SOCIAL BEHAVIOR RESEARCH AND PRACTICE

Open Journal

Openventio
PUBLISHERS

Brief Research Report

What Were They Thinking? Analytic and Cognitive Language in Instagram Captions

Sheila Brownlow, PhD[1*]; Makenna Pate, BA[2]; Abigail Alger, BA [Student][2]; Natalie Naturile, BA[2]

[1]Department of Psychology, Catawba College, Salisbury NC 28144, USA

Corresponding author
Sheila Brownlow, PhD
Professor and Chair, Department of Psychology, Catawba College, Salisbury NC 28144, USA; E-mail: sbrownlo@catawba.edu

ABSTRACT

Background
We examined content and expression of Instagram captions of major celebrities who differed according to sex and status, with a focus on determining whether these variables influenced the use of analytic language and cognitive content.

Method
Instagram captions (n=942) were analyzed with the linguistic inquiry and word count (LIWC), which delineated percentage of language reflecting analytical thought and various cognitive mechanisms, such as causality and discrepancy.

Results
Men and low-status persons used more functional analytic language, demonstrating critical thought; in contrast, high-status celebrities showed more causality. Women more than men "qualified" their speech with discrepancy. These findings were not a function of sentence length.

Conclusion
Status increased the tendency to construct and explain, perhaps because higher status celebrities (particularly women) knew that they could hold followers' attention with complex content. The tendency to write captions that were concrete was seen in those lower-status persons who may have perceived that followers would not wade through a lot of complicated thoughts. Thus, status contributes to the manner in captioning based, perhaps, on having a broader audience willing to read more complex language.

Keywords
Language use; Status; Instagram; Sex differences in linguistics.

INTRODUCTION

A simple Google search of "sample Instagram post" leads to 276,000,000 results, confirming that social media is both important to people and perhaps more complicated than it would appear on the surface. Given that Instagram is designed to share visual content, why is there so much advice about what words to use to caption Instagram posts?

A wealth of research shows that words do, indeed, matter. Several avenues for the analysis of linguistic content are available to understand how people use language based on several personal and social variables. For example, language can be classified into naturally-occurring groupings that reflect a user's current affective state, as well as more enduring traits, *via* an "open" vocabulary analysis approach.[1] The "closed" approach, exemplified by the linguistic inquiry and word count (LIWC),[2] analyzes content words and speech devices by categorizing these into Adictionary of nearly 6400 different words, including text and internet-related abbreviations for words. The LIWC automatically produces more than 90 dependent measures per file of text; most of these measures take the form of a percentage. So, for example, the LIWC will compute the percentage of analytic language, or pronouns, or language of cognitive processes in each file.[2] The

Openventio

PUBLISHERS

most recent LIWC dictionary has been generated by the analysis of over 80,000 separate speakers and writers who have produced over 230 million words.[2] These samples have been taken across several contexts, from personal expressive writing to the Internet and social media postings, to books and plays, to news articles and speeches.[2] The closed approach using the LIWC has revealed ways that language reflects a speaker's motivation, state, sex, status, and ways-of-thinking.

Emotional (affective) language, for example, signals important aspects of the speaker, including sex and status. Women more than men use language with positive emotion,[1,3] as do those high in extraversion,[4] conscientiousness,[5] and agreeableness.[6,7] Not surprisingly, celebrities using Twitter most often discuss their preferences,[8] yet affect language among celebrities on Twitter also differs, with positive emotion seen in less "followed" (i.e., lower social status) celebrities, regardless of sex,[9] perhaps because those persons wanted to be seen as agreeable, and light—therefore, more fun to follow.

Status and sex also influence the use of the function words that mirror a focus on the self, as seen *via* the first-person pronoun "I". On Twitter, for example, women and lower-status persons used "I" in their Tweets, but "I" was equal (and lesser-likely) in both women and men of higher social status.[10] The use of an informal and friendly style that includes self-referencing is seen in men of lower social status, perhaps in effort to engage more people and enhance their own status through more self-referencing. The use of "I" signals a more "narrative" or "story-like" approach to provide perspective, omitting sophisticated language, and avoiding cognitive mechanisms.[11]

Language signaling thought, causality, and insight is used in predictable circumstances, such as when people wish to transmit facts or reconstruct events and provide explanations for them.[12] Such use may also suggest that a story or argument is well-known because it does not contain hedges and fillers. So, for example, cognitive processes are reflected when people use words such as "know" or "because". Discrepancy and tentativeness are also markers of thought, and may be seen when people use "should" or "maybe". Differentiation may qualify what is said ("but"), although certainty may reinforce what is said (i.e., "always").[2] People show more cognitive sophistication in their language as they age,[13] and men more than women use prepositions and articles across different types of communication contexts,[14] although women employ discrepancies (should, could) more than men in most communication contexts.

Analytic thought as expressed in language is decreasing as technology, issue-complexity, and global political shifts toward populism are increasing.[15] Analytic language includes the use of articles (e.g., the, and) and prepositions (e.g., above, with) designed to show connections and critical relations among points. On the other hand, the "informal" style includes more narrative, ideas, actions, and stories.[15] Although shifts away from analytic language are occurring most acutely in the political realm, other common

textual information (such as in movies and on television) also trends toward informality.[15] Thus, linguistic simplicity may be a key marker of the speech of persons in the public eye. Our study examined the analytic expression, as well as evidence of cognitive thought, in speech of celebrities who have the attention of the public, in this case by examining their language used on a specific type of social media: Instagram.

Instagram is a social media platform that includes text captioning along with the primary focus of sharing visual content (photos and videos). People post videos or photos and write text to "caption" their visual posting. According to https://www.web-sitehostingrating.com/instagram-statistics/,[16] Instagram posts garner more reactions or responses (termed "engagements") than any other social media site—58 times more than Facebook. Moreover, 1 billion people worldwide use Instagram, and 38% of users in the US visit multiple times a day, spending 30 min on average on the site, often "liking" posts at a rate of 4.2 billion per day. Over 100 million photos are added to Instagram each day. As the videos and photos are intended to be the primary communication, the captions may, therefore, allow for analysis of expression, given that the visual provides the content. We thus examined how sex and status influenced the use of analytic and cognitive language in Instagram captions.

METHOD

Sample

The "Most Powerful Celebrities with Highest Social Ranking" was obtained from www.ranker.com[17] in August 2018. The original list included 100 celebrities, but some of those were members of groups (such as rock bands) and some did not have verifiable Instagram accounts, which we checked by locating the small blue "check mark" for verified accounts. Our sample thus included 41 celebrities (15 women, 26 men) who produced 942 Instagram posts (with captions) in a six-week period (August 14, 2018 to September 28, 2018). Each Instagram caption was taken verbatim from the post and placed into a word document, where we then applied the Linguistic Inquiry and Word Count 2015.[2]

Following Beach et al[10] we used a median split of number of followers to delineate high and low status, and then used a 2×2 (Sex×Status Median Split) between-subjects ANOVA with number of followers as the dependent measure to check our manipulation. Table 1 shows the mean number of followers according to our independent variables. This analysis was statistically significant for status, $F_{(1, 37)}=22.79$, $p<0.001$, $\eta_p^2=0.38$, as the 21 persons designated as high status had significantly more followers (M=53,066,667, SD=45,571,541) than the 20 persons designated as low status (M=2,873,800, SD=2,401,524). However, neither the main effect for sex, $F_{(1, 37)}<1$, $p>0.05$, nor the interaction, $F_{(1, 37)}<1$, $p>0.05$, were statistically significant. The number of Instagram captions that comprised the sample according to sex and status can be seen in Table 2.

Table 1. *Means and Standard Deviations for Number of Followers as a Function of Sex and Status*

	Status	
	High	Low
Women	63,277,778 (43,899,367)	2,369,333 (1,907,432)
Men	45,408,333 (47,170,743)	3,090,000 (1,907,432)

Note. Followers differ between high- and low-status celebrities.

Table 2. *Number of Instagram Captions According to Sex and Status*

	Status	
	High	Low
Women	329	166
Men	285	162

RESULTS

One concern in the examination of analytic expression and content was that these may have been a function of words given per sentence, as longer sentences might suggest a higher degree of analytical expression and cognitive thought. Words per sentence (WPS) were therefore used as a dependent measure in a 2×2 (Sex×Status Median Split) between-subjects ANOVA. There was a statistically significant main effect of status, $F(1, 938)=9.75$, $p<0.003$, $\eta_p^2=0.01$, but no statistically significant effect of sex, $F(1, 938)=3.60$, $p>0.05$. High status (HS) celebrities (M=8.84, SD=6.11) used significantly fewer WPS than did low-status (LS) celebrities (M=10.22, SD=6.87), although the effect was qualified by a statistically significant interaction, $F(1, 938)=6.44$, $p<0.012$, $\eta_p^2=0.007$. Post-hoc tests (simple effects) showed that men's WPS was relatively regardless of their status ($M_{LS}=9.25$, SD=5.51 *vs.* $M_{HS}=8.99$, SD=5.78), $t(938)<1$, $p>0.05$, but that high-status women used significantly far shorter captions (M=8.71, SD=6.39) than did lower-status women (M=11.18, SD=7.87), $t(938)=4.08$, $p<0.001$. Additionally, lower-status women had higher WPS than their low-status men counterparts, $t(938)=2.75$, $p<0.05$. Therefore, WPS was used as a covariate in the subsequent analyses.

Analytic Language

The means and standard deviations for the 2×2 (Sex×Status) between-subjects ANCOVA on analytic language can be seen in Table 3. There was a statistically significant main effect of status, $F(1, 937)=12.05$, MSE=965.35, $\eta_p^2=0.013$, and sex, $F(1, 937)=9.49$, $p<0.003$, $\eta_p^2=0.010$. Men (M=79.97, SD=29.80) used significantly more analytic language with prepositions and articles more than women did (M=66.33, SD=32.69). Additionally, low-status (LS) persons used significantly more analytic language than did high-status (HS) persons (MLS=75.27, SD=28.33; MHS=67.12, SD=32.84). The interaction was not statistically significant $F(1, 937)=2.44$, $p>0.05$. Observed power was 0.93 and 0.87 for status and sex, respectively.

Table 3. *Means and Standard Deviations for Percentage of Various Aspects of Language as a Function of Sex and Social Status*

	Sex			
	Men		Women	
	Status			
	Low	High	Low	High
Analytic Language	76.63 (28.34)	72.47 (30.55)	73.95 (28.33)	62.49 (34.07)
Cognitive Mechanisms Insight	0.97 (2.18)	1.70 (3.56)	1.50 (2.93)	1.48 (2.91)
Causation	0.29 (1.06)	0.93 (2.98)	0.85 (2.30)	1.18 (3.71)
Discrepancy	0.33 (1.25)	0.40 (1.41)	0.87 (2.26)	0.58 (2.02)
Tentativeness	1.20 (2.97)	0.91 (2.92)	1.31 (2.78)	1.23 (3.13)
Certainty	1.23 (2.71)	1.57 (3.93)	1.26 (2.55)	1.52 (3.27)
Differentiation	0.87 (2.06)	0.87 (2.31)	1.42 (2.87)	1.29 (3.59)

Note: Numbers represent percentage of language classified in the LIWC category.

Cognitive Mechanisms

The linguistic variables that comprise the category of cognitive thought (differentiation, discrepancy, causation, insight, tentativeness, and certainty) were entered as dependent variables in a 2×2 (Sex×Status Median Split) MANCOVA, holding WPS constant. The means and standard deviations from the individual variables in this category are displayed in Table 3.

The MANCOVA produced statistically significant main effects of sex of celebrity, $F(6, 932)=2.50$, $p<0.022$, Wilks' $\lambda=0.98$, $\eta_p^2=0.016$, and status, $F(6, 932)=2.33$, $p<0.032$, Wilks' $\lambda=0.99$, $\eta_p^2=0.015$. The interaction was not statistically significant, $F(6, 932)<1$, $p>0.05$. Observed power was 0.84 and 0.81 for the main effects of sex and status, respectively.

Follow-up ANCOVAs with a Bonferroni correction for six dependent variables were performed. Significantly more discrepancy (e.g., "should" or "maybe") was seen in women (M=0.68, SD=2.10) than in men (M=0.37, SD=1.36), $F(1, 937)=7.51$, $p<0.039$, $\eta_p^2=0.008$. Observed was power 0.78. Additionally, the language of high-status persons (M=1.07, SD=3.39) showed a trend toward more causality (e.g., "because") than that of low-status celebrities (M=0.57, SD=1.82), $F(1, 937)=6.78$, $p=0.052$, $\eta_p^2=0.007$, observed power=0.74. No other statistically significant differences were found located for sex or status.

The foregoing analyses highlight differences between expression and content, and suggest there should be an inverse relationship between analytic language (content) and cognitive mechanisms (transmission). Indeed, analytic language and the language of cognitive processes were statistically inversely related, $r(942)=-298$, $p<0.001$.

Openventio
PUBLISHERS

DISCUSSION

The results of this study revealed men and low-status persons, in comparison to women and high-status celebrities, used language that showed more "analytic" expression designed to demonstrate critical thought. Yet, high-status celebrities captioned their visuals with more causality. Women employed discrepancy "hedging" their language somewhat. The findings were not a function of words per sentence.

The findings with regard to causality and analytic language are on the surface paradoxical: the content of the captions of high-status celebrities were more likely to include facts and, perhaps, their own stories, as language with causation may reflect personal experiences.[14] Thus, the manner in which they wrote their captioning was different from that of low-status celebrities and men. The latter used analytical language, which typically includes articles and prepositions, termed "function" words.[15] As Jordan and colleagues[15] note, these linguistic mechanisms reflect a simplification and unpacking of more complicated ideas. Thus, low-status persons and men communicated analytically, perhaps omitting personal experiences. Indeed, this argument is supported by an analysis of 150 Instagram posts of three major women celebrities that showed that the most popular captions (measured by likes) were those that dealt with personal issues or one-sided (i.e., parasocial) relationships with fans.[18]

Pennebaker and colleagues[12,14] have shown that men use more prepositions and articles across several types of writing, and our data confirm that men provided more simplification of their ideas on the captions, as measured through analytical language. Low-status celebrities did the same, which may also provide a reason why their words-per-sentence were high. While low-status celebrities and men used more words, the content of captions from high-status celebrities may have been more complex because they have sufficient followers to retain attention, regardless of the complexity of their content.

As in previous research,[14] women's language included more discrepancies, a style that is generally more polite. However, discrepancies are somewhat context dependent, varying according to whether the writing is spontaneous or planned, with sex differences more likely in language that is unconstrained,[14] such as with Instagram captions. In this research, captioning did not really include much of this polite, hedging language compared to the standards present by Pennebaker et al[2] analysis of over 200 million words, generated across multiple contexts by about 80,000 speakers, demonstrated hedging at a rate 2- to 3-times the rate we found, which is not surprising considering people were captioning their own visual content. Thus, the small amount of differentiation was present but women were using it slight more than men, thereby qualifying their speech. It should be noted that the standard deviations in all our measured categories were relatively high. Yet, even with this type of free-form content, women still used more polite language than men.

Several fruitful avenues of research exist for future analyses of Instagram captioning. Most notably, Instagram captions of non-celebrities should be examined, particularly given our large proportion of captions by high-status women. Additionally, there are several other research questions connected to affect, given that similar work using 140-word Tweets has shown that low-status celebrities and men "keep it light" and positive in that social media platform.[10] Additional variables influencing language use include age[13] and personality.[1,4,7] Most importantly, the nature and valence of the visual—whether it is explained, amplified, complementary, or in contrast to—the text content will also be essential. Do men and women of varying status caption their visual presentation differently, and is their visual presentation different? Perhaps their pictures differentially mirror their content, helping us to understand further how people communicate what is on their minds.

CONCLUSION

Our findings support previous findings with regard to sex and language, and also shed light on how status may mitigate sex-linked language effects. Sex and status contribute to the language used in captioning Instagram posts, perhaps based on the likelihood that an audience will read complex content from high-status celebrities and women, with more concrete language seen in lower-status persons whose followers may not wade through complicated ideas. Regardless of content, women's captions were more polite.

CONSENT

There is no consent required for research with public documents, and research used public-domain material (public Instagram accounts), which does not require Institutional Review Board (IRB) approval.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

1. Kern M, Eichstaedt J, Seligman M, et al. From 'Sooo excited!!!' to 'So proud': Using language to study development. *Dev Psychol*. 2014; 50(1): 178-188. doi: 10.1037/a0035048

2. Pennebaker JW, Boyd, RL, Jordan K, Blackburn, K.The development and psychometric properties of LIWC2015. 2015. Austin, TX: University of Texas at Austin.

3. Mehl M, Pennebaker JW. The sounds of social life: A psychometric analysis of students' daily social environments and natural conversations. *J Pers Soc Psychol*. 2003; 84(4): 857-870. doi: 10.1037/0022-3514.84.4.857

4. Park G, Schwartz HA, Eichstaedt JC, et al. Automatic personality assessment through social media language. *J Pers Soc Psychol*. 2015; 108(6): 934-952. doi: 10.1037/pspp0000020

5. Golbeck J. Negativity and anti-social attention seeking among narcissists on Twitter: A linguistic analysis. First Monday. 2016; 21(3): Accessed August 2, 2016.

6. Pennebaker JW, King L. Linguistic styles: Language use as an individual difference. *J Pers Soc Psychol.* 1999; 77(6): 1296-1312. doi: 10.1037//0022-3514.77.6.1296

7. Schwartz HA, Eichstaedt JC, Kern ML, et al. Personality, gender, and age in the language of social media: The open vocabulary approach. *PLoS One.* 2013; 8(9): e73791. doi: 10.1371/journal.pone.0073791

8. Stever GS, Lawson K. Twitter as a way for celebrities to communicate with fans: Implications for the study of parasocial interaction. *North American Journal of Psychology.* 2013; 15(2): 339-354.

9. Brownlow S, Beach JC, Silver NC. How social status influences "affect language" in Tweets. *Psychol Cogn Sci Open J.* 2017; 3(4): 100-104. doi: 10.17140/PCSOJ-3-130

10. Beach JC, Brownlow S, Greene, M, Silver NC. The "I"s have it: Sex and social status differences on Twitter. *Soc Behav Res Pract Open J.* 2016; 1(1): 17-21. doi: 10.17140/SBRPOJ-1-104

11. Pennebaker J, Chung C, Frazee J, Lavergne G, Beaver D. When small words foretell academic success: The case of college admissions essays. *PLoS One.* 2014; 9(12): e115844. doi: 10.1371/journal.pone.0115844

12. Tausczik Y, Pennebaker J. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal Of Language And Social Psychology.* 2010; 29(1): 24-54. doi: 10.1177/0261927X09351676

13. Pennebaker J, Stone L. Words of wisdom: Language use over the life span. *J Pers Soc Psychol.* 2003; 85(2): 291-301. doi: 10.1037/0022-3514.85.2.291

14. Newman M, Groom C, Handelman L, Pennebaker J. Gender differences in language use: An analysis of 14,000 text samples. *Discourse Processes.* 2008; 45(3): 211-236. doi: 10.1080/01638530802073712

15. Jordan KN, Sterling J, Pennebaker JW, Boyd RL. Examining long-term political trends in politics and culture through language of political leaders and cultural institutions. *Proceedings Of the National Academy of Sciences.* 2019; 116(9): 3476-3481 doi: 10.1073/pnas.1811987116

16. Ahlgren M. 28+ Instagram statistics and facts for 2019. Web site. https://www.websitehostingrating.com/instagram-statistics/. Accessed May 15, 2019.

17. Most Powerful Celebrities with Highest Social Ranking. Web site. www.ranker.com. Accessed August 10, 2018.

18. Ward J. A content analysis of celebrity Instagram posts and parasocial interaction. *Elon Journal of Undergraduate Research in Communications.* 2016; 7(1): 44-51.