A statistical analysis of satirical Amazon.com product reviews

Stephen Skalicky, Scott Crossley

ABSTRACT

A corpus of 750 product reviews extracted from Amazon.com was analyzed for specific lexical, grammatical, and semantic features to identify differences between satirical and non-satirical Amazon.com product reviews through a statistical analysis. The corpus contained 375 reviews identified as satirical and 375 as non-satirical (750 total). Fourteen different linguistic indices were used to measure features related to lexical sophistication, grammatical functions, and the semantic properties of words. A one-way multivariate analysis of variance (MANOVA) found a significant difference between review types. The MANOVA was followed by a discriminant function analysis (DFA), which used seven variables to correctly classify 71.7 per cent of the reviews as satirical or non-satirical. Those seven variables suggest that, linguistically, satirical texts are more specific, less lexically sophisticated, and contain more words associated with negative emotions and certainty than non-satirical texts. These results demonstrate that satire shares some, but not all, of the previously identified semantic features of sarcasm (Campbell & Katz 2012), supporting Simpson’s (2003) claim that satire should be considered separately from other forms of irony. Ultimately, this study puts forth an argument that a statistical analysis of lexical, semantic, and grammatical properties of satirical texts can shed some descriptive light on this relatively understudied linguistic phenomenon, while also providing suggestions for future analysis.

KEYWORDS

satire; statistical analysis; online product reviews

REFERENCES

Carvalho, P., Sarmiento, L., Silva, M. & de Oliveira, E. (2009). ‘Clues for detecting irony in user-generated contents: Oh ... It’s “so easy”’. in TSA ‘09: 1st International CIKM Workshop on Topic-Sentiment Analysis for Mass Opinion (Hong Kong, 6 November 2009), New York: Association for Computing Machinery, pp. 53-56.

FULL TEXT: VIEW FULL TEXT HERE

FURTHER REFERENCES

corpus of 750 product reviews extracted from Amazon.com was analyzed for specific lexical, grammatical, and semantic features to identify differences between satirical and non-satirical Amazon.com product reviews through a statistical analysis. The corpus contained 375 reviews identified as satirical and 375 as non-satirical (750 total). Fourteen different linguistic indices were used to measure features related to lexical sophistication, grammatical functions, and the semantic properties of words. A one-way multivariate analysis of variance (MANOVA) found a significant difference between re...