

The Language of Western Beauty in Online Women's Magazines

Cassandra Kane

May 6th, 2020

Research Questions

This project explores the language surrounding western female beauty culture in modern online media. This project focuses specifically on how online language surrounding beauty in women's magazines has evolved over the past decade, including changes in trends, topics, and vernacular. This project also examines how advertising may affect online language centered around beauty. Online media brands often maintain their popularity by appearing reliable and honest with their readers, but they may also accept lucrative brand deals to promote certain products. Usually these media channels [claim](#) to only promote products that the editors genuinely enjoy themselves, but this project investigates how online magazine language may shift when an organization intends to sell a product while still maintaining an authentic-feeling connection with their consumers. All code relating to this project may be found [here](#), and the preliminary results of this project may be found [here](#).

The relationship between beauty, language, and advertising in women's magazines has been previously researched through many academic lenses. A 2014 article published in the *Journal of Intercultural Communication Research* titled "The Globalization of Beauty: How is Ideal Beauty Influenced by Globally Published Fashion and Beauty Magazines?" specifically

examines differences in the cover stories of women's magazines across different countries, focusing on four international publications – *Elle*, *Cosmopolitan*, *Vogue*, and *Glamour* – over a 36-month period (October 2007 to October 2008). The study discovered that the topics discussed in the cover stories for these magazines were dominated by western ideals and therefore relatively similar across countries, but the imagery within these cover stories remained more unique to the culture of each region.¹ Another 2017 study published in the *Tell-Us Journal* titled “Analysis of Linguistic Features of Beauty Product Advertisements in Cosmopolitan Magazine: A Critical Discourse Analysis” hand-picked a sample of ten *Cosmopolitan* product advertisements and qualitatively examined the linguistic features of each, noting the use of imperative speech, modalities, targeted pronouns, and adjectives with positive connotations.² The study details how each of these linguistic techniques can effectively persuade consumers towards purchasing an advertised product.

This project aims to build upon and distinguish itself from this prior research in a number of ways. First, this project studies online women's magazines over a broader period of time (2012 to 2020) and examines how these magazines have evolved within the past decade. Second, this project focuses on all of the online beauty articles published by these magazines as opposed to their cover stories or advertisements exclusively, providing a more exhaustive overview of how these publications discuss and portray beauty online. Third, this project focuses on beauty content created for primarily western audiences (i.e. the American websites for each of these magazines). While some of these differences significantly broaden the scope of the project and

¹ Yan Yim and Kim Bissell, “The Globalization of Beauty: How Is Ideal Beauty Influenced by Globally Published Fashion and Beauty Magazines?” *Journal of Intercultural Communication Research*, vol. 43, no. 3, May 2014, pp. 194–214

² Zainal Arifin Renaldo, “Analysis Of Linguistic Features Of Beauty Product Advertisements In Cosmopolitan Magazine: A Critical Discourse Analysis.” *Tell-Us Journal*, vol. 3, no. 2, Sept. 2017, pp. 141–154

therefore limit the depth of the analysis that can be performed (such as the qualitative methodology described in the *Tell-Us Journal* study, which would prove time-consuming to manually execute for over ten thousand articles), the study ultimately benefits from a larger, more comprehensive data set given the research questions driving this project.

Data Sources

The online women's magazines examined for this project were *Elle*, *Cosmopolitan*, and *Seventeen*. Articles were considered to be beauty-centric if they were listed under the "beauty" section of their magazine's website. While each of these magazines targets a female audience, they do differ somewhat in content and style. *Elle* – [self-described](#) as "the world's largest fashion magazine and media brand inspiring women" – produces [beauty content](#) that consists of tutorials, product reviews, celebrities news, and general beauty resources with a relatable and relatively professional tone; examples of *Elle*'s online beauty headlines include "The 12 Best Natural Face Moisturizers For Skin That Glows" and "Everything You Should Know Before Getting A Dip Powder Manicure." *Cosmopolitan* – [self-described](#) as "the biggest young women's media brand in the world" – produces similar [beauty content](#) as *Elle* but targets more of a millennial audience while employing casual, humorous language with headlines such as "Um, Yeah, Hi, Your Lip Balm Actually Needs SPF." *Seventeen* – [self-described](#) as "the ultimate content destination for Gen Z and young millennials" – uses similarly casual and relatable language in its [beauty content](#) but focuses on more teenage-specific topics such as prom and acne.

This specific collection of magazines was selected based on a number of reasons, the first being their diverse styles. Second, these magazines are well-established and widely read in the

United States; *Elle*, *Cosmopolitan*, and *Seventeen* have existed for decades and still remain prominent media sources for young women, particularly in online spaces. Third, complete beauty article archives are accessible for each of the selected online publications, while other popular women's magazines like *Vogue* and *Glamour* do not provide a way for users to comprehensively view all of their beauty content. Fourth, each of these magazines shared the same website layout (HTML structures, tags, etc), which allowed for information from these websites to be scraped more efficiently. The most likely reason as to why these publications share the same website layout is that *Elle*, *Cosmopolitan*, and *Seventeen* are all owned by the same mass media conglomerate [Hearst Communications](#), which may have a single organization developing the websites for all of its asset media brands.

Data Collection and Management

Article data from the beauty sections of *Elle*, *Cosmopolitan*, and *Seventeen*'s websites were acquired using a [Python web scraper](#). The web scraper built for this project relied on [BeautifulSoup](#) and [Selenium](#) to traverse the contents of a given webpage and to gather information for a specified number of articles determined by the user, storing data such as the article's headline, authors, sub-headline, publication date, parent category, thumbnail, sponsor if applicable, and a link to the full article. The [resulting data](#) scraped from the [beauty section](#) of *Elle*'s website consisted of 6999 articles published between September 29th, 2009 and April 13th, 2020. The [resulting data](#) scraped from the [beauty section](#) of *Cosmopolitan*'s website consisted of 5999 articles published between November 4th, 2008 and April 13th, 2020. The [resulting data](#) scraped from the [beauty section](#) of *Seventeen*'s website consisted of 6000 articles

published between May 18th, 2007 and April 13th, 2020. All of the data described above was later filtered based on year in the analysis process. [Data](#) was also collected from the [grooming section](#) of the men's magazine *Esquire* (which was the closest equivalent to the beauty content produced by women's magazines), but this information was not analyzed or explored for the current version of this project.

After this raw data was scraped, each article headline was [lemmatized](#) and stored in an [updated dataset](#) using a [Python program](#) that relied on the lemmatizing tools within the [Python Natural Language Toolkit](#) library. Prior to lemmatization, all of the words within a headline were converted to lowercase, and the punctuation within each word was removed with the exception of hyphens, which were replaced by spaces. This process of text cleaning was imperfect in the lemmatization process, especially considering the unique syntactic styles of modern online magazine headlines. For instance, the *Cosmopolitan* headline “This Rubber-Band Ponytail Is E-v-e-r-y-t-h-i-n-g” would correctly have “rubber” and “band” lemmatized as separate words by the program, but the word “everything” would not be lemmatized correctly since each of its letters would be separated apart. Abbreviated slang such as “OMG” would also not be correctly lemmatized by the Python program implemented. Despite these imperfections, lemmatizing these headlines allowed for more meaningful analysis to be performed in later steps of the research process.

Exploratory Data Analysis

The exploratory data analysis for this project consisted of four parts: examining all of the online beauty articles published over the past decade, examining the general subjects of these

articles, examining word frequencies within these articles, and examining how these frequencies changed depending on whether the article was sponsored.

For the first part of the exploratory data analysis process, a histogram was built to measure the number of online beauty articles published over the past decade, either for an individual publication or for all of the publications collectively. While the number of articles published in a given month typically ranged from 20 to 100, the histogram also displayed a sharp spike in the number of articles published during June 2011 for *Elle* (349 articles) and during July 2010 for *Seventeen* (324 articles). Further investigation revealed that 308 of the 349 articles published by *Elle* in June 2011 were published on the same day (June 14th, 2011), and 301 of the 324 articles published by *Seventeen* in July 2010 were published on the same day (July 13th, 2010). These findings suggested that *Elle* and *Seventeen* may have uploaded a backlog of previously written material to their websites on a single day, especially considering that these spikes occurred in the early 2010s when many popular publications were still transitioning to online formats. Therefore, given that the content of those articles may not have been representative of the year in which they were published, the time period examined for this project was narrowed down to be between January 2012 and January 2020.

For the second part of the exploratory data analysis process, the general subjects of these online beauty articles were examined. An article's subject was determined based on the parent category the article belonged in. However, this exploratory analysis revealed how each magazine differed greatly in their categorizations. *Elle* classifies their online beauty content as either "Hair," "Makeup & Skin," "Health & Fitness," or general "Beauty." *Seventeen* classifies their online beauty content as either "Hair," "Makeup & Skincare," "Nails," "Celeb Beauty," "Beauty

Quizzes,” or general “Beauty.” *Cosmopolitan* does not categorize their online beauty content at all. These discrepancies made comparing beauty subjects between these magazines significantly more difficult. For instance, content relating to hair could not be compared across magazines because *Cosmopolitan* does not explicitly label certain beauty articles as relating to hair, and there existed no other convenient means of determining whether a given article was about hair (articles could have been searched for keywords like “hair” or “braid,” but this would not have been a foolproof means of categorization). Another issue was that while *Elle* classifies “Health & Fitness” articles as a part of their beauty content, *Seventeen* contains their health and fitness articles in a separate section of their website, making it difficult to determine which magazine may discuss a certain subject more often than the others; *Elle*’s inclusion of health and fitness articles in their beauty content reduces the relative prominence of other beauty subjects. Given these difficulties, there was no meaningful way to examine any individual subject based on the inconsistent categorization of articles across these online women’s magazines. Another approach to determining the subject of a given article could be considered in a future iteration of this project.

The third part of the exploratory data analysis process involved examining word frequencies within the headlines of online beauty articles. In order to investigate these frequencies, a [visualization tool](#) was built that would construct an area chart based on a word entered by the user, measuring the proportion of headlines containing that word for a given month between January 2012 and January 2020. The curves calculated for these area charts were [cubic splines that preserved monotonicity](#), meaning that each curve continuously passed through every given data point without interpolating points that were more extreme than the data

provided. In other words, every measured proportion was charted in the curve, and the maximum and minimum proportions indicated by the area chart existed in the original data itself. This visualization tool also allowed for data to be filtered by publication, for frequencies to be measured by either proportion or count, and for a word to be searched within either the original or lemmatized version of a headline. With this tool, the evolution of different word frequencies could be visualized more quickly, and meaningful trends could be recognized more easily.

For the fourth part of the exploratory data analysis process, word frequencies were examined based on whether an article was sponsored. Using [this Python program](#), the number of sponsored articles for each magazine was first determined in order to ensure a large enough sample size for evaluation; of the online beauty articles published between January 2012 and January 2020, 236 of the 5574 articles published by *Elle* were sponsored (~4.2%), 282 of the 4693 articles published by *Cosmopolitan* were sponsored (~6.0%), and 102 of the 3967 articles published by *Seventeen* were sponsored (~2.6%). Given the sufficient quantity of sponsored articles for each publication, the program then evaluated for a given word the proportion of sponsored article headlines containing that word and the proportion of unsponsored article headlines containing that word. These proportions and the factor by which a sponsored headline was more likely to feature a given word than an unsponsored post (either for a specific publication or for all of the publications collectively) were then stored to [an output file](#). This tool allowed for convenient testing of whether certain words were more likely to appear in beauty article headlines for sponsored content.

Preliminary Findings

The data visualizations accompanying these preliminary findings may be located [here](#) (please note that these visualizations may take a moment to render and are best viewed on a desktop screen). The first preliminary finding focused on the number of online beauty articles published each month by *Elle*, *Cosmopolitan*, and *Seventeen* magazine between January 2012 and January 2020, as illustrated by this [histrogram overview](#). Collectively, these publications experienced a steady growth of online beauty content published from 2012 through 2015, followed by an overall decline from 2016 through 2019. This trend is apparent for *Elle* and *Seventeen* individually as well, while *Cosmopolitan* shows multiple peaks around 2013, 2016, and 2020. Although more online beauty content was published during these periods, it is not evident whether beauty content also made up a higher proportion of these magazines' articles or whether more online content was being published overall. To determine which may be the case, all of the online articles published by these magazines between January 2012 and January 2020 would have to be scraped, a task which could be pursued in a future iteration of this project.

A number of other preliminary findings focused on word frequencies within online beauty headlines over time, using the [word frequency visualization tool](#) previously described. For instance, the usage of “you” (and other second-person pronouns like “your” and “yours”) in headlines for online beauty articles was found to have increased overall between January 2012 (~2.1% of headlines) and January 2020 (~23.3% of headlines) with a minor decline during 2017, as illustrated in [this area chart visualization](#). This trend was also observed for each publication separately, indicating that no individual magazine skewed the overall results. This increase in the

usage of second-person pronouns suggests that forming a sense of connection with their readers has become more of a priority for online women's magazines within the past decade, as well as a growing emphasis on the idea that a magazine's beauty content is personally pertinent to the readers themselves.

Additionally, the usage of "how" in online beauty article headlines experienced a sharp increase in mid-2013 (peaking at ~25.5% of headlines in October 2013) followed by a steady decline after 2015 (~5.3% of headlines in January 2020), while the usage of "need" (and its conjugations) experienced a steady increase from January 2014 (~2.5% of headlines) through January 2020 (9.0% of headlines), as illustrated in [these area chart visualizations](#). Similar trends were also observed for the publications individually. While the inclusion of "how" or "need" in an article's headline may not be completely indicative of an article's content, the "how" headlines suggest that those articles contain some kind of explanation or instruction, while the "need" headlines imply more urgency, whether it be for information or for a product. This decline in "how" online beauty content and growth in "need" over the past eight years suggest that online women's magazines have become more focused on grabbing the attention of readers as opposed to providing tutorials or explanations of beauty trends and techniques.

The remaining preliminary findings involved examining the likelihood of certain words appearing in the headline for an online beauty article based on whether the article was sponsored, and the data relating to these findings may be located [here](#). All of the words examined were confirmed to have appeared in at least three headlines for sponsored online beauty articles. Given the proportions and ratios calculated by [the Python program](#) previously described, the occurrences of "easy," "perfect," and "you" proved to have notable results. Specifically,

occurrences of “easy” and “perfect” appeared more than twice as often in headlines for sponsored online beauty articles than for unsponsored content, likely emphasizing the convenience of a given product and the beauty ideals the product can help a consumer achieve. Occurrences of “you” (and other second-person pronouns) appeared almost twice as often in sponsored content than in unsponsored content, likely emphasizing how a product may personally relate to the magazine's reader. Similar patterns were also recognized for each publication separately, indicating that no individual magazine skewed the overall results. The significance of these results could be further evaluated in a future iteration of this project based on whether the evolving popularity of a word corresponded with changes in the number of sponsored articles published over time (which would affect the likelihood ratios described above), as well as on how the occurrences of certain words compare between headlines for sponsored and unsponsored articles that focus specifically on products, given that an article’s topic may be a confounding variable when examining the language employed in its headline.

Future Work

There are many ways to move forward with this project. One aspect of this research that could be expanded is within the data collection process, where all of the online articles published by these women’s magazines could be scraped in order to better understand how the prominence of online beauty content has changed over the past decade. Scraping all of a magazines’ online content does present its own challenges, however, such as the computational limits of the standard personal computer (reliance on a third-party cloud computing service such as [AWS](#) would mostly likely be necessary) and whether a complete archive of articles for all sections of a

magazine's website is available (the [opinion section](#) of *Cosmopolitan*, for example, does not currently present a way for users to traverse all of its online opinion pieces). Data could also be collected from more online women's magazines in addition to *Elle*, *Cosmopolitan*, and *Seventeen* if comprehensive article archives could be located.

The lemmatization process for this data could also be improved in a future iteration of this project. As previously mentioned, the lemmatization process currently implemented does not account for the unique syntactic styles of modern online article headlines or for the growing use of abbreviated slang. Solutions to these issues could involve further testing of text cleaning techniques to determine which methods produce the best results, as well as using [slang dictionaries](#) for mapping abbreviated terms to English phrases. Other textual information could be lemmatized and examined as well, such as sub-headlines or the full text of a given article.

Additional computational linguistic processes could also be implemented like [sentiment analysis](#) or [compression algorithms for detecting repetition](#) to further evaluate the language of online magazine beauty content. These techniques in particular could more concretely build upon the findings from prior research on language within women's magazines (such as the *Tell-Us Journal* study previously described) by detecting repeated uses of phrases with positive connotations. Reliance on other [natural language processing libraries](#) could allow for more in-depth analyses relating to prior research as well, for instance by examining verb conjugations and linguistic modalities in article headlines.

In terms of expanding the findings from this project, each of the results involving word frequencies within article headlines could be studied in relation to more variables other than the article's publication, publish date, and sponsors. Additional factors could include the article's

topic (whether it focuses on a particular brand, celebrity, trend, etc), format (such as listicles or galleries), and thumbnail; all of these variables could also be examined on their own. Data collected from online women's magazines could be compared against the language used in popular online men's magazines such as *Esquire*, where data has already been scraped for this project but not yet analyzed. Although the websites for many men's magazines do not contain their own beauty sections, language could be compared across similar categories like [grooming](#). All of these results could be more effectively communicated in this [preliminary site](#): the findings relating to sponsorships could be visualized as bar charts, and the other visualizations could be better annotated and organized. The site could also be heavily redesigned as a visual essay focused on digital story-telling (inspired by the type of work done by [The Pudding](#) and other data visualization publications), guiding a reader through the results of the project with a more coherent structure and engaging format. Lastly, more rigorous tests of statistical significance could be performed to confirm whether the trends perceived in these preliminary findings were accurate or meaningful.

American women's magazines and media brands are only one type of online space where beauty is discussed and portrayed; future projects could investigate all sorts of online environments where beauty content is prominent including social media platforms like Instagram or entertainment services like YouTube. As of now, this project demonstrates only a small fraction of the work that can be done in this research field.

Sources Cited

- Renaldo, Zainal Arifin. "Analysis Of Linguistic Features Of Beauty Product Advertisements In Cosmopolitan Magazine: A Critical Discourse Analysis." *Tell-Us Journal*, vol. 3, no. 2, Sept. 2017, pp. 141–154., doi:10.22202/tus.2017.v3i2.2628.
- Yan, Yan, and Kim Bissell. "The Globalization of Beauty: How Is Ideal Beauty Influenced by Globally Published Fashion and Beauty Magazines?" *Journal of Intercultural Communication Research*, vol. 43, no. 3, May 2014, pp. 194–214., doi:10.1080/17475759.2014.917432.