decade, and a box-and-whisker plot is created (figure 3 and 4). Most outliers are cut out of the visualization below so as to have a clearer view of the boxes. The boxplot for the science fiction corpus shows a peak in the 1910s, followed by a sudden decrease in the 1930s and the 1940s. Interestingly, the mean occurrence of urban terms at the end of the century is lower than that at the beginning of the century, going from around 350 words per 100,000 in the 1900, to less than 300 in the 2000s. On the other hand, in the random corpus, the presence of urban terms steadily grows over the century. It has to be kept in mind that in the corpus there are only 5 novels from the 1910s and they may not be representative of all science fiction works published in that decade. Nonetheless, it is still worth some reflection that those few books show such a high presence of urban tokens.

The second feature is the number of occurrences per 100,000 words of an urban term in each book of both collections. The chi-square test is performed on all the lemmas in the collection. To reduce the amount of processing time, only 10% of the random corpus is considered. All 59 urban lemmas are found to be distributed differently between the corpora, meaning that the pattern of occurrence of one term in one collection is different from the pattern of occurrence of the same term in the opposing collection. For instance, the minimum, the median and the maximum value in the distribution of the term "city" in the random corpus are around 1, 30 and 125 respectively. Whereas, the same values in the distribution of the same term in the sci-fi corpus are around 1, 45, 180, respectively. Thus, in one book of the random

collection “city” occurs only once, and that happens to be the minimum number of occurrences of the term in a book of the random corpus. To establish whether some urban lemmas are more different in their distributions than non-urban lemmas, the p-value of urban lemmas is then compared to that of lemmas that have a similar

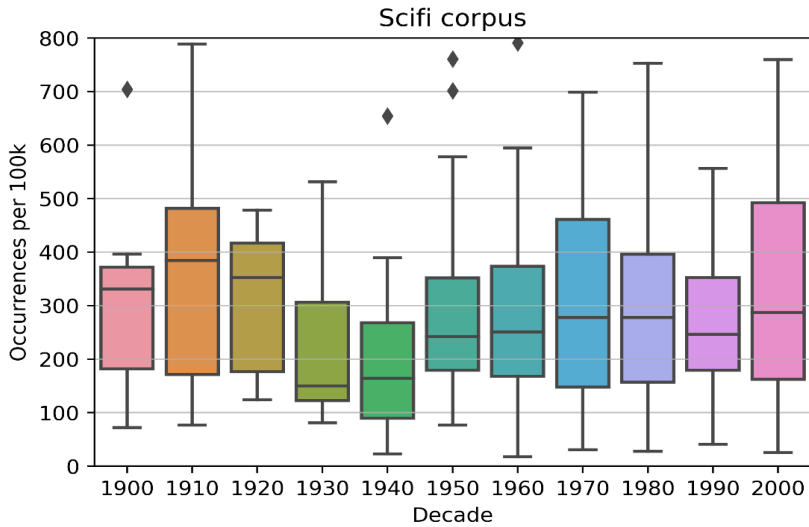


Figure 3: box-and-whiskers plot representing the distribution of urban terms at the decade level.

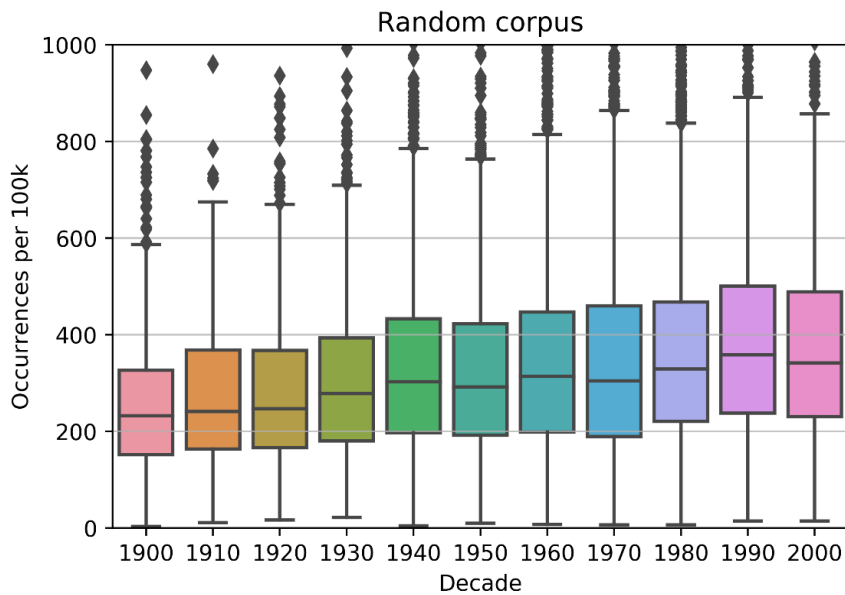


Figure 4: box-and-whiskers plot representing the distribution of urban terms at the decade level.

number of occurrences per 100,000 words. Only the distributions of “city” in the two corpora were significantly and statistically more different than the distributions of words with a similar number of occurrences. Looking at the distributions of urban lemmas in both collections, the lemmas with the lowest p-value correspond to the most frequently occurring ones. After all, the chi-squared test prioritizes small differences between big values over big differences between small values. On the other hand, “suburb” and “metropolitan” were rare in both corpora but were found to be significantly different because of their peculiar distributions. As a result, urban tokens collectively were not found more different than other tokens of similar frequency across the two collections, nor were their differences caused by their higher presence in one of the two corpora.

To evaluate the effectiveness of this model at detecting urban space, it is now useful to review the highest scoring books in terms of occurrences of urban tokens. With regard to the science fiction corpus, this model based on token counts provides results that are comparable to those obtained implementing LDA topic modeling. If we define urban space as the technologically advanced metropolis, only nine of the top twenty novels for number of occurrences of urban terms are set in such areas. Similarly, of the top twenty novels for proportion of topic 19, only nine are set in highly populated modern cities. In addition, the two rankings greatly overlap, as they both include the following titles: Priest’s *The Inverted World* (1974); Brunner’s *The Squares of The City* (1965); Cummings’ *The Shadow Girl* (1929); Doctorow’s *Ragtime* (1975); Burroughs’ *A Princess of Mars* (1912); Bacigalupi’s *The Windup Girl* (2009); Robinson’s *The Years of Rice and Salt* (2002); Burroughs’ *Carson of Venus* (1939) and Silverberg’s *Lord Valentine’s Castle* (1980). Of the remaining books in the tokens’ ranking, two are set in rural areas: Ray Bradbury’s *Something Wicked This Way Comes* (1962) is set in a fictional small midwestern town surrounded by prairies and Cyril M. Kornbluth’s *Not This August* (1955) is mostly set in rural Chiunga County, NY. Even though big cities are occasionally mentioned in the latter, they only appear in the background. Some books in the list are set in worldly metropolises, like Walter Tevis’ *Mockingbird* (1980) and George R. Stewart’s *Earth Abides* (1949). The first describes NYC with a sense of decadence and despair. The second narrates the collapse of civilization due to the outbreak of a pandemic and describes the cities’ decadent conditions with nostalgia. Other books are set in invented cities, like Bradbury’s famous *Fahrenheit 451* (1953) and Leigh

Brackett's *The Long Tomorrow* (1955). Both *Fahrenheit 451*'s "The City" and *The Long Tomorrow*'s Bartorstown are symbols of technological advancement. In the first one, technology is used for violence and oppression, and the city is juxtaposed to a restorative and welcoming wilderness. In the second, by contrast, science enables the betterment of society and the city is set in contrast to underdeveloped rural villages.

On the other hand, in the random corpus, the top three books for number of occurrences of urban tokens are all children's books. The repetition of few urban terms and the brevity of these narrations cause the number of occurrences per 100,000 words to spike. The books identified do pay significant attention to urban spaces. Jerrold Beim's *Andy And the School Bus* (1947) tells the story of a child who lives in the countryside and wants to take the bus to go to school in the city. *Kermit, Save the Swamp!* (1992) by Richard Chevat recounts Kermit's efforts to save his native swamp from being transformed into a shopping mall and amusement park, warning the reader about the disruptive forces of urbanization. Nonetheless, these books only contain two or three of all the urban lemmas in the list. Therefore, a second ranking is created consisting of books that contain at least twenty different urban terms. All the top novels in this list are set in modern cities. Marina Snow's *Ailanthus Park* (2008) is set in Sacramento and Stephen Peter's *The Park Is Mine* (1981) is set in New York City. In Jim Crace's *Arcadia*, urbanization is once again portrayed as a corruptive force as its rich protagonist wants to tear down the city's local market to build a shopping center. The collection of short stories *Marcovaldo* by Italo Calvino (1963) tells of a man born in the countryside living in an industrialized city, connecting urban areas to environmental issues, poverty and consumerism. Pier Paolo Pasolini's *Stories from the City of God* (1995) is a mosaic and heterogeneous representation of the cultural and social life of the outskirts of Rome in the 50s and 60s.

Looking at these titles, it is clear that the urban tokens' model effectively detects books set in modern metropolises from the general corpus. Consequently, the lack of urban novels among the sci-fi rankings is not due to the ineffectiveness of the model, but rather seems to result from the nature of science fiction works themselves. After all, all the books in the sci-fi rankings do pay *some* attention to *some* urban spaces. Their stories may take place in metropolises turned to ruins and pre-modern

city-states or move across a variety of landscapes. Urban spaces are not completely absent from these novels - which caused the model to retrieve novels that were not set in populated modern cities. They are recollected in memory or used as backdrops for the main story. As a result, urban environments end up coexisting with rural ones, setting up a contrast between cities and countryside. In the retrieved novels, this conflict appears to be instrumental in criticizing urbanism's corrupting and destructive forces, but also in praising its prosperity in terms of culture, wellness and economy. After all, the city-country antinomy comes from a long tradition of pastoral texts and is not uncommon in science fiction either.^{23,8} The two spaces have been in relentless dialogue in Western culture for a long time.

This section's conclusions can be summarized into four points. First, science fiction novels generally include fewer urban tokens than do other types of fiction. Secondly, science fiction shows a high number of urban tokens at the beginning of the century, a decrease in the 1930s and 40s and a rise in the 1950s. Third, urban tokens are not exceptionally different in the distribution of their occurrences across the two corpora. Lastly, reviewing the book ranking for the general fiction corpus reveals that the token counts' model is effective at retrieving books concerned with urban spaces.

Urban space as geographic locations

Even though many science fiction works locate their stories in fictional settings or use fictional names to address them, science fiction's fascination with existing cities is undeniable, as several works attempt to capture their fate and possible futures (e.g. *Neuromancer's* (1984) North-Eastern cities, the Bay Area of *The Scarlet Plague* (1912) and *Earth Abides* (1949), the New York City of countless works). Therefore, in order to accurately measure urban spaces' presence in the genre, it is essential to also take into account those instances where wordly locations are mentioned, by measuring the number of occurrences of urban geolocations alongside topic modeling and urban terms' occurrences.

The geographic locations contained in both corpora are easily accessed thanks to the Textual Geographies project, which makes available the geographic data related to the HathiTrust digital library.^{13,29} For the purposes of this study, I define urban

locations as those associated with a third-level administrative area or lower (including all localities).³⁰

The first feature to analyze is the number of occurrences of urban locations per 100,000 words in a novel. In the box-and-whisker plots (figure 5), general fiction novels show a higher number of occurrences. In order to assess whether there is a statistically relevant difference in the distribution of occurrences of urban locations per each novel, the Welch's t-test, the Mann Whitney U test and the chi-squared test are performed on the sci-fi and general corpora. A statistical difference in the number of urban locations per 100,000 words in a book is found across the corpora, and we can thus conclude that science fiction novels generally contain fewer named urban places.

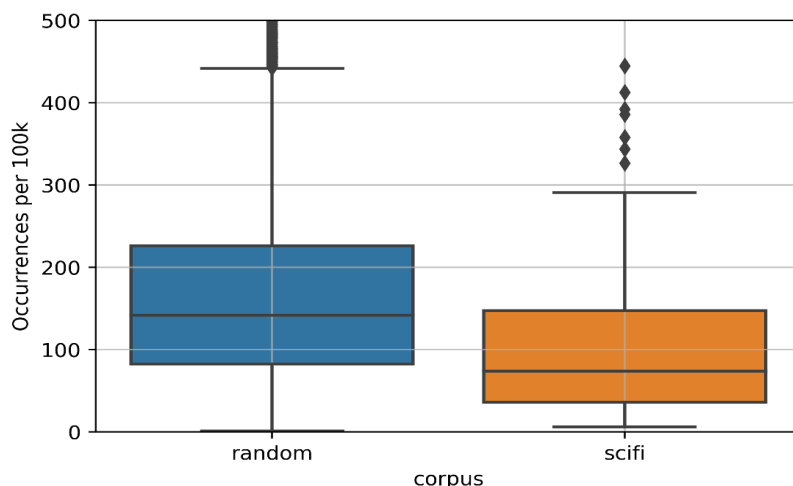


Figure 5: boxplots representing the distribution of occurrences of urban locations per 100,000 words in each corpus.

Subsequently, the transformation in the number of occurrences of urban locations over time is taken in consideration. All the occurrences for each novel are summed and grouped by decade; the resulting box-and-whisker plots are shown in figure 6 and 7. The decade with the highest median value for the science fiction corpus is the 1910s, followed by a decrease in the 1930s and 1940s. In analyzing the data, it is important to keep in mind that the first few decades of the century are the least represented in the collection. Nonetheless, it is impressive that a small number of books show such an intense interest in cityspace. This trend mirrors the one

described above for the number of occurrences of urban terms per 100,000 words. On the other hand, in the general fiction corpus, the number of urban locations per 100,000 words remains stable throughout the analyzed period of time.

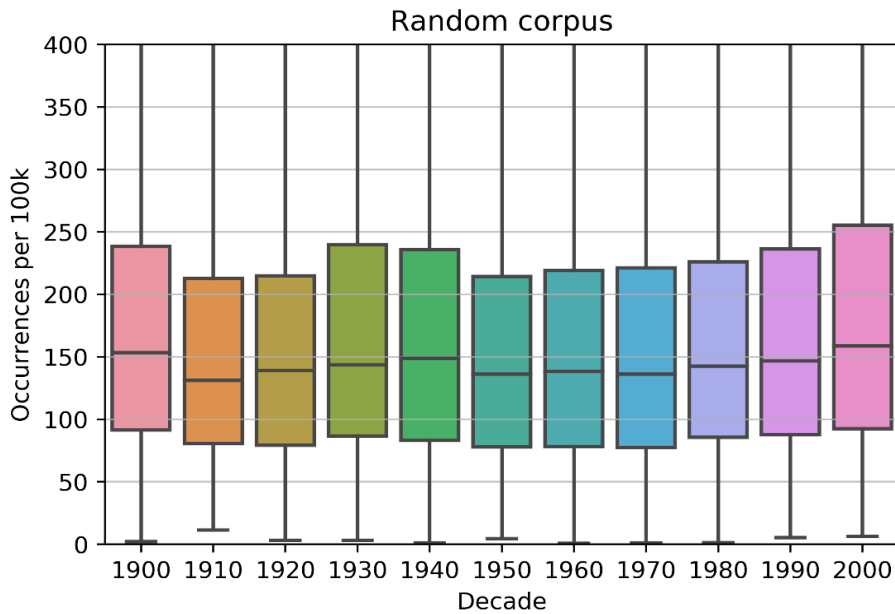
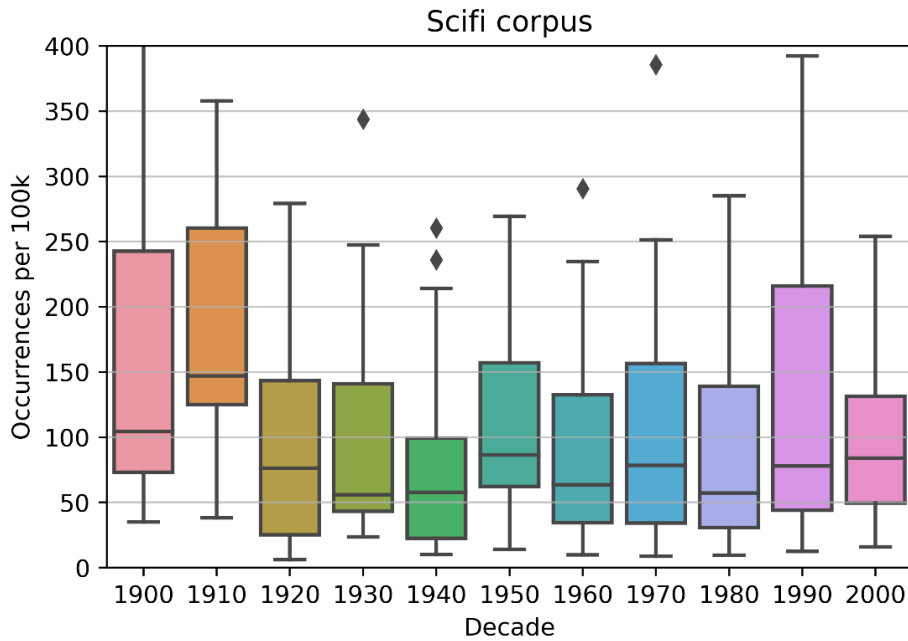


Figure 6&7: box-and-whiskers plot representing the distribution of the occurrences of urban locations at the decade level.

Lastly, simple linear regression is performed on the occurrences of urban locations and the occurrences of urban terms for each novel to test the possibility of a correlation between these two features. Specifically, the correlation between urban geodata and urban tokens is first tested in both corpora. Then the correlation is also tested in the science fiction corpus only. A residual plot is created to ensure that linear regression is a good fit for the data. Once that is confirmed, the data is passed through a logarithmic function so that the plot and regression line appear more clearly (figure 8). The regression line is slightly skewed. In the two corpora, Pearson's coefficient and the coefficient of determination for the two variables are 0.17 and 0.04 respectively. Whereas, in the sci-fi corpus only, the two coefficients are 0.23 and 0.05. Therefore, the two variables seem to slightly move together. However, their relationship can hardly explain the variation of each of them. Three main conclusions can be drawn. First, science fiction novels generally contain fewer occurrences of urban geographical locations. Secondly, science fiction shows a peak in urban location in the 1910s, followed by a decrease in the 1930s. Lastly, no strong correlation was found between number of occurrences of urban terms and number of occurrences of urban locations.

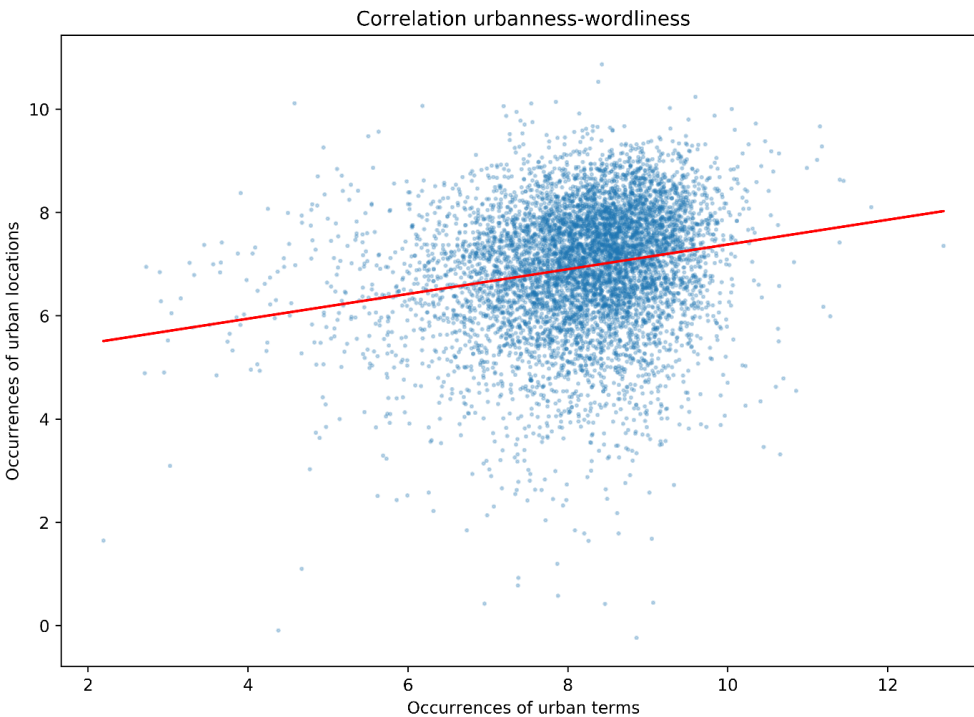


Figure 8: scatterplot of the occurrences' data points with the regression line

Conclusions

Lack of urban novels and urban space in science fiction

Among the top twenty science fiction novels with the highest presence of the city topic or the highest number of occurrences of urban tokens, only a few are set in densely populated and technologically advanced cities. There are two possible causes for these results: 1.) there may be fewer science fiction novels set in urban areas than critics have generally believed, or 2.) sci-fi novels set in metropolises may render urban space via idiosyncratic or neologistic language.

In order to address the first point, I turned to Abbott's *Imagining Urban Futures*.⁷ Of the many novels he mentions for their representation of urban areas, only two titles appear in the rankings summarized above: Bacigalupi's *The Windup Girl* (2009) and Priest's *The Inverted World* (1974). Surprisingly, most of Abbott's novels were not present in the corpus, as they had not been nominated for any of the three awards aforementioned. Nonetheless, other novels mentioned by Abbott and contained in the corpus did not appear in the rankings: in Isaac Asimov's Foundation series, Trantor, the capital of the Galactic Empire, is a technological advanced metropolis as vast as a planet; and William Gibson's *Neuromancer* (1984) is set in "BAMA, the Sprawl, the Boston-Atlanta Metropolitan Axis".³¹ Even though these novels were in the corpus, none was among the highest scoring ones for presence of urban spaces. This might be due to the lack of material descriptions of cityspace in these novels.

To address the second cause, I turned to the number of occurrences of urban tokens and urban locations. With regard to both of these measures, science fiction shows lower numbers than general fiction. Even the sci-fi novels that have the highest number of occurrences of urban tokens in the genre's corpus do not describe their settings as thoroughly as general fiction works do. The top fifty general fiction books for occurrences of urban terms show twice as many occurrences as the top science fiction novels. Looking at the books reviewed above, it seems that the city is both a symbol of cultural and technological advancement and a site of decadence and destruction in sci-fi novels. In Jack London's *The Scarlet Plague* (1912), the disease that will depopulate the world first appears and spreads in big cities. The epidemic

recounted in George R. Stewart's *Earth Abides* (1949) follows the same dynamic. However, the protagonists of both books reminisce about the cultural richness and wellness that cities and progress offered. In Leigh Brackett's *The Long Tomorrow* (1955), cities and the scientific knowledge they nurture are seen as the cause for the outbreak of a nuclear war. However, the protagonists are fascinated by technology and decide to live in a forbidden city where scientific research is allowed. Spinrad's *The Iron Dream* (1972), Bacigalupi's *The Windup Girl* (2009) and Bradbury's *Fahrenheit 451* (1953) are all set in capitals at the rise of an authoritarian regime. Such totalitarianisms cause civil and international wars that almost wipe out humanity. In Kornbluth's *Not This August* (1955), cities are connected to a low quality of life and warfare. In these books, cityspace symbolizes both the best and worst of humankind's progress. Another group of texts is formed by Burroughs' Barsoom series, Silverberg's *Lord Valentine's Castle* (1980), Williamson's *The Reign of Wizardry* (1940) and Robinson's *The Years of Rice and Salt* (2002). They do not present dystopian visions, but here too cities are the centers of political and cultural power.

It appears that science fiction novels mention cityspace more because of its symbolic value than as a realistic setting. Cities are the ground where major events take place. Sci-fi novels are not focused on cityspace's social dynamics as much as works of general fiction are. Exceptions to this trend seem to be Brunner's *The Squares of The City* (1965) and Doctorow's *Ragtime* (1975). Here, the metropolis is not the birthplace of human destruction, but the place of enactment of social inequality, discrimination and violence.

Historical trend in urban attention

Science fiction shows a high presence of urban spaces at the beginning of the century, followed by a decrease between the 30s and 40s, and a consecutive increase in the 50s. These dates match major events in United States' history that transformed the face and soul of the country. First, the massive urbanization of the United States started in the 1870s and the nation reached urban-majority status between 1910 and 1920. This growth came almost to a halt between the 1930s and the 1940s.³² Additionally, the Immigration Act of 1924 limited the number of European immigrants and banned Asian and Indian immigration, almost forcing to a stop the

influx of immigrants that took place from the 1890s to 1920.^{33,34} Since the great urban areas of the Northeast did not receive as many immigrants after that point, they grew at a slower pace. Cities suffered another hit when the Great Depression struck the US and the whole world in 1929. After decades of migration from rural to urban areas, the trend was reversed as millions headed back to their family farms.³⁵ This was due to the collapse of the construction and manufacturing industries, which caused millions to lose their jobs, both high-paying and low-paying ones.³⁶

Looking at the science fiction novels published in the 1910s, not all of them have a high number of occurrences of urban terms. Conan Doyle's *The Lost World* (1912) is set in an unexplored part of the Amazon basin and *A Columbus of Space* (1911) by Garrett P. Serviss tells of an expedition in outer space. Both show a number of occurrences that is lower than the corpus' average. Nonetheless, J. D. Beresford's *Goslings* (1913) and Jack London's *The Scarlet Plague* (1912) show an above average number of occurrences even though no functioning city is represented. Both recount the spreading of a disease which decimates the world population. Furthermore, Burrough's *A Princess of Mars* (1912) is the book with the highest number of occurrences in the Barsoom series. Looking at these titles, it seems that in the 1910s science fiction cannot imagine a better form of human settlement than the city. In contrast, in the 1930s, novels are concerned with the future of humanity in its ethical and social aspects, but their number of occurrences of urban tokens is below the corpus' average: Olaf Stapledon's *Last and First Men* (1930) tells the history of humankind over two billion years starting from the early 21st century; Aldous Huxley's *Brave New World* (1932) narrates about humankind in the year AF (After Ford) 632 (AD 2540 in the Gregorian calendar); H. G. Wells' *Star Begotten* (1937) recounts how a group of men suspects Martians are genetically modifying the human race.

An intriguing pattern seems to emerge from these works: science fiction from the 1930s is not just concerned with a nation's fate, but with the fate of humankind. The events of these novels have global consequences and present a global consciousness. After all, by the mid-1930s, the Great War had taken place, totalitarianisms had risen all over Europe, and the atomic bomb had become part of the public's imagination. It thus comes as no surprise that the science fiction of the 1930s is worried about humans' survival. Despite the drought starving the countryside and the economic

crisis taking over cities, science fiction novels appear not to address such prosaic issues. Such subjects are being covered – either realistically or allegorically – by William Faulkner, John Dos Passos, Pearl S. Buck, Margaret Mitchell, John Steinbeck, Nathaniel West, Zora Neale Hurston, and others. On the other hand, science fiction seems to answer the public’s anxieties by either speculating on potential global events or providing an escape. Two more books of the 1930s are an example of the first trend: Sinclair Lewis’ *It Can’t Happen Here* (1935) speculates on the consequences of the rise of a despotic leader to the position of President of the United States; and J. B. Priestley’s *The Doomsday Men* (1938) tells of three brothers’ plan to blow up the entire world and three strangers’ inability to stop them. An example of the second trend is T. H. White’s *The Sword in the Stone* (1938), a fantasy novel revisiting the legend of King Arthur. An exception to this trend is Schuyler’s *Black No More* (1931), which stood at the intersection of the Harlem Renaissance and science fiction and now belongs to Afrofuturism.

The genre partially regains interest in urban areas from the 50s onward. This increase in urban space might seem to challenge my previous assertion. Due to the beginning of the Cold War, the period that followed WWII is characterized by intense paranoia and anxiety. The tension between the United States and the Soviet Union, and the countries’ arms race, turns human extinction into an increasingly realistic threat in the minds of the American people.³⁷ At the same time, post-war wealth and economic growth gives rise to a feeling of optimism.³⁸ More and more families come out of poverty into a growing middle class that can sustain a high standard of living. This new purchasing power combined with federal aid and the development of suburbs that offer cheap standardized houses for the masses results in a housing boom and suburbanization.³⁹ This contradiction between enthusiasm for an apparently never-ending urban and economic expansion, and worry of mass extinction likely comes to be rooted in Americans’ everyday life. It also appears to permeate the science fiction of that time. In these novels, nuclear wars and urban areas come to coexist. In *Not This August* (1955), cities survive the conflict between the Western and Eastern blocks. In *The Long Tomorrow*, despite all American cities having been destroyed by a nuclear war, new towns and an underground city reflowerish in the war’s aftermath. In *Fahrenheit 451* (1953), Montag and the rebels plan to rebuild civilization from the City’s ruins after nuclear bombs have obliterated it. Lastly, *The Iron Dream* (1972)’s Heldhime has been rebuilt after a global nuclear

war had destroyed the previous civilization. As the possibility of human extinction is increasingly normalized, it seemingly ceases to exclude urban areas from sci-fi narration. Cities are depicted as the cause and aftermath of human annihilation, forming a cycle that sees no alternative, and urbanism is represented as a necessary condition for the reconstruction of civilization.

The results of this study have several implications for the investigation of urban space and of science fiction. Contrary to what has long been assumed, cities are rather unpopular in sci-fi. While I have speculated on the reasons both for that assumption and for the observed lack of urban orientation, there is obviously a need for additional work. The expectation of an urban orientation might be due to the influence that other media (film in particular) in the genre has exerted on the public. Furthermore, the lack of urban terms in novels that are considered to be distinctively urban raises questions as to what characterizes an urban novel. Which criteria should we use in order to classify urban novels? Is having an urban setting enough, or should that setting be thoroughly described? Moreover, it has often been assumed that urban spaces in sci-fi are always technological and vertical metropolises. In light of the genre's fascination for other kinds of urbanism as well, it is essential to reconsider this stereotypical vision of the science fiction city. More attention should be paid to other less-known and less-stereotypically-futuristic cityscapes. In addition, comparisons should be drawn among sci-fi works belonging to different decades, as science fiction transforms and re-adapts itself to new historical conditions over time. This approach might unveil shifts in the public's expectations toward the genre and how the latter has fulfilled the needs of its readership. I hope for this study to encourage other researchers and to help overcome the countless paralyzing biases that have long discredited this genre and its study.

Appendix 1

The following are the 77 lemmas that have been retrieved from the novels and counted in order to measure the occurrence of urban terms.

city	alley	tar
chinatown	traffic	homeless
highway	subway	uptown
intersection	gentrification	shantytown
skyscraper	neighborhood	backstreet
metropolitan	outskirt	theater
metropolis	avenue	bitumen
enclave	streetlight	commuter
freeway	bus	barrio
crossroad	crowd	slum
building	suburb	promenade
suburban	plaza	mall
megalopolis	ave	urban
ghetto	lamppost	commute
speedway	concrete	burbs
wayside	uptowner	road
park	borough	sidewalk
suburbanize	riverfront	market
conurbation	boulevard	urbanize
hood	hydrant	pedestrian
street	asphalt	driveway
lane	downtowner	curbside
metro	downtown	library
gentrify	shanty	urbanization
town	blvd	sewer
midtown	port	

References

1. I thank Matthew Wilkens (Cornell University), whose guidance and supervision has been of incommensurable value. I also thank Kristoffer L. Nielbo (Aarhus University) and Ross D. Kristensen-McLachlan (Aarhus University) and the researchers of the Center for Humanities Computing Aarhus for their instrumental help and support. Lastly, I thank Mads Rosendahl Thomsen (Aarhus University) for the hospitality and having created a stimulating environment at Aarhus University.
2. Soja, Edward W. *Postmetropolis: Critical Studies of Cities and Regions* (Blackwell, 2000): 9.
3. Foucault, Michel "Questions on Geography." In *Power/Knowledge: Selected Interviews and Other Writings 1972-1977*, translated by C. Gordon. (New York: Pantheon): 63-77. Quoted in Edward W. Soja. *Thirdspace: A Journey through Los Angeles and Other Real-and-Imagined Places* (Blackwell, 1996: 69-70)

4. Lefebvre, Henri. *Le droit à la ville/Espace et politique* (Paris: Anthropos, 1968). Quoted in Edward W. Soja, *Postmetropolis: Critical Studies of Cities and Regions*, (Blackwell, 2000): 19.
5. Jameson, Frederic. *Archaeologies of the future: The desire called utopia and other science fictions* (Verso, 2005): 286-8.
6. Heise, Ursula K. *Sense of place and sense of planet: The environmental imagination of the global*. (Oxford University Press, 2008): 3.
7. Abbott, Carl. *Imagining urban futures: cities in science fiction and what we might learn from them*. (Wesleyan University Press, 2016): 24-35.
8. Stableford, Brian M, David Langford, and John Clute. "Cities," *The Encyclopedia of Science Fiction*. Eds. John Clute, David Langford, Peter Nicholls and Graham Sleight. Gollancz, 31 Aug. 2018. Accessed 28 Apr. 2020. <http://www.sf-encyclopedia.com/entry/cities>. The Encyclopedia of Science Fiction, which contains all sorts of information about science fiction authors, books, films and themes and has been periodically updated from 1979 up to this day.
9. Dean, John "Science Fiction City," *Foundation* 23 (October 1981): 64-65.
10. Jameson, Fredric. "Science Fiction as a Spatial Genre: Generic Discontinuities and the Problem of Figuration in Vonda McIntyre's "The Exile Waiting"," *Science Fiction Studies* (1987): 54.
11. Graham, Stephen. "Vertical noir: Histories of the future in urban science fiction," *City* 20.3 (2016): 389-390; Shaw, Debra Benita. "Strange zones: Science fiction, fantasy and the posthuman city," *City* 17.6 (2013): 778,781.
12. Soja, Edward W. *Thirdspace: A Journey through Los Angeles and Other Real-and- Imagined Places*. Blackwell, 1996.
13. Wilkens, Matthew *et al.*, Textual Geographies. Accessed May 18, 2020. <http://www.txtgeo.net/>.
14. Underwood, Ted. *Distant Horizons: Digital Evidence and Literary Change* (University of Chicago Press, 2019).
15. Wilkens, Matthew. "[Genre, Computation, and the Varieties of Twentieth-Century U.S. Fiction](#)," *Cultural Analytics* (Nov. 1, 2016): 2-3. DOI: 10.22148/16.009.
16. Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, *et al.*, [Scikit-learn: Machine Learning in Python](#), *JMLR* 12 (2011): 2825-2830
17. London, Jack. *The Scarlet Plague* (London: Mills, 1916): 29-30.
18. See the first section "Urban space and science fiction".
19. Zoran, Gabriel. "Towards a theory of space in narrative." *Poetics today* 5.2 (1984): 312-3.
20. Ronen, Ruth. "Space in fiction." *Poetics Today* 7.3 (1986): 421-438.
21. *Ibid.* 421,430.
22. Tenen, Dennis Yi. "Toward a computational archaeology of fictional space." *New Literary History* 49.1 (2018): 119-147.
23. Williams, Raymond. *The country and the city*. Vol. 423. (Oxford University Press, 1975).
24. Eisner, Will. *New York: The Big City* (WW Norton & Company, 2006).

25. Macmillan Dictionary, “Areas in Towns or Cities - Synonyms and Related Words.” Accessed May 18, 2020. <https://www.macmillandictionary.com/thesaurus-category/british/areas-in-towns-or-cities>
26. The words “outskirts” and “crossroads” are typed without the final “s”, as WordNet Lemmatizer wrongly transforms them in a non-existent singular form when encountering them in the corpus. Unfortunately, several noun groups (e.g. “city center”) are left out since words are stored without their linguistic context in the HathiTrust Research Center Extracted Features Dataset.
27. Capitanu, Boris *et al.* The HathiTrust Research Center Extracted Feature Dataset (1.0) [Dataset]. HathiTrust Research Center, 2016. Last modified May 14, 2020. Accessed May 19, 2020. <http://dx.doi.org/10.13012/J8X63JT3>.
28. Underwood, Ted, Boris Capitanu, Peter Organisciak, Sayan Bhattacharyya, Loretta Auvil, Colleen Fallaw, J. Stephen Downie, Word Frequencies in English-Language Literature, 1700-1922 (0.2) [Dataset]. HathiTrust Research Center. 2015. Last modified June 25, 2019. Accessed May 18, 2020. <http://dx.doi.org/10.13012/J8JW8BSJ>.
29. The data is produced by first extracting the named locations from the volumes using [Stanford’s named entity recognizer](#) (NER). Then, the resulting location strings are associated with geographic data from Google’s Places and Geocoding APIs. Once the geographic data is retrieved, the addresses are manually checked for greater accuracy. A more complete description of this process can be found in Evans, Elizabeth, and Matthew Wilkens. "Nation, Ethnicity, and the Geography of British Fiction, 1880-1940." *Cultural Analytics* (July 13, 2018): 16 <https://doi.org/10.31235/osf.io/psg2z>
30. Google, “Place Types | Places API | Google Developers.” Accessed May 18, 2020. https://developers.google.com/places/supported_types.
31. William Gibson, *Neuromancer* (New York: Ace Books, 1984), 57.
32. U.S. Census Bureau. “2010 Census of Population and Housing, Population and Housing Unit Counts, CPH-2-5: United States Summary,” (U.S. Government Printing Office, 2012): 20-6.
33. Higham, John. *Strangers in the land: Patterns of American nativism, 1860-1925* (Rutgers University Press, 2002), 301-330.
34. U.S. Department of State, “The Immigration Act of 1924 (The Johnson-Reed Act).” Accessed May 18, 2020. <https://history.state.gov/milestones/1921-1936/immigration-act>.
35. Boyd, Robert L. "A "migration of despair": unemployment, the search for work, and migration to farms during the Great Depression." *Social Science Quarterly* 83.2 (2002): 554-567
36. Bernstein, Michael A. *The Great Depression: delayed recovery and economic change in America, 1929-1939* (Cambridge University Press, 1987), 74.
37. Buck, Stephanie. “Fear of Nuclear Annihilation Scarred Children Growing up in the Cold War, Studies Later Showed.” *Medium. Timeline*, August 29, 2017. Accessed May 18, 2020. <https://timeline.com/nuclear-war-child-psychology-d1ff491b5fe0>.
38. Wiefek, Nancy. *The impact of economic anxiety in postindustrial America* (Greenwood Publishing Group, 2003), 3.
39. Kenneth T. Jackson, *Crabgrass Frontier: The Suburbanization of the United States*. (New York: Oxford University Press, 1987).