

Lexichrome: Examining Word-Color Associations with Visualization

Chris K. Kim*

University of Ontario Institute of Technology

Christopher Collins**

University of Ontario Institute of Technology

ABSTRACT

Based on a comprehensive word-color association lexicon, we present *Lexichrome*: a web-based application that offers its visitors an ability to browse a catalogue of visual relationships between words and colors. Additionally, the application seeks to allow literary scholars, brand managers, and writers to further examine a trend in the popular notion of word-color associations through user-provided texts.

1 INTRODUCTION

Color evokes a response, describes an object, and shapes a culture. Color has historically been a significant part of human lives: Colors tell stories about those who exploit them, 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 [1]; in particular, emotional color associations are widely used in brand design [2].

The relationship between words and colors is a fluid and dynamic one, influenced by time, geography, and culture. Shakespeare, in his numerous plays, describes color green as both a positive and negative symbol for youth and innocence, and occasionally decay and envy respectively [3]; and this duality holds true even in today's context: some associate color green with money and greed, while others consider it a symbol for life, environment, and springtime.

Many recognize the linguistic and symbolic importance of color—particularly in product marketing, web design, and information visualization—and strive to reap the benefits of further strengthening the message and triggering the desired emotional response from potential audiences.

However, with the possibility of accidentally eroding the original message with improper use of color, there has been a lack of reliable lexicons containing word-color associations [4]. This gap has recently been addressed by National Research Council Canada's (NRC) comprehensive lexicon. The lexicon is comprised of more than 25,000 unique associations, and each entry contains a primary word and the name of the associated color.

Lexichrome (<http://lexichrome.com>), whose name is a portmanteau of “lexicon” and “chrome” (derived from the Greek word “chroma”), is an application that seeks to map the NRC dataset to a web-based catalogue that users can browse and analyze. In addition, *Lexichrome* allows its visitors to view the “chromatic makeup” of a user-provided text by applying the associated colors to each relevant word token. This application is an exercise in appropriation of linguistic dataset and implementation of user-centered design techniques in a scholarly context.

* chris.kim@uoit.ca

** christopher.collins@uoit.ca

2 BACKGROUND AND INSPIRATION

As an interactive visualization of popular opinion surrounding word-color associations, *Lexichrome* appropriates a number of graphical techniques, motifs, and datasets.

The application primarily sets out to visualize the data in the word-color association lexicon, created and published by Mohammad. Produced using Amazon's crowdsourcing platform Mechanical Turk, the lexicon contains numerous words that may be associated with multiple colors in varying degrees. For instance, the term “caution” is popularly linked to both colors red and yellow, but may be more strongly associated with yellow based on participant votes.

In order to illustrate this array of associative data, *Lexichrome* relies on a series of traditional visualization techniques, including stacked bar chart, treemap, and color maps, and looks to other predecessors for inspiration. Harris' visualization project “We Feel Fine” parses and 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 [5]; Wattenberg's “Color Code” establishes an interactive map of more than 30,000 nouns, each equipped with a color based on the average color of image search results [6].

Finally, Keim and Oelke's “literature fingerprinting” approach to literary analysis serves as a visualization method for displaying color features of each word—as individual entities or as part of a longer text [7].

3 DESIGN AND DEVELOPMENT

Lexichrome is a web-based application that utilizes various server-side and client-side technologies to parse, process, and display the word-color lexicon.

PHP is used to access the lexicon database and process user queries, and Python and NLTK are used on an adhoc basis to clean up each instance of user-provided texts before querying the database for word-color associations [8].

In the front-end realm, a series of code libraries have been used to facilitate the development process. In addition to the jQuery library, D3.js has been implemented in order to quickly generate interactive visualizations of word-color data [9].

3.1 Bidirectional Exploration: *Palette* and *Words*

Each visitor to *Lexichrome* is invited to explore the application's database in two ways: interacting with an eleven-color palette and searching for a particular word.

Palette view, as illustrated in Figure 1, allows the users to see



Figure 1: Frequencies of color associations in the NRC dataset through as a treemap

herbal 8 out of 9	remittance 8 out of 9	mint 15 out of 17	frog 14 out of 16
habitat 7 out of 8	meadow 7 out of 8	monetary 7 out of 8	pay 7 out of 8
wealth 9 out of 11	plantation 8 out of 10	abundance 8 out of 9	cost 12 out of 15

Figure 2: List of English terms associated with green, sorted in order of relevance

the distribution of word-color association frequencies in the database using treemap visualization. Each tile contains a color, numerical statistics, as well as a random sample of words associated with that particular color.

Upon clicking a tile, the user can view the complete list of associated words, as seen in Figure 2. Each of the small bar charts illustrates the strength and the relevance of each word-color association: For instance, if the word “frog” has been labelled as “14 out of 16” in color green, it means that the other two survey participants did not agree with the association between “frog” and color green—ultimately resulting in a slightly weaker but nonetheless dominant association between the two.

The *Words* view functions as a queryable dictionary of word-color associations, as it allows visitors or analysts to search for multiple terms and the subsequent list of associated colors. By hovering over each item in the search results, each 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 ratio of different colors: if the word “damage” displays red and black with the red bar significantly longer than the other, it means that the majority of the original survey participants agree that the word is associated with color red.

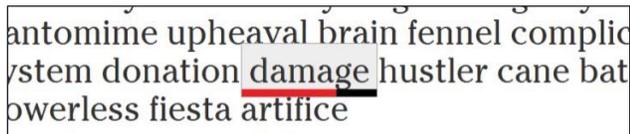


Figure 3: Layer displaying the colors associated with the term “damage” in *Words* view

3.2 Text-Color Analysis through “Fingerprinting”

In addition to allowing the visitors and analysts to explore the NRC word-color dataset, Lexichrome has an ability to process user-provided texts and extract the relevant colors.

As shown in Figure 4, Lexichrome allows writers, marketing and PR professionals, and literary linguists to paste their own texts and view their “chromatic makeup”: this view reveals what color associations a text will evoke in readers. Upon pressing “Analyze,” the application sends the provided text to NLTK which processes it using normalization, lemmatization, and function word removal and extracts an array of lexical (content) words. This array is then matched to make associations to the relevant colors, whose results are subsequently displayed using a variant of the “fingerprinting” motif as described [7].

This associative process of transposing words to colors yields interesting results: Edgar Allan Poe’s somber narrative poem “The Raven” contains more instances of black and gray colors than Shakespeare’s love poem; the first line from Albert Camus’ “The Stranger” an interesting mix of colors pink and black, as the text contains references to the protagonist’s mother and her untimely death; Wikipedia’s article about “agriculture” yields a predominantly green chart, as the article contains words including development, climates, and farming—all associated with color green. Coca Cola’s company mission statement results in a mix of colors, led by white, with their company signature color red being sixth.

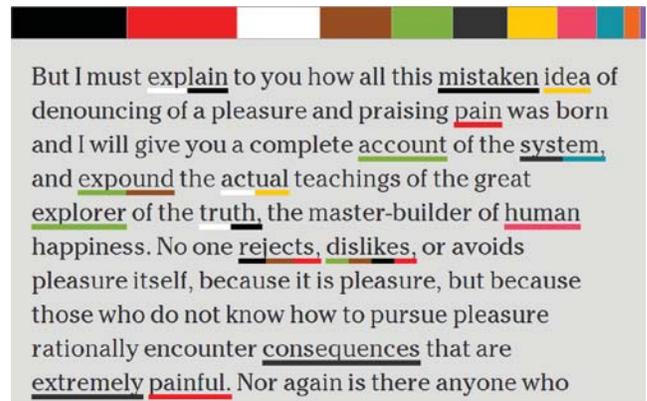


Figure 4: “Chromatic fingerprint” of a sample text based on word-color associations

Positive terms such as happiness, accountable, and passion all associate with red and appear in the Coca-Cola brand statement.

4 DISCUSSION

Lexichrome is an interactive culmination of numerous visualization techniques that seeks to bridge the gap between lexical semantics and popular notion of colors. Its web-based interface allows the users to freely explore the comprehensive lexicon of word-color associations, and further contextualize the implications of this dataset by “fingerprinting” their own texts. Lexichrome will be applicable to a range of academic and creative personnel, including literary linguists, corporate brand managers, and writers; the application will continue to improve as additional visualization techniques and interactive components are implemented in the future.

ACKNOWLEDGMENTS

Funded by the Natural Sciences and Engineering Research Council of Canada (NSERC), Lexichrome was created with the support of NRC’s Research Officer Saif Mohammad.

REFERENCES

- [1] Hutchings, J. (2004). Colour in folklore and tradition - The principles. *Color Research & Application*, 29(1), 57–66.
- [2] Ciotti, P. (2014). The psychology of color in marketing and branding. *Entrepreneur*. Retrieved from: <http://www.entrepreneur.com/article/233843>.
- [3] Heather, P. (1948). Colour Symbolism: Part I. *Folklore*, 59(4), 165–183.
- [4] Mohammad, S. (2011). Colourful language: Measuring word-colour associations. *Proceedings of the 2nd Workshop on Cognitive Modeling and Computational Linguistics*.
- [5] Harris, J. (2006). We Feel Fine. Retrieved from <http://wefeelfine.org/>
- [6] Wattenberg, M. (2005). Color Code. Retrieved from <http://www.bewitched.com/live/colorcode/index.html>
- [7] Keim, D. a., & Oelke, D. (2007). Literature Fingerprinting: A New Method for Visual Literary Analysis. *2007 IEEE Symposium on Visual Analytics Science and Technology*, 115–122.
- [8] Bird, S. (2006). NLTK: the natural language toolkit. *Proceedings of the COLING/ACL on Interactive presentation sessions*, 63–70.
- [9] Bostock, M., Ogievetsky, V., & Heer, J. (2011). D³: Data-Driven Documents. *IEEE Transactions on Visualization and Computer Graphics*, 17(12), 2301–9.