Lexichrome: Text Construction and Lexical Discovery with Word-Color Associations Using Interactive Visualization

Chris Kim[1], Uta Hinrichs[2], Saif M. Mohammad[3], Christopher Collins[1]
[1]Ontario Tech University, Oshawa, Canada
[2]University of St Andrews, St Andrews, United Kingdom
[3]National Research Council Canada, Ottawa, Canada
chris.kim@ontariotechu.ca, uh3@st-andrews.ac.uk, saif.mohammad@nrc-cnrc.gc.ca, christopher.collins@ontariotechu.ca

ABSTRACT

Based on word-color associations from a comprehensive, crowdsourced lexicon, we present *Lexichrome*: a web application that explores the popular perception of relationships between English words and eleven basic color terms using interactive visualization. Lexichrome provides three complementary visualizations: "Palette" presents the diversity of word-color associations across the color palette; "Words" reveals the color associations of individual words using a dictionary-like interface; "Roget's Thesaurus" uncovers color association patterns in different semantic categories found in the thesaurus. Finally, our text editor allows users to compose their own texts and examine the resultant chromatic fingerprints throughout the process. We studied the utility of Lexichrome in a two-part qualitative user study with nine participants from various writing-intensive professions. We find that the presence of word-color associations promotes awareness surrounding word choice, editorial decision, and audience reception, and introduce a variety of use cases, features, and future opportunities applicable to creative writing, corporate communication, and journalism.

Author Keywords

visualization; qualitative user study; word-color associations; text construction; lexical discovery; creative writing

CCS Concepts

•**Human-centered computing** → **Visualization;** *Human computer interaction (HCI);* •**Applied computing** → Word processors;

INTRODUCTION

Color has historically been a significant part of human lives: colors tell stories about those who use them in their works, and invite emotional responses from those who see them. These associations form a series of symbols, which comprise the oral and linguistic traditions specific to each culture [17], and our emotional connections to these colors play a pivotal role in branding and purchase decisions [43].

The relationship between words and colors is a fluid one, influenced by time, geography, and culture. Shakespeare, in his numerous plays, describes the color green as both a symbol for youth and innocence, and occasionally decay and envy [16]. This holds true even today, as colors that are used on special occasions vary by region and culture. The dichotomy between East-Asian and European cultures, in use of white or black respectively for mourning, seems less significant in comparison to funeral ceremonies in Mexico that feature coffins and apparel adorned with bright colors [17]. Despite this variability, however, within a region and time period there is a large set of concepts that are strongly consistent in color associations. For example, in the United States, *money* is associated with green, *love* is associated with red [33].

In addition, many stakeholders recognize the linguistic and symbolic importance of color—particularly in product marketing and layout design—and strive to reap the benefits of further strengthening the message and triggering the desired emotional response from potential audiences. While the majority of studies in color psychology indicate that color is subject to "strong individual and cultural differences" [51], where the characteristics, the prior experiences, and the associated cultural upbringing of each individual affect the way he or she interprets different colors, studies have shown that consumers are responsive to color in recognizing familiar brands [27].

Despite the evident importance of color in inviting emotional response, there has been a lack of reliable lexicons containing word-color associations: perhaps a lost opportunity in more effectively communicating the message. This gap has since been addressed by a wealth of algorithms and approaches to accumulating word-color association, ranging from identifying co-occurence of color terms and n-grams [42] to extracting colors from semantically relevant image assets [31]. One of many seminal datasets is the comprehensive lexicon published by National Research Council (NRC) Canada [32, 33], comprised of more than 25,000 unique word-color associations, with each entry consisting of a word, the list of associated word senses, and the voting results for all word-color associations.
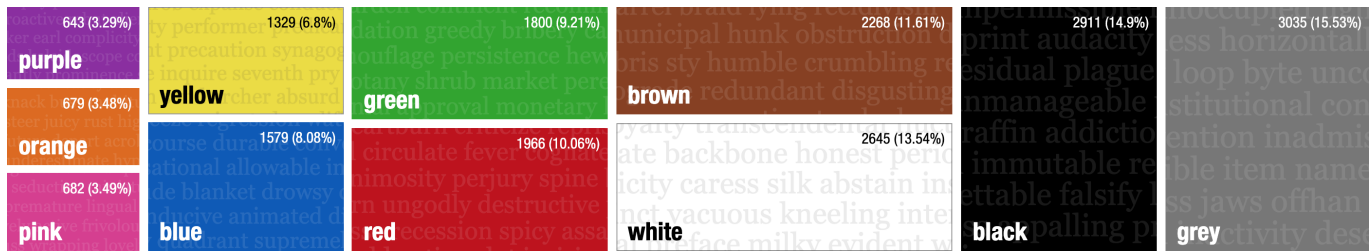
Figure 1. *Palette* view color tiles display the number of word-color associations, as well as a partial list of associated words.

Lexichrome is an interactive visualization inspired by the NRC dataset, mapping it to a web-based interface that offers its visitors the ability to examine the popular opinion on word-color associations and leverage this information to inspire their own text compositions and creative writing. The application invites visitors to freely explore the dataset through a wide offering of interactive visualization techniques, and to discover insights from their own work by applying the color associations to their own text creations. Lexichrome seeks to inform the work of literary scholars and creative writers as a "provocateur" as well as a source of inspiration.

We discuss the potential and associated challenges of supporting creative writing processes by incorporating word-color associations based on our two-part qualitative user study with nine participants from various writing disciplines. From the study, we find that Lexichrome inspires a more exploratory approach to writing or tailoring texts, as well as mindfulness of audience reception. We also outline a variety of use cases and scenarios in which Lexichrome might be used beyond creative writing scenarios to enhance the work of literary scholars, brand managers, educators, and writers. Our work at the intersection of visualization and interface design also provides a critical view on how approaches like ours meant to enhance creativity in the widest sense, can also introduce new and enforce learned biases, potentially with a negative impact on users and society at large.

BACKGROUND AND INSPIRATION

Visualization of literary texts is not a new concept to both academic and artistic domains. Applying natural language processing techniques, such as lexical scoring (e.g. *tf-idf*) and topic modeling, to a collection of texts has become a common practice in literary text analysis, and the results are often mapped to a number of visualization tools that offer statistical and anecdotal insights about each text [19]. They are not only applied to literary works of the distant past, but also used to enable journalistic discoveries from informal conversations on social networks [12] and large corpora [8].

In-depth surveys of the state of the art in text visualization and computational text analysis have been published in both the visualization community [2, 26], and in the humanities, where the TAPoR Project catalogs an array of related scholarly projects as a hub that explores the possibilities of computer-assisted analysis [40]. TextArc [37] and PoemViewer [1] are two of the prominent applications featured as part of the TAPoR project, as both applications seek to visualize structural and relational information of each text using visual encodings.

At the forefront of creative application, literary visualization takes an unexpected turn: visualization designers used edge bundling to offer a summary of character interactions in novels [4], a color signature—as interpreted by the artist—to represent books [44], and even the visualization of relationships between ingredients of each meal introduced throughout a novel [49]. News outlets embrace the trend of using computational text analysis and interactive visualization to tell a story as well: one group applies the Flesch-Kincaid readability test to every presidential address and visualizes the results to claim that the linguistic standard of the State of the Union has been declining [18], while a popular webcomic *xkcd* and its academic counterpart formalize a technique of "storyline visualization" intended to depict the temporal dynamics of social interactions occurring in each work [34, 45].

The "literature fingerprinting" approach to literary analysis introduced the application of the pixel-oriented visualization method to displaying features of words as individual entities or as part of a longer text [24]. This approach is also thoroughly explored in other works that feature color overlays on otherwise static and monochrome texts without transforming the layout structure of the original text [14], and analysis functions of our work are inspired by such prior visualizations.

Serendipity and playfulness have also served as goal for text visualization, as in the Bohemian Bookshelf project, which applies coordinated views to visualize multiple attributes of books in a collection and allows its users to search for books not based on the title or the author, but based on unlikely attributes such as cover color, page count, and content era [48]. We borrow the playful aesthetics in this work, to encourage exploration for the purposes of discovery and enjoyment.

The application of natural language processing and interactive visualization techniques, in investigation of academically valid insights, is not without its own share of critics. As the domain of digital humanities embraces computational text analysis—consisting of natural language processing, statistical computing, and corpus linguistics—as the preferred form of literary investigation, the existing field of literary criticism argues that the resultant empirical facts and patterns are insufficient to qualify as the "basis for judgment" without the critic's own interpretation, rhetoric, and politics [39]. Another literary critic takes a harsher tone, as the author claims that the computational literary analytic techniques are futile in emulating human intuition and that "machines just don't get it" [30].

This heated debate is partially attributed to the current lack of comprehensive resources that empirically capture relations

between language and cognition, but despite this scarcity, a number of visualization projects rely on encoding linguistic characteristics using colors. Harris' visualization *We Feel Fine* harvests emotion words from a large number of blog posts, which are then mapped to a series of differently colored dot particles, each corresponding to the tone of the instilled emotion [21]. Wattenberg's *Color Code* establishes an interactive map of more than 30,000 nouns, each depicted by a color block based on the average color of image search results [50]. Lee introduces a visual comparison between two languages, Chinese and English, and their ways to describe color [29].

INTRODUCING LEXICHROME

Lexichrome allows the visitors to uncover different facets of word-color associations using various interactive visualizations, each with a number of improvements that build on existing approaches. This section presents a number of design considerations for the application, along with features offered by Lexichrome and potential use cases. The most recent version is publicly accessible at `http://lexichrome.com`.

Note about Source Dataset

Our approach is not tied to a specific lexicon, as Lexichrome has been developed with interoperability, modularity, and compatibility in mind with the capacity to accommodate alternative datasets. The prototype described herein is driven by the NRC dataset which is framed around the eleven most common color terms. We recognize that there is extensive literature surrounding color perception and color-concept associations beyond eleven basic color terms. As various papers on color preference [36, 52] and mood associations [20] suggest, lightness and saturation play a significant role in the way users perceive and associate with presented colors.

By restricting the cardinality of the palette to a limited number, we increase the probability of agreements on word-color association, ensure familiarity of the palette, and reduce the complexity of the views. The subsequent paragraphs discuss the NRC dataset as a primary source of data for Lexichrome.

NRC Dataset: Overview

Studies in linguistic relativity by Berlin and Kay [3, 23] have shown that often color terms appeared in languages through an evolutionary pattern: if a language has only two color terms, then they are white and black; if a language has three terms, then they are white, black, and red; and so on up to eleven colors. From these groupings, the colors can be ranked as follows: white and black, then red, green, yellow, blue, brown, pink, purple, orange, and grey. They are referred to as basic colors in cognition research.

The dataset was collected from a survey of 2,000 US-based participants on a crowdsourcing platform, to create a dataset of 25,000 word-color associations. The annotators were presented common English words, one at a time, and asked to select which of the eleven basic colors is most associated with a given word sense, as a single word may have different color associations in different senses. Post-annotation analyses showed that even though the color options were presented in random order, the order of the most frequently associated colors is identical to the Berlin and Kay order. About 32%

Figure 2. Partial list of English terms associated with the selected color (green), sorted in row-major order by participant agreement level.

of the words had strong color associations ("the colour is a salient feature of the concept the word refers to." [32]), and still more had weak to moderate associations.

Strong associations were discovered not only for concrete terms (e.g., water) but also for more abstract concepts (e.g., jealousy) [33]. While the resulting data tells us the associations in the contemporary US context, many of them more common to human experience (e.g., *water–blue*) will likely hold even for other cultures. This dataset has been used by several other researchers for a variety of research purposes (e.g., [10]).

NRC Dataset: Color Palette

There may be some variation amongst hues that share the same name: the *red* of a fire engine differs from the *red* of a rose. The NRC dataset does not associate words with actual hues, but rather with names of eleven basic colors (black, white, red, etc.). As a result, one challenge was to select hues for the visualizations which best matched the color names in the word-color association data. While color map design is an active research area in visualization and cartography [9, 47], and custom palettes are shared and critiqued amongst designers [28], there are no agreed canonical hues for various named colors to our knowledge. Thus, rather than selecting an existing palette or designing our own, we turned instead to a data-driven approach to select the hues for Lexichrome.

The hues used to represent the eleven-color palette throughout Lexichrome are based on the data generated by the "color names" experiment conducted by Dolores Labs, a crowdsourcing research company now known as Figure Eight [35]. Similar to the World Color Survey [22] in its collection method, the crowdsourced dataset was collected by showing participants a hue and inviting them to provide an English name for the hue, such as *blue*, *turquoise*, or *lavender*. 10,000 color name-hue pairings were collected in the RGB and HSV color spaces. From this dataset, we extracted all hues matching each of the eleven color names of the NRC lexicon and averaged them as a mean of circular quantities. The resultant eleven hues were implemented as the Lexichrome color scheme. Below we describe Lexichrome's individual features—*Palette*, *Words*, *Detail*, *Roget's Thesaurus*, and *Text Editor*—that each provide a unique perspective on the relationship between words and associated colours. Present potential use case scenarios with each feature that informed our design process.

Palette: An Overview of Word-Colour Associations

The "Palette" view, which the visitors first encounter when they access Lexichrome, displays the distribution of color
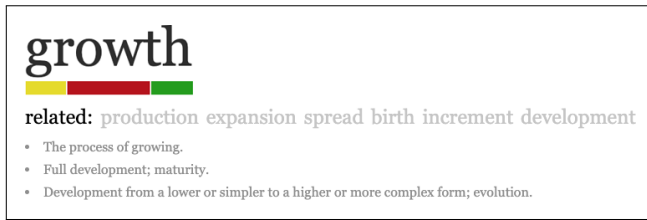
Figure 3. Chromatic makeup, dictionary definitions, and related terms for the term *growth* in the "Detail" view.
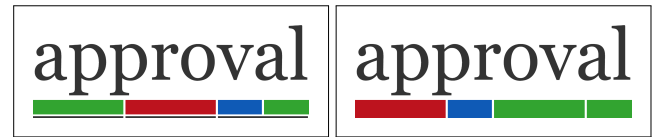

Figure 4. Two ways to sort the color association blocks: the first image illustrates the clustering of color blocks by word-senses, and the second demonstrates the clustering by hue.
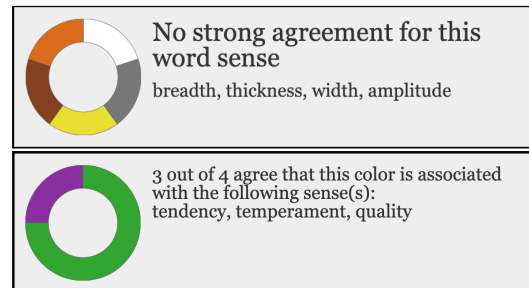

Figure 5. Tooltips that show vote details for *latitude* (top) and *nature* (bottom), displayed upon hovering over a color block in the "Detail" view.

terms in the database using a single-level treemap visualization. As illustrated in Figure 1, each of the eleven treemap "tiles" represents a color. Each tile also displays the number of word-color associations pertaining to the color, as well as a randomly generated sample of associated words.

Selecting a tile reveals the complete list of associated English terms, as partially shown in Figure 2. Each bar in the chart illustrates the strength and the relevance of a word-color association, while the small caption below the word represents the number of associated word senses involved. For instance, the word *wealth* is labeled as "3 out of 4 senses strongly associated" to green, meaning that a single sense of *wealth* has a less-than-significant association with the color green, ultimately resulting in a slightly weaker but nonetheless dominant association. The length of the bar in the chart presents the strength of the word-color association in a more granular fashion: for each word, the agreement levels across the all associated senses are averaged and represented as a percentage. For example, this granularity allows visitors to discover that the word *salad* has a higher level of agreement with the color green than the more abstract concept *headway*, despite all of the senses associated with both words indicated as being strongly associated with the same color.

Potential use case scenario: A marketing writer for a consumer goods company chooses color green in the "Palette" view to see words commonly associated with the company's brand color. The user confirms that green is commonly associated with words including *forest*, *botanical*, and *prosperity*, and decides to incorporate them in the next project.

Words: Lexicon Search with User Query
The "Words" view functions as a dictionary of word-color associations, as it allows visitors to search for multiple terms or examine the randomly populated list in order to view associated colors. By hovering over each entry, the visitor can access the extra layer that displays the "chromatic makeup" of each word, as illustrated in Figure 3. The bar displayed below the highlighted word represents the number of associated senses, their color associations, and the degree of disagreement for each sense-color association. If the word has five senses, five blocks of equal length, each colored with the associated hue, are displayed. For example, the word *growth* displays a series of yellow, red, and green blocks with the presence of the color red particularly stronger than others, meaning that the majority of the original survey participants agree that at least half of the associated word senses are associated with red.

Potential use case scenario: A brand copywriter enters the "Words" view to confirm assumptions regarding project key-

words. She is happy to learn that words including *baby* and *organic* are commonly associated with colors pink and green, but is surprised to find out that the word *safe* has a strong disagreement, mainly due to various usage contexts of the word.

Detail: Revealing Additional Word Information
Upon clicking on an instance of a valid lexicon entry anywhere on Lexichrome, visitors are redirected to the "Detail" view for that word, which reveals additional information including the list of related terms, definitions, and the level of participant agreement for each sense-color association.

The list of definitions that displays below the word is generated through a series of API calls to Wordnik, an online dictionary service that aggregates multiple sources of word definition and thesaurus data [53]. The resultant definitions are not sense-specific, and are shown to the visitors as a potential reference that informs the displayed color associations. In addition, a set of "related terms" — as identified during the collection of the initial data — that serve to disambiguate the senses of the selected word are displayed. As these secondary words sometimes have a set of sense words themselves, the visitors can further explore the dataset by clicking through these related terms one after another.

Color blocks displayed below the primary word indicate the number of associated senses, their color associations, and the degree of disagreement for each sense-color association. For example: if the word has five senses, five blocks of equal length, each colored with the associated hue, are displayed.

In case of disagreement or tie amongst the color associations for a single sense, the original block representative of that particular sense is further split into additional blocks that represent the "colors in dispute." For instance, if a certain sense-color association has an equal number of votes for three colors — red, blue, and green, for example — the original "sense block" will be divided once again into three smaller blocks of equal length to represent this disagreement. This
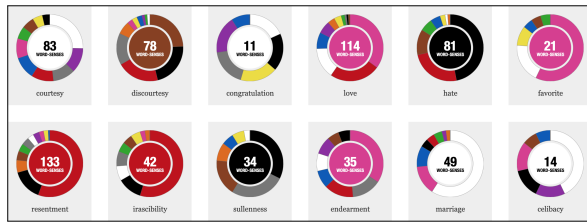
Figure 6. Navigation in the "Roget's Thesaurus" view, representing the section "Sympathetic Affections" and its categories.
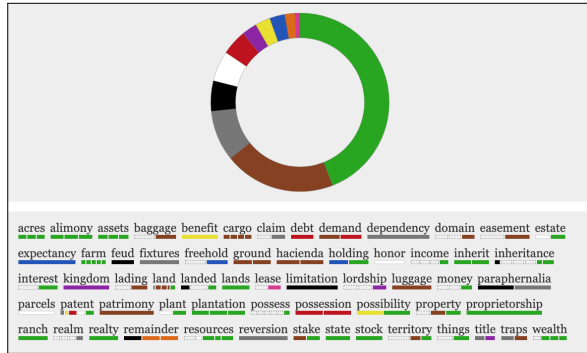

Figure 7. List of words and their color association blocks that belong in the lexical category "Property."

design decision was made in order to expose sense-color association uncertainty without giving extra prominence to any particular word sense. As evident in the left image of Figure 4, the word *approval* has a total of three senses, and the third has a disagreement in the color association.

The resultant color blocks of varying lengths can be further sorted in two ways: they can be clustered based on the associated senses or colors. "Cluster by sense" clarifies the details of sense-color disagreements for each sense individually, but results in colors appearing in multiple locations in the color bar. "Cluster by color" (Figure 4, right) allows the visitors to view the overall distribution of color associations for a particular word, and is applied by default across Lexichrome.

Color blocks that are marked with dashed borders indicate that there is no strong agreement for those particular word senses in terms of color association (no color won the "majority vote"). To reveal the original voting data, the user can "hover" over each block and open a tool tip window, as shown in Figure 5. Tool tips contain a donut chart illustrating the distribution of votes for across various sense-color associations. This information is available not just for the "undecided" blocks, but for all the color blocks in the "Detail" view as well.

Potential use case scenario: To seek inspiration for a project, a poetry author clicks on the word *nature* and enters the "Detail" view to discover a surprising association with the color purple (although not as strong as with green). She decides to further explore the word list for "purple" and discovers other terms—amethyst and grapes she considers for her project.

Visualizing Roget's Thesaurus

Roget's Thesaurus was originally created in 1852 and included about 15,000 English words. Roget's taxonomic structure groups words into six classes. These six classes are further
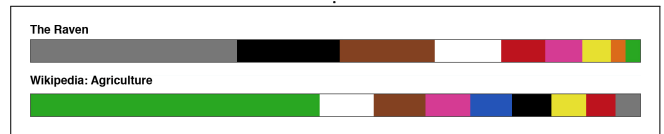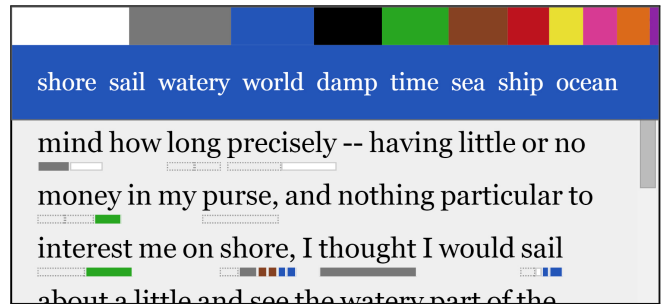

Figure 8. "Chromatic fingerprint" based on excerpts of *Moby Dick*. The "word drawer" displays the list of in-text words that are associated with a particular color (top). Fingerprints of excerpts found in Edgar Allan Poe's "The Raven" and the Wikipedia entry for "agriculture" (bottom).

partitioned into thirty nine sections, which are in turn divided into one thousand categories. Each category lists about 20 to 200 related words and expressions.

The "Roget's Thesaurus" view allows the visitors to explore and examine the distribution of colors across the semantic categories of the 1911 edition of Roget's Thesaurus [41], by accumulating color associations bottom-up across the hierarchy. Two distinct visualization techniques are used in this view: the donut chart and the "chromatic makeup" color bar. First, a small multiples view of donut charts allows the visitors to explore and examine the distribution of colors across the semantic categories of Roget's Thesaurus. For example, this view shows the predominant color associated to "borrowing" terms is green whereas for "stealing" terms it is black.

Analysts can drill into deeper levels of the hierarchy by clicking on a desired category, as illustrated in Figure 6, and traverse back up to explore other branches. The user can hover over each slice to reveal more information about the presence of a specific color pertaining to the selected class, section, or category. For instance, a category slice named "Property" that belongs in the aforementioned "green" section indicates that this category is predominantly green, as about 44% or exactly 48 of all relevant word senses are marked as green. The users can also access the list of lexicon words included within each category, as noted in Figure 7.

Potential use case scenario: A student journalist explores the "Roget's Thesaurus" view, and discovers that the two words that are completely opposite in meaning can share the same color: words that belong to a category named "resentment" are predominantly red, while another category named "affections" contains a large number of words associated with red as well. Also, the journalist remembers that the color green is also strongly associated with the concept of money, as notices that the "Possessive Relations" words in the view are generally marked as green as well. Fascinated by the duality of these colors in meaning, the journalist enthusiastically makes note of these findings and proceeds with preparing for the meeting with the editors to discuss the magazine cover art.
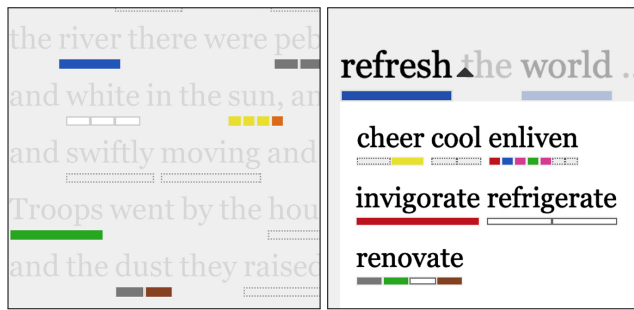
Figure 9. The left image demonstrates toggling between different presentation modes of "Text Editor" view: spatial information remains intact when words are hidden away. The right image features the synonym window where each term is equipped with its own set of color blocks.

Text Editor: Generating "Chromatic Fingerprints"

Lexichrome has an ability to process user-provided texts and extract the relevant colors. As shown in Figure 8, Lexichrome allows creative writers, journalists, marketing and PR professionals, and literary linguists to generate and edit new texts or paste existing ones into Lexichrome's Text Editor. Upon entering new or pasting existing text of any length into the text box, or choosing a sample excerpt from the list of preset texts, the application sends the provided text to the parsing script which processes it through normalization, lemmatization, and function word removal and extracts an array of lexical (content) words. It is then matched to make associations to relevant colors, whose results are subsequently displayed using a variant of the "fingerprinting" motif of Keim and Oelke [24].

The top bar in the fingerprint, consisting of multiple color blocks of different lengths (see top colored line in Fig. 8), represents the overall distribution of in-text colors. Figure 8(top) also demonstrates the ability to activate the "word drawer", accessible by clicking on each color block, to view the list of text words that are associated with each color. For example, one can confirm that the majority of "blue" words in the first paragraph of *Moby Dick* are generally associated with the concept of sea. The fingerprint can give an overall aesthetic impression. Compare *Moby Dick* to the fingerprints shown in Figure 8 (bottom): Poe's "The Raven" (predominantly black, grey) and the Wikipedia entry for "agriculture" (green).

As illustrated in Figure 9, the visitors can "toggle" the words and the color blocks present in the chromatic fingerprint. Hiding the words highlights the text's chromatic fingerprint over its semantic meaning (see Fig. 9(l)). The distance between words remains intact, however, and this results in an encoding of both spatial and frequency information of colors evoked by different words. The users can also move the cursor over the words to reveal their synonyms, each equipped with its own set of color blocks, which can be clicked to replace the original term. Figure 9(r), for example, shows synonyms and related terms for the word "refresh," from the Coca Cola brand statement. This allows writers and brand managers to revise their texts on-the-fly, for example to include "invigorate" which invokes the brand color red, as they proceed with writing articles designed to encode certain evocative qualities.

Potential use case scenario: A creative writer tests out a number of sample preset texts provided by the application and discovers that Poe's "The Raven" does contain numerous words she assumed to be dark and achromatic and that the Wikipedia article on "agriculture" is predominantly green.

Curious to find out more, the writer submits the famous "to be or not to be" soliloquy from William Shakespeare's *Hamlet* and notices that the colors are more diverse than initially predicted, and attributes this characteristic to a number of metaphors that Hamlet makes throughout the monologue. The writer triggers the "word drawer" feature for the color blue and notices that he uses the words *seas* and *currents*, and that the large amount of "blackness" in the text is a result of various negative words including *die*, *death*, and *sins*.

Implementation

Lexichrome uses a number of existing server-side and client-side technologies to parse, process, and display the word-color lexicon. PHP is used to access the MySQL-driven lexicon database and process user queries, while the JavaScript implementation of the Snowball algorithm is used to clean up each instance of user-provided texts before querying the database for word-color associations [38]. In the front-end realm, the jQuery library and D3.js were used in order to quickly generate interactive visualizations of word-color data [6].

QUALITATIVE USER STUDY

Lexichrome's features cover a variety of potential use case scenarios as outlined above. As a first step to investigate the potential impact of this approach of highlighting word-colour associations, we decided conduct a qualitative study with people from a range of creative writing disciplines: if and how does the presence of word-color associations influence creative writing processes? We were particularly interested in how writers would make use of and experience Text Editor.

Participants

We recruited nine participants for our study from local academic and leisurely creative writing groups. All our participants are storytellers by training and compose texts as part of their professional everyday work. Two of our participants are journalists (**J**), two write articles as part of their jobs in independent media (**M**), three regularly engage in creative writing activities (e.g., writing literature and poetry) (**W**), and two participants hold professions in corporate communication (**C**) and regularly compose texts as part of this. No participants had used or knew of Lexichrome prior to our study.

Study Procedure

Our study consisted of two phases, a grounding interview about the participant's writing workflow, and a prototype testing session to investigate how the intervention of Lexichrome specifically and word-color associations in general would affect one's writing process.

Phase 1: We first conducted individual interviews with each participant in order to gain insights into their general approach to creative writing and their writing processes. As part of these interviews we asked participants to describe the type of texts they typically work on and their preferred writing style and genre. We asked about their inspirations when writing and the role of visual aesthetics (e.g., conveyed by visual features
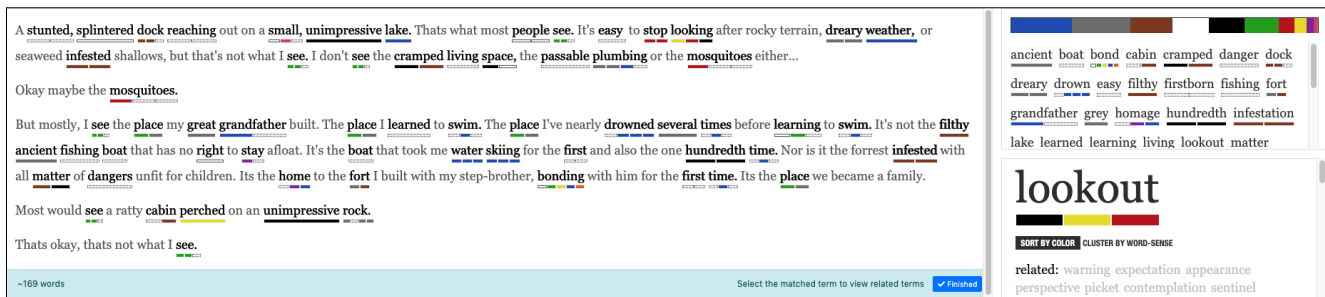
Figure 10. "Text Editor" view with an article by one of the study participants. The user can access word-color associations for individual words on-the-fly and toggle other features including word definition and related terms by clicking on individual terms.

of the text and/or the content) as part of this. Interviews also included questions about preferred writing tools and any obstacles or frustrations encountered during the process.

Phase 2: We scheduled a second individual meeting with participants where we first introduced them to Lexichrome, and then asked them to write a brief text passage (150–250 words) of their choice in Lexichrome's Text Editor. We did not provide participants with instructions on what to write—the writing was purely driven by participants' preferred writing genre and subject matter. This creative writing phase was followed by an interview where we asked participants to reflect on their experience. Interviews included questions on how Lexichrome's features influenced participants' writing process in a positive and/or negative way (if at all), also in comparison to their typical writing process and tools.

Data Collection & Analysis
Data collection included audio recordings of interviews (both Phase 1 and 2), participants' ratings of Lexichrome, and screen recordings of participants writing processes (Phase 2) [46].

We collected approximately 7 hours of audio and 4 hours of video recordings across all 9 participants. All data were transcribed word-by-word and then analyzed using a thematic coding approach [7, 13], iteratively developing and refining a qualitative coding scheme. As part of this, one researcher qualitatively coded participant statements per interview.

These codes were then discussed and refined with two additional researchers, before they were expanded to additional interviews. Initially, the coding scheme was directly informed by interview questions (e.g., regarding participants' typical writing process, compared to that described in Lexichrome), but additional themes emerged, for example, with regard to how particular observed word-color associations influenced participants' experience with the tool.

This analysis of interviews in Phase 2 guided our analysis of video recordings of participants' use of Lexichrome. We first conducted a high-level analysis of all video recordings, noting down interaction times, and particular Lexichrome features explored [15]. This was followed by a more in-depth analysis of selected video sequences, prompted by the interview analysis. For example, when participants commented on their writing process in Lexichrome or their reaction to certain word-color associations, we reviewed these episodes in the video.

Our in-depth analysis of the interviews, complemented with our video analysis led to interesting findings regarding participants' use of Lexichrome and how the presence of word-color associations has influenced their writing process and experience, as described in the following two sections.

LEXICHROME'S INFLUENCE ON WRITING PROCESSES
We begin the description of our findings by first outlining participants' general use of the Lexichrome editor and the types of texts that they composed. We then describe in more detail how Lexichrome and in particular the representation of word-color associations integrated into the Text Editor feature influenced participants' writing processes.

Participants' Use of Lexichrome
Participants spent an average of 15 minutes writing their texts in Lexichrome's Text Editor (7 minutes min.; 44 minutes max.). The word counts of texts that participants wrote ranged from 93 to 277 words (167 words on average). We found that participants' choice of topic was somewhat driven by their professional background and ongoing projects as well as personal experiences. Three participants wrote texts that were ***project-driven***, that is, directly linked to ongoing work projects. For example, M1 (working in independent media) created a brief introduction to the their next podcast episode. J2 (a journalist) wrote an opening paragraph to their upcoming newspaper article. Finally, C1 (working in corporate communication) paraphrased a company press release.

In contrast, three participants created text excerpts that were more personal and even ***autobiographical*** in nature, in that the writing was rooted in their lived experience or personal pondering. J1 (a journalist) recounted on starting a new role at an organization. W3 (a creative writer) described a summer vacation at a family cottage. Finally, C2 (working in corporate communication) wrote a personal essay, revolving around various instances where people mispronounce their name.

Finally, three participants focused on describing ***concrete and abstract concepts***; their texts were not directly linked to a personal experience or current professional project, but certainly linked to their professional background in the widest sense. W2 (a creative writer) wrote a brief, self-referential text reflecting on how to write an effective essay; M2 (working in independent media) wrote a comment on the pleasant experience of riding a bus. Finally, W1 (a creative writer) produced a prose poem about a specific word, as they were inspired to create a piece of writing with heightened visual imagery.

This shows the wide variety of texts produced as part of the study sessions. Most participants' writing inspiration was not driven by Lexichrome's features, at least not initially. However, notably, one of our participants was immediately and directly inspired by the word-color associations presented in Lexichrome: "*The whole idea around words and colors is still in my mind, so I wanted to write based off of one word. I started with the word "bliss" and then I just kind of started going off on a random . Yeah, I was just inspired by words and colors and tried to think about what's a piece of writing that I can create that is also very visual.*" [W1].

Naturally given the study setup, participants focused mostly on composing their text within Lexichrome's Text Editor. However, both our video and interview analysis revealed that participants' also took note of the Editor's Chromatic Fingerprint feature (see Fig. 8), the Related Terms feature (see Fig. 9(r)), as well as the Definitions feature (see Fig. 3), and out of four participants who used all of the features at least once, two participants actively incorporate these additional features in their writing process. These features, combined with the presence of color associations below individual words had an influence of how participants experienced their writing process and how they reflected on their text composition during the writing process and in retrospect as outlined below.

Word-Color Associations: Inspiration & Interruption

Statements from participants suggest that the presence of color associations below individual words was a source of inspiration. Our video analysis revealed that the word-color associations made participants pause their writing process from time to time and re-consider their writing. This would not necessarily result in actual word changes, but seemed to trigger mental ideation processes, as reflected in the following statement: "*Every time I started a new sentence, when it would load up the colors associated with each word, I would stop for like a second and go 'oh, interesting' and take a second to see what colors were associated with the words[...] I think, I was reconsidering not my train of thought, per se, but: 'Maybe I phrase it like this.' [...] Maybe like an epiphany: 'Oh, this is how I can continue writing.' It was surprisingly inspiring in a way that it didn't impede my writing.*" [W2].

Other participants felt that even a perceived misrepresentation of word-color associations would direct their writing process. For example, W1 commented: "*It's really interesting to see what kinds of colors are associated with words, and I think it provides like a challenge to your writing, because if you don't see a certain word that way, then it kind of makes you work hard enough to even think of a different word or to project the color that you do see through the writing and to make it that color, even if it's not normally associated with that.*"

However, the same participant also pointed out a potential negative effect of this type of influence on the writing and creative process: "*One may be overthinking it a little bit too much. [...] It may interrupt you writing naturally. It might influence what you decide to write.*" [W1].

Another creative writer commented: "*I think it [Lexichrome] would cause me to second guess myself a lot. Maybe I (will) try too hard to fit the theme and lose the flow of what I've written, lose the purity of (content) straight from your mind.*" [W2] Similarly, a participant from corporate communication said, "*It would influence my writing and I don't know if I would want it to in that moment.[...] For me if it [presence of word-color associations] was an optional button that I could get on turn on after I'm done... That would be helpful.*" [C1].

Word-Color Associations as Triggers of Reflection

Interviews with participants after their writing experience indicate that the presence of word-color inspired reflection about the writing process and the meaning of their text.

Reflection on Personal Writing Style & Word Choices: The presence of word-color associations made participants critically reflect on their writing style and word choices. J1 highlighted "*It [Lexichrome] is making me think about: 'Is my writing too conceptual?' And: 'I may be using boring words all the time.' Or: 'Should I be describing things that excite people and put things into their actual heads?'*" W3 reflected: "*It made me realize how much, I guess, I'm a 'blue' and a 'brown' writer, because those words kept coming up. I changed a few of them but there were a few that I didn't want to change.*"

For example, W3 replaced experimented with "get past" and "stop looking," the two similar phrases, and chose the latter upon reviewing the resultant color associations. In addition, W3 replaced the phrase "dirty old" with "filthy ancient" upon reviewing their related terms.

Another creative writer highlighted how the reflection on word choices through Lexichrome triggered positive emotions: "*It was kind of exciting. It made me feel good about the words that I was using, because I felt like I was using a variety of different emotions, I guess, with the word that I was choosing, and it made it [the text] feel more lively, and like the piece had a sense of its own.*" [W1]. Similarly J1 stated: "*I think 'issue' surprised me. I just thought it was weird that it was red. I thought it might be another boring word. I found, I was using extremely boring words the whole time, and then if 'issue' came up as red and I was like okay that's cool because I hadn't had a red yet except for a little bit.*"

Reflection on Audience Perception & Interpretation: There was also clear evidence that word-color associations made participants think of how the readers of their texts would emotionally react to the choice of certain words. That is, color and emotion were intrinsically linked.

For example, one journalist participant stated: "*I want people to feel something more physical, internally—so bolder colors. Maybe I want them to slow down and feel a bit sad for a little bit, so maybe I'll search for more blues.*" [J1]. Another agreed: "*[For longer stories] I really want to focus on 'How are people going to take this?' [...] I can really see myself flopping something in [a text into Lexichrome] and being: 'Okay interesting, this is how people might read this.'*" [J2].

Similarly, C1 mused about using Lexichrome to select words associated with certain colors to trigger specific emotional reactions. "*I think, I will be able to switch out words... 'That's not the color, emotion I'm really going for... this is what I'm*

going for.' [C1]. As somebody writing in a corporate context, however, C1 also highlighted that they saw the potential of Lexichrome more in the creative writing industry: "*It (Lexichrome) does not affect the work I do just because the audience is different. If I was writing fiction, for example, the novel, this would be very useful. Colors are I think are critical in the words you write in the story you compose.[...] If it is something more creative, of course, I definitely want it. If I was writing something more cut and dry, not necessarily.*"

One participant pointed out that the presence of word-color associations made them aware of different possible interpretations of a word: "*When I wrote the word 'bonding', and I saw all these different colors pop up, and then I realized the different connotations of that word and how the different connotations will definitely bring up different colors.*" [W3].

All of these examples highlight how much our participants associated colors with emotions which, in turn, facilitated (1) self-reflection their own writing, (2) translating of emotions into text, and, closely related, (3) estimating their audiences' potential emotional reactions to their texts.

Word-Color Associations as Provocations
As stated above, the presence of associated colors within the texts was not something that participants could ignore—the colors appearing with individual words were experienced by participants as a visual provocation that triggered emotions and inspirations, but also felt like an interruption to some.

All our participants commented at least once on specific word-color associations and how these did or did not match their own expectations and associations with the word in focus. Participants would often try to rationalize why a certain word was linked to a color. For example, rationalizing popular color choices based on the colors of physical objects or real-world phenomena related to the word in question, as illustrated in the following statements: "*'Spark' is red, which makes you think about fire.*" [C2]; "*I don't know if it's because [the public bus company] is red. I'm not sure I would assign yellow to this [the word 'bus']*" [M2]; "*'Funk' being blue, for whatever reason, that makes sense to me, more than the other words.*" [M1], "*'Company', I guess, I expected gray like cubicles. It was really interesting to see how people might feel when they're reading a word even if they don't know [the color], because you are not thinking 'Company? I'm bored already!'*" [J2]. This latter two statements, again, highlight how some colors are inextricably linked to certain emotions.

In particular, words representing more abstract concepts triggered more surprises and disagreements with popular associations as represented in the underlying dataset. This, in turn, made participants reflect on why these color associations were so popular and whether cultural connotations may have an influence, but also why they personally disagreed with them. For example, J2 mused: "*Things like 'fashion'—like it was a thing that I can immediately place like: 'This is a color!' Why particularly is it brown? What is it that makes people associated that with brown?*" C2 reflected on their own color associations that disagreed with popular choices: "*'Bad' I wouldn't think as black, I would think of as red. So this kind*

of just made me think about my own color associations: what would have led me to think that?" Interestingly, word associations with the color black often triggered some disagreement among participants. For example, C1 stated: "*Black for me is a very dark color, being associated with death or illness, sickness. I can see 'hub' here [in their text]. It's black and gray. 'Hub' could mean a bunch of different things, but in this context, it was a positive connection.*"

Participants also commented on the possibility of Lexichrome triggering stereotypes: "*'Pretty' didn't actually show up as pink. I thought it [Lexichrome] may reinforce certain stereotypes that some people have. Especially if we think about gender and colors and how they were so linked in the past.*" [C2].

It may have been these disagreements, an urge to change their text in order to bring in certain colors, or simply the search of synonym that made participants turn toward the "Related Terms" feature of Lexichrome (see Fig. 9(r)). In general, participants liked the idea of including this feature directly into Lexichrome's text editor, as illustrated by the following statement: "*For me the really useful part [of Lexichrome] is clicking on words and having alternative words. It would be useful [in my usual text editor], if I don't have to go to an alternative site for that [thesaurus].*" [C1]. However, interview statements revealed that while participants were quite excited about this feature, they would like to see more synonyms for existing words, rather than related terms: "*When I clicked it [a word], I only saw words that are similar but not synonyms. That would be the first thing that I would look for. When I clicked it, I wanted to see synonyms that would make my writing more colorful, but these were all totally different words, and I went 'Okay, that doesn't help.'*" [J1].

DISCUSSION
Distilling the findings from our analysis of user study results, we present a number of insights surrounding word-color associations and their impact on creative writing processes.

Influence of Word-Color Associations
The surfacing of word-color associations can influence the writing process. Although the authors' approach typically did not change in terms of focus and content, the tool supported careful vocabulary selection while minimizing interruption.

Approach: We found that the presence of word-color associations has little to no effect on the author's approach to the text in terms of its content-related focus, in particular if the work is defined by a set of requirements or factual reporting. However, such associations are useful as part of more exploratory approaches to writing or tailoring of texts to evoke a specific set of colors, and in turn, imagery, in the reader's mind.

Vocabulary: While the majority of authors may continue to rely on traditional methods of exploring synonyms and other related terms, the presence of word-color associations may influence the selection of alternate terms not only to avoid repetition, but also to capture a certain feeling or mood. Such visual feedback also allows authors to be mindful of audience reception. Surprisingly, disagreements and weak associations in the dataset did not have a negative effect on word choice.

Rather, they prompted thoughtful reflection on why survey respondents may have had a particular association. This leads to the possibility of even messy crowd-data to be a useful tool for brainstorming and prompting during creative writing.

Flow: It is not yet clear whether one's writing process is impeded in the presence of word-color association due to the author's awareness of additional insights or potential usability issues in Lexichrome, but the author's process was interrupted, however minimal, by on-the-fly visual feedback. This effect could be minimized by disabling the visualization and allowing the user to explicitly request it. More subtle and fluid animations could reduce the disruption caused by flickering.

Use Cases
We discovered that word-color associations are most useful when used for experimenting with words, proofreading completed texts, and establishing tone. Lexichrome promotes discovery of words potentially useful in artful exploration; the popular perception of words and their corresponding colors helps editorial teams to be aware of their potential audiences and tweak paragraphs accordingly; finally, word-color associations help to achieve mindfulness of individual words and visualize a specific tone that permeates the larger text.

Embedded Biases in Source Dataset
Recent advances in machine learning suggest that computer-aided systems are becoming more human-like in their predictions, and in turn perpetuate human biases. Some learned biases may be beneficial for the downstream application. Other biases can be inappropriate and result in negative experiences for some users. Examples include loan eligibility and crime recidivism systems that negatively impact people of a certain race [11] and resumé sorting systems that believe that men are more qualified to be programmers than women [5].

Similarly, sentiment and emotion analysis machine learning systems can perpetuate and accentuate inappropriate human biases. Recent work has shown that a majority of systems consistently give different emotionality scores to sentences mentioning different races and genders [25], for example, marking near-identical sentences that mention African Americans names as being more angry than those mentioning European American names, and sentences that mention women as being more emotional than sentences mentioning men.

As evident in one of the comments collected during the study, Lexichrome offers opportunities to examine stereotypes that may be embedded in the original dataset. Despite the potential pitfall of preserving and even perpetuating such inappropriate biases, the application also presents an opportunity for its users to discover, examine, reflect their own biases.

Limitations
While the study was designed to capture rich qualitative insights in engaging with a diverse group of participants from different writing-intensive disciplines, we acknowledge that there are some limitations to our findings in their generalizability. Each participant used Lexichrome for a short period of time on a laptop device configured specifically for the study, away from their own usual work environments. While we did not provide specific instructions for text composition, we recognize that this user study may not completely capture editorial and behavioral nuances common in each participant's usual writing process. In the future, we hope to further expand on this work to study how our participants interact with Lexichrome when incorporating it as part of their usual workflow.

CONCLUSION AND FUTURE WORK
Lexichrome is an interactive culmination of numerous visualization techniques that seeks to bridge the gap between lexical semantics and popular opinion on colors. Its interface allows the users to freely explore the comprehensive lexicon of word-color associations, and further contextualize the implications of this dataset by using their own texts. Lexichrome is applicable to a range of academic and creative personnel from various disciplines, including literary analysts, corporate brand managers, and writers as they refer to Lexichrome for quantitative insights or simply second opinions on word-color associations.

Lexichrome also presents a number of ways to improve the existing visualization techniques, including a variant of "literature fingerprinting" to display the "chromatic distribution" of words in a given text [24]. Through our two-part user study, we found that Lexichrome promotes awareness surrounding text composition, and introduce a variety of use cases and features applicable to different disciplines. Future work will include studies of Lexichrome in additional contexts (as outlined in our scenarios) as well as longitudinal studies required to investigate the long-term impact of this type of approach.

A number of future improvements are possible: the application may benefit from implementing a user account system that allows each visitor to contribute to expanding the original lexicon and implement his or her own color palette with a customizable list of words and color associations; the "fingerprinting" algorithm could be improved to present a more context-reflective set of "chromatic fingerprints"; an AI-driven mechanism that suggests editorial changes that promote a more specific chromatic makeup could be deployed; finally, the visitors could potentially apply the generated "chromatic fingerprints" directly on their scanned documents without distorting the spatial information embedded in the original text.

While Lexichrome features a simplified color palette, the presented techniques can be easily applied to other more complex datasets. We recognize opportunities to accommodate more varied color palettes beyond the eleven basic terms, and that Lexichrome could benefit from receiving alternative datasets and visualizations for a more tailored experience.

Initially intended as a lexicon visualization and an exploration of agreement across crowd-sourced opinions, Lexichrome paves the way to applying the word-color association data across a diverse range of text analysis and creation domains. Identifying the semantic qualities of each text may invite creative ways to characterize an author or a genre—and achieve a new level of mindfulness as storytellers visualize their texts.

REFERENCES

[1] Alfie Abdul-Rahman, Julie Lein, Katherine Coles, Eamonn Maguire, Miriah Meyer, Martin Wynne, Chris R Johnson, A Trefethen, and Min Chen. 2013. Rule-based Visual Mappings–with a Case Study on Poetry Visualization. In *Computer Graphics Forum*, Vol. 32. 381–390. DOI: http://dx.doi.org/10.1111/cgf.12125

[2] Aretha B. Alencar, Maria Cristina F. de Oliveira, and Fernando V. Paulovich. 2012. Seeing beyond reading: A survey on visual text analytics. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 2, 6 (Nov. 2012), 476–492. DOI: http://dx.doi.org/10.1002/widm.1071

[3] Brent Berlin and Paul Kay. 1969. *Basic Color Terms: Their Universality and Wvolution*. Univ of California Press.

[4] Natalia Bilenko and Asako Miyakawa. 2014. Visualization of Narrative Structure. http://nbilenko.com/projects/narrative.html. (2014).

[5] Tolga Bolukbasi, Kai-Wei Chang, James Y Zou, Venkatesh Saligrama, and Adam T Kalai. 2016. Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. In *Proc. Advances in Neural Information Processing Systems (NIPS)*. 4349–4357.

[6] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. 2011. D^3 data-driven documents. *IEEE Trans. on Visualization and Computer Graphics* 17, 12 (2011), 2301–2309.

[7] Richard E. Boyatzis. 1998. *Transforming Qualitative Information: Thematic Analysis and Code Development*. Sage Publications.

[8] M. Brehmer, S. Ingram, J. Stray, and T. Munzner. 2014. Overview: The design, adoption, and analysis of a visual document mining tool for investigative journalists. *IEEE Trans. Visualization and Computer Graphics (Proc. InfoVis)* 20, 12 (2014), 2271–2280. DOI: http://dx.doi.org/10.1109/TVCG.2014.2346431

[9] Cynthia A Brewer. 2015. ColorBrewer 2. http://www.ColorBrewer2.org. (2015).

[10] Elia Bruni, Gemma Boleda, Marco Baroni, and Nam-Khanh Tran. 2012. Distributional semantics in technicolor. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, Vol. Long Papers 1. Association for Computational Linguistics, 136–145.

[11] Alexandra Chouldechova. 2017. Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big Data* 5, 2 (2017), 153–163.

[12] Nicholas Diakopoulos, Mor Naaman, and Funda Kivran-Swaine. 2010. Diamonds in the rough: Social media visual analytics for journalistic inquiry. In *IEEE Symp. on Visual Analytics Science and Technology (VAST)*. IEEE, 115–122.

[13] Graham R Gibbs. 2007. Thematic coding and categorizing. *Analyzing Qualitative Data* 703 (2007), 38–56. DOI: http://dx.doi.org/10.4135/9781849208574

[14] Catherine Havasi, Robert Speer, and Justin Holmgren. 2010. Automated color selection using semantic knowledge. In *2010 AAAI Fall Symposium Series*.

[15] Christian Heath, Jon Hindmarsh, and Paul Luff. 2010. *Video in Qualitative Research*. Sage.

[16] P.J. Heather. 1948. Colour Symbolism: Part I. *Folklore* 59, 4 (1948), 165–183.

[17] John Hutchings. 2004. Colour in folklore and tradition–The principles. *Color Research & Application* 29, 1 (2004), 57–66.

[18] Guardian US interactive team. 2013. The state of our union is . . . dumber: How the linguistic standard of the presidential address has declined. http://www.theguardian.com/world/interactive/2013/feb/12/state-of-the-union-reading-level. (2013).

[19] Martyn Jessop. 2008. Digital visualization as a scholarly activity. *Literary and Linguistic Computing* 23, 3 (2008), 281–293.

[20] Domicele Jonauskaite, Betty Althaus, Nele Dael, Elise Dan-Glauser, and Christine Mohr. 2019. What color do you feel? Color choices are driven by mood. *Color Research & Application* 44, 2 (2019), 272–284.

[21] Sepandar D Kamvar and Jonathan Harris. 2011. We feel fine and searching the emotional web. In *Proc. of ACM Int. Conf. on Web Search and Data Mining*. ACM, 117–126.

[22] Paul Kay, Brent Berlin, Luisa Maffi, William R Merrifield, and Richard Cook. 2009. *The World Color Survey*. CSLI Publications Stanford, California.

[23] Paul Kay and Luisa Maffi. 1999. Color appearance and the emergence and evolution of basic color lexicons. *American Anthropologist* 101, 4 (1999), 743–760.

[24] Daniel A Keim and Daniela Oelke. 2007. Literature fingerprinting: A new method for visual literary analysis. In *Proc. IEEE Symp. on Visual Analytics Science and Technology*. IEEE, 115–122.

[25] Svetlana Kiritchenko and Saif M. Mohammad. 2018. Examining Gender and Race Bias in Two Hundred Sentiment Analysis Systems. In *Proc. of the Seventh Joint Conf. on Lexical and Computational Semantics*. New Orleans, Louisiana, 43–53.

[26] Kostiantyn Kucher and Andreas Kerren. 2014. Text Visualization Browser: A Visual Survey of Text Visualization Techniques. In *Proc. IEEE Conf. on Information Visualization, poster session*.

[27] Lauren I Labrecque and George R Milne. 2012. Exciting red and competent blue: The importance of color in marketing. *Journal of the Academy of Marketing Science* 40, 5 (2012), 711–727.

[28] Creative Market Labs. 2019. COLORlovers. http://www.colourlovers.com. (2019).

[29] Muyueh Lee. 2013. Green Honey. http://muyueh.com/greenhoney/. (2013).

[30] Nicholas Lezard. 2009. Scientific study of 'literary fingerprinting' reveals only the bleeding obvious. http://www.theguardian.com/books/booksblog/2009/dec/23/literary-fingerprinting-obvious. (2009).

[31] Albrecht Lindner, Nicolas Bonnier, and Sabine Süsstrunk. 2012. What is the color of chocolate?–extracting color values of semantic expressions. In *Conference on Colour in Graphics, Imaging, and Vision*, Vol. 2012. Society for Imaging Science and Technology, 355–361.

[32] Saif Mohammad. 2011a. Colourful language: Measuring word-colour associations. In *Proc. of the 2nd Workshop on Cognitive Modeling and Computational Linguistics*. Association for Computational Linguistics, 97–106.

[33] Saif Mohammad. 2011b. Even the abtract have color: Consensus in word-colour associations. In *Proc. of the Annual Meeting of the Association for Computational Linguistics*. 368–373.

[34] Randall Munroe. 2009. xkcd: Movie Narrative Charts. https://xkcd.com/657/. (2009).

[35] Brendan O'Connor. 2008. Where does "Blue" end and "Red" begin? http://www.crowdflower.com/blog/2008/03/where-does-blue-end-and-red-begin. (2008).

[36] Li-Chen Ou, M Ronnier Luo, Andrée Woodcock, and Angela Wright. 2004. A study of colour emotion and colour preference. Part I: Colour emotions for single colours. *Color Research & Application* 29, 3 (2004), 232–240.

[37] W Bradford Paley. 2002. TextArc: Showing word frequency and distribution in text. In *Proc. IEEE Symp. on Information Visualization (posters)*, Vol. 2002.

[38] Martin F Porter. 2001. Snowball: A language for stemming algorithms. http://snowball.tartarus.org/texts/introduction.html. (2001).

[39] Stephen Ramsay. 2007. Algorithmic criticism. In *A Companion to Digital Literary Studies*, Ray Siemans and Susan Schreibman (Eds.). Blackwell.

[40] Geoffrey Rockwell, Stefan Sinclair, and Kirsten C. Uszkaloand Milena Radzikowska. 2006. TAPoR - Text Analysis Portal for Research. http://www.tapor.ca/. (2006).

[41] Peter Mark Roget. 1911. *Roget's Thesaurus of English Words and Phrases*. TY Crowell Company.

[42] Vidya Setlur and Maureen C Stone. 2015. A linguistic approach to categorical color assignment for data visualization. *IEEE Trans. on Visualization and Computer Graphics* 22, 1 (2015), 698–707.

[43] Satyendra Singh. 2006. Impact of color on marketing. *Management Decision* 44, 6 (2006), 783–789.

[44] Liz Stinson. 2013. Infographics: The Colors Mentioned Most in 10 Famous Books. http://www.wired.com/2013/09/data-visualization-the-color-signatures-of-famous-book. (2013).

[45] Yuzuru Tanahashi and Kwan-Liu Ma. 2012. Design considerations for optimizing storyline visualizations. *IEEE Trans. on Visualization and Computer Graphics* 18, 12 (2012), 2679–2688.

[46] John C Tang, Sophia B Liu, Michael Muller, James Lin, and Clemens Drews. 2006. Unobtrusive but invasive: Using screen recording to collect field data on computer-mediated interaction. In *Proc. Conf. on Computer-Supported Cooperative Work*. ACM, 479–482.

[47] M. Tennekes and E. de Jonge. 2014. Tree Colors: Color Schemes for Tree-Structured Data. *IEEE Trans. on Visualization and Computer Graphics* 20, 12 (Dec 2014), 2072–2081.

[48] Alice Thudt, Uta Hinrichs, and Sheelagh Carpendale. 2012. The Bohemian Bookshelf: Supporting serendipitous book discoveries through information visualization. In *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*. ACM, 1461–1470.

[49] Johnny Waldman. 2012. Visualizing the meals in Haruki Murakami's 1Q84. Spoon & Tamago. (2012).

[50] Martin Wattenberg. 2005. Color Code: A color portrait of the English language. http://www.bewitched.com/live/colorcode/index.html. (2005).

[51] T.W. Whitfield and T.J. Whiltshire. 1990. Color psychology: A critical review. *Genetic, Social, and General Psychology Monographs* (1990).

[52] Lisa Wilms and Daniel Oberfeld. 2018. Color and emotion: effects of hue, saturation, and brightness. *Psychological Research* 82, 5 (2018), 896–914.

[53] Wordnik. 2009. Wordnik Developer. http://developer.wordnik.com/. (2009).