

What makes an idea worth spreading? Language markers of popularity in TED talks by academics and other speakers

Kate MacKrill¹ | Connor Silvester¹ | James W. Pennebaker² | Keith J. Petrie¹ 

¹Department of Psychological Medicine, Faculty of Medical and Health Sciences, University of Auckland, Auckland, New Zealand

²Department of Psychology, University of Texas, Austin, Texas

Correspondence

Keith J. Petrie, Department of Psychological Medicine, Faculty of Medical and Health Sciences, University of Auckland, Private Bag 92019, Auckland, New Zealand.
Email: kj.petrie@auckland.ac.nz

Abstract

TED talks are a popular internet forum where new ideas and research are presented by a wide variety of speakers. In this study, we investigated how the language used in TED talks influenced popularity and viewer ratings. We also investigated the differences in linguistic style and ratings of talks given by academics and non-academics. The transcripts of 1866 talks were analyzed using the Linguistic Inquiry and Word Count program and eight language variables were correlated with number of views and viewer ratings. We found that talks with more analytic language received fewer views, while a greater use of the pronoun “I,” positive emotion and social words was associated with more views. Talks with these linguistic characteristics received more emotional viewer ratings such as inspiring or courageous. When comparing talks by academics and non-academics, there was no difference in the overall popularity but viewers rated talks by academics as more fascinating, informative, and persuasive while non-academics received higher emotional ratings. The implications for understanding social influence processes are discussed.

1 | INTRODUCTION

TED talks have become an internet success story since their online launch in 2006. By 2012, the combined talks on the Ted.com website had been watched over one billion times. Originating from a conference on Technology, Entertainment, and Design, TED talks have grown to cover a wide variety of topics, with over 2,600 talks now available online that collectively receive 1.5 million views a day (TED, 2012). TED talks have been given by world leaders, technology entrepreneurs, Nobel Prize winners, university academics, musicians and other speakers from a wide variety of backgrounds. The talks have a similar format that usually takes less than 18 minutes and focus on personal storytelling, often with humor and images designed to explain concepts and applications to a lay audience. The tone of the presentations is generally

optimistic, passionate and promotes a sense of shared social activism (Ludewig, 2017).

The tagline of TED is “ideas worth spreading” but the question of what makes some talks more popular than others has received surprisingly little attention, given the amount of information available for each talk. For example, the TED website includes basic statistics about each talk, including the date when the talk was uploaded to the website and the total number of views. The database also includes the videos from which you can get the length of the talk, the complete transcripts of the talks in English with some translations into other languages, as well as viewers' ratings. These rating allow viewers to choose from a number of descriptors (e.g., fascinating, funny, informative, inspiring). From the website and other sources, it is also possible to get the speakers' age, sex, education, field of study and current occupation.

1.1 | TED talks and viewer responses

There has been some previous research looking at the associations between various TED talk features and viewer responses. TED word use has been found to have a similar word vocabulary coverage to newspapers and novels (Coxhead & Walls, 2012). Özmen and Yucel (2019) found that talks rated persuasive were 35% longer than talks rated as ingenious and talk titles that include the word “brain” are more likely to be rated as fascinating. Studies looking specifically at TED talk topics find that science and technology talks are more liked on YouTube than talks on art and design (Sugimoto & Thelwall, 2012). On YouTube, talks about inequality are more likely to receive negative comments (Schwemmer & Jungkunz, 2019). There has also been research on the association between background characteristics of presenters and the number of talk views. For example, in general talks by male presenters are viewed more on YouTube but not on the TED website (Sugimoto et al., 2013). In one analysis, female presenters received more negative comments on YouTube than male speakers (Schwemmer & Jungkunz, 2019), while another showed that female presenters received more emotional comments in general and male presenters more neutral comments (Veletsianos et al., 2018).

The rise of the interest in TED talks has coincided with a push to make scientific findings more accessible to the public and an examination of what enhances this process (Dudo, 2012). The internet allows a more direct communication between the scientist and the public without the need for a journalist. This means findings can be transmitted quickly to a large social media and online platform audience, but only if the message is in a format that is appealing and readily understood by the public. Work has begun to look at the factors that make this scientific communication more effective, such as asserting the credibility of the speaker, lower levels of jargon and the use of a narrative format (Fontaine et al., 2019; Liang et al., 2014). Studies have examined the impact of an academic speaker on TED talk popularity. TED talks have provided a new forum for the work of scientists to gain exposure and have their work popularized (Romanelli et al., 2014).

The TED platform has provided a ready audience for new scientific ideas and concepts and has turned relatively unknown academics into internet celebrities. Whether a speaker is an academic or not appears to be a relevant factor to investigate as some of the most viewed TED talks have been delivered by academics. For example, the Swedish academic epidemiologist Hans Rosling became internationally known following a series of captivating TED talks and use of visualizations that explained

worldwide population and economic changes (Gallo, 2017). Similarly, social psychologist Amy Cuddy's 2012 talk on power posing has been viewed over 40 million times and has possibly had the unexpected consequence of intensifying questions concerning the reliability of her initial findings (Credé & Phillips, 2017; Cuddy et al., 2018; Simmons & Simonsohn, 2017).

However, Sugimoto et al. (2013) found that while academic presenters received more likes and comments on YouTube, they were not necessarily viewed more than non-academics. Moreover, they also found that the academics' standing, in terms of their citation rates, was not associated with number of views, but it should be noted that most academic TED presenters are senior faculty from American academic institutions. Academic presenters have been shown to receive fewer negative reactions than non-academic presenters, perhaps illustrating the public's trust in their authority (Sugimoto & Thelwall, 2012). Another study found that inspiring and courageous ratings are more often given by viewers to personal emotional stories, while academic talks on social issues often receive informative and persuasive ratings (Meza & Trofin, 2015).

1.2 | Psychological styles and talk response

Beyond the background characteristics of the speakers and the content of the talk itself, we were interested to investigate whether it is possible to identify psychological styles of the speakers through their language use in order to predict how frequently their talks are viewed. In other words, to examine whether certain speaking styles were more popular than others. Of particular interest are the ratings viewers assign to each talk. While some talks become “viral” due to receiving a large number of views by being shared rapidly and widely on the internet (Wuebben, 2016), there may also be something inherent in the speaker's language style that influences the ratings and views the talk receives. Due to the large amount of views some TED talks by academics have received and past research examining the effect of academic presenters, we were also interested in how the language styles of academics and non-academics differed.

In recent years, an increasing number of studies have linked social and psychological processes to linguistic styles (Tausczik & Pennebaker, 2010). Of particular relevance have been findings concerning words that reflect peoples' writing or speaking styles rather than the content of what they are communicating. Whereas content-related words tend to be nouns, regular verbs, adjectives, and most adverbs, style-related words are made up of a

relatively small number of short but commonly used “function” words (Miller, 1995). Examples of function words include pronouns (e.g., I, them, it), prepositions (to, for, of), auxiliary verbs (was, have, am), articles (a, an, the), conjunctions (and, but, since), negations (no, not, never), and a small group of non-referential adverbs (so, really, very).

Interestingly, different groups of function words have been found to be related to overarching thinking styles, honesty and self-focus, status and power, and emotional tone. In an analysis of over 25,000 college admissions essays, a factor analysis of function word categories yielded a single coherent language factor with positive loadings of articles and prepositions and negative loadings of pronouns, auxiliary verbs, and the other function word dimensions. This bi-dimensional factor has been labeled as an analytic or narrative index of word use (Boyd & Pennebaker, 2015; Pennebaker et al., 2014). In addition, it is also possible to identify emotion in language use by looking at the proportion of positive and negative emotion words (Chung & Pennebaker, 2014).

Writing samples high in analytic thinking tend to be formal, logical, and hierarchical; those low in analytic thinking (or high in narrative, also referred to as authentic) tend to involve personal stories, more action of characters and events, and a sense of immediacy. Greater authenticity is characterized by greater first-person pronoun use (I, me) and lower negative emotion words (Newman et al., 2003). Interestingly, high analytic scores in admissions essays correlated with 4-year grade point averages and higher College Board scores (Pennebaker et al., 2014). Similarly, highly analytic written personal introductions have been shown to predict final course performance (Robinson et al., 2013).

Other dimensions of language-based speaking styles can reveal additional aspects of speakers' psychological states. For example, people use pronouns and other function words differently when they are honest than when they are duplicitous (Hancock et al., 2008; Newman et al., 2003). People high in clout or a sense of social hierarchy—either naturally or experimentally-induced—tend to use words like “we” and “you” at high rates compared to people lower in status, who have elevated use of “I” words (Kacewicz et al., 2013). People who are happier tend to use more positive emotion and “we” words and fewer negative emotion and “I” words than those who have greater levels of neuroticism (Yarkoni, 2010) or depression (Rude et al., 2004).

Ironically, almost all of the research on language use has focused on how words reflect the psychological state of the speaker. With few exceptions, very little research has explored how the language from a broad array of speakers may influence perceptions of the listeners or

audience (Berry et al., 1997; Khazaei et al., 2017; Larrimore et al., 2011). We applied this bi-dimensional theory of linguistics to TED talks and investigated whether it was possible that the subtle word choices of speakers influence the talk popularity and viewer ratings that they receive.

1.3 | The current study

The current study investigates how language style measured by a standard text analysis program Linguistic Inquiry and Word Count (LIWC; Pennebaker et al., 2015), which categorizes each word into psychologically meaningful categories to examine how these impact on people's perceptions of a speaker's TED talk. LIWC is a commonly used text analysis program that has been used to examine a wide range of questions including the relationship between language markers and narrative structures (Boyd et al., 2020) as well as the relationship between grammatical and psychological word categories with personality (Holtzman et al., 2019), psychological coping (Pennebaker et al., 2003) and social relationships (Chung & Pennebaker, 2014).

The advantage of using the TED paradigm is that we have access to over 1800 speakers who have been viewed and rated by millions of people. It is unusual to have a database that allows the same stimulus, in terms of a talk, to be rated by so many individuals. This allows a more fine-grained analysis of the relationship between language style and outcomes that is possible with most database approaches. Further this approach can help explain what makes TED talks popular and what implications this may have for making communication more appealing and compelling to audiences such as a more narrative style, which may have important lessons for communication in science, education and politics.

Based on prior research we predicted that certain types of language used would be associated with talk popularity. The literature shows that language can be differentiated based on the bi-dimensional factor of analytic/authentic. TED talks are a highly narrative “story-telling” style of presentation, designed to convey emotion and spoken with authenticity (Ludewig, 2017; Romanelli et al., 2014). As such, we hypothesized that:

Hypothesis H1. *TED speakers who have a more analytic linguistic style would have less popular talks.*

Hypothesis H2. *TED speakers who have a more authentic linguistic style and use more positive words would have more popular talks.*

Hypothesis H3. *Greater use of the word “I” would be associated with greater talk popularity.*

Due to the increase in science communication by experts, our final hypotheses concerned the differences between talks by academics and non-academics. Past research has shown that academic presenters receive fewer negative reactions than non-academic speakers (Sugimoto & Thelwall, 2012). However, linguistic analyses have shown that greater academic ability is associated with a more analytic speaking style (Pennebaker et al., 2014), which does not fit with the narrative style of the TED talk medium and therefore may result in fewer views and positive ratings. Academia also has a strong hierarchical organization (Martin, 1998). As such we predicted that:

Hypothesis H4. *TED talks by academics would have a greater use of analytic words as well as clout.*

Hypothesis H5. *TED talks by non-academics would use more authentic words, characterized by a higher use of “I” and more positive emotion words than academics.*

Hypothesis H6. *Due to academic talks having a greater analytic linguistic style, they would be rated more negatively by viewers than non-academic talks and on average would attract less views overall.*

2 | METHOD

All talks uploaded to the TED website (<https://www.ted.com/>) from June 2006 to February 2017 and fitting the study inclusion criteria were selected for analysis. Talks were required to have a minimum of 200 words and those featuring songs, dance routines and interviews were excluded, as these speaking styles do not reflect the typical TED talk presentations. For speakers who had given more than one talk, only their most viewed was included so as to not have duplicate speaking styles in the data analysis. This resulted in a sample of 1866 talks. Data collected from each talk included the title, speaker, talk length, date of upload to the TED website, total views the talk had received, viewer ratings, and the talk transcript. Thirteen talks were delivered in a language other than English, therefore a translated transcript was analyzed. In addition, each speaker was categorized as an academic (the criteria being that they had to be affiliated with a university and with a higher education degree) or non-academic at the time of their TED talk. Overall, 27.4% of the talks were delivered by academics. The start

point of 2006 was chosen as this was when the TED talks online platform was first created. Unfortunately, around 2017 TED removed the viewer rating feature so we were unable to collect more talk data beyond this time point.

The popularity of a talk was determined by the total number of views it had received as of March 1, 2017. The study also examined the viewer ratings of each talk as a secondary form of talk popularity. On the TED website viewers could rate each talk by clicking on a maximum of three of the following descriptors: beautiful, confusing, courageous, fascinating, funny, informative, ingenious, inspiring, jaw-dropping, longwinded, obnoxious, okay, persuasive and unconvincing. TED calculates the number of times a talk has been rated as each of these descriptors and converts the score to a percentage of the talk's total ratings.

2.1 | Data preparation and analytic strategy

The transcript of each talk was cleaned, removing any transcription information (e.g., number of lines, words not understood by translator), audience interaction (e.g., laughter, applause) and videos. Each transcript was analyzed using Linguistic Inquiry and Word Count (LIWC; Pennebaker et al., 2015). LIWC is a computerized text analysis program that calculates the percentage of words in a text that reflect different parts of speech, thinking styles, emotions and social concerns (Pennebaker et al., 2001). LIWC has been used in many previous studies to analyze different types of text such as the lyrics of Beatles songs (Petrie et al., 2008), poetry (Stirman & Pennebaker, 2001), patients' perceptions of their illness (Jordan et al., 2019), and interviews by presidential candidates (Pennebaker et al., 2005).

Although LIWC provides analyses for approximately 80 dimensions of language and punctuation, we focused on only the eight dimensions relevant to our hypotheses—all of which have been the subject of multiple previous studies: analytical thinking, clout, authenticity, I, we, positive emotion, negative emotion, and social processes (see Table 1 for descriptions and examples).

We conducted some initial analyses to determine whether there were any variables that needed to be controlled for. Kolmogorov–Smirnov tests for normality show that the total number of views and rating percentages were not normally distributed so nonparametric tests were used for all analyses. There was a significant association between the date a talk was uploaded to the TED website and total views, $r_s(1864) = .16$, $p < .001$. Controlling for date of upload did not influence the

strength or significance of the relationship between views, viewer ratings and the LIWC variables. Given this and the fact that the correlation was weak, talk upload date was not controlled for in further analyses.

We also examined whether there were any differences in the outcome variables for normal TED talks compared to TEDx talks. TEDx events are local gatherings where TED-like talks are shared with the community and may be less curated than the usual TED format. There were 345 TEDx talks in the dataset. No differences between TED and TEDx talks were found for total views. There were differences at the $p \leq .001$ level for the viewer ratings of informative and persuasive, and linguistic variables of analytic and clout. TEDx talks were significantly more informative (median = 18.0, interquartile range [IQR] = 16.0), persuasive (median = 9.0, IQR = 10.0) and were higher in clout (median = 78.62, IQR = 14.58) than regular TED talks (informative median = 15.0, IQR = 15.0; persuasive median = 7.0, IQR = 9.0; clout median = 76.92, IQR = 16.33). TEDx talks were lower in analytic style (median = 48.80, IQR = 23.28) than TED talks (median = 55.52, IQR = 26.26). Controlling for TEDx talks did not affect the strength or significance of the relationships between the variables and was therefore not controlled for in further analyses.

To investigate the relationship between a speaker's use of language and a talk's popularity and viewer ratings, Spearman correlations were conducted using the eight LIWC variables, total views and rating percentages. Consistent with de Winter et al. (2016), we relied on a Spearman correlation because of the heavy tailed distribution of the TED talk viewership data. The popularity, viewer ratings and language of academics versus non-academics were compared using a series of Mann–Whitney U tests

TABLE 1 Linguistic enquiry and word count variables

LIWC variable	Definition or examples
Analytical thinking	Language that is formal, logical
Clout	Language that is authoritative, confident, shows leadership
Authentic	Language that is personal, honest, vulnerable
I	I, me, mine
We	We, us, our
Positive emotion	Love, nice, sweet
Negative emotion	Hurt, ugly, nasty
Social processes	Mate, talk, they

with total views, rating percentages and LIWC variables as the outcome variables. Medians and interquartile range are reported. Due to the large number of statistical comparisons, only effects at $p \leq .001$ were considered significant in order to reduce the risk of a Type 1 error.

3 | RESULTS

3.1 | Talk popularity and language

The five most popular and five least popular TED talks, as measured by their number of views, are shown in Table 2. The median number of views a talk had received was 1,139,255 (IQR = 920,319). We examined which aspects of a speaker's language were associated with a talk's popularity (H1–H3) and these data are presented in Table 3. As predicted, speakers who used more analytic language in their talks received fewer views (H1), whereas those that used more authentic language were more popular (H2). Greater use of the pronoun “I” was associated with a greater popularity score (H3). We also found a greater use

TABLE 2 The five most popular and five least popular TED talks (as of March 1, 2017)

Talk title	Speaker	Total views
<i>Most popular</i>		
Do schools kill creativity	Ken Robinson ^a	43,773,514
Your body language shapes who you are	Amy Cuddy ^a	39,417,669
How great leaders inspire action	Simon Sinek	30,879,283
The power of vulnerability	Brené Brown ^a	28,604,837
10 things you didn't know about orgasm	Mary Roach	28,604,837
<i>Least popular</i>		
Plant fuels that could power a jet	Bilal Bomani	126,391
Smelfies, and other experiments in synthetic biology	Ani Liu	130,787
The racial politics of time	Brittney Cooper	165,451
The balancing act of compassion	Jackie Tabick	165,539
A new way to stop identity theft	David Birch	168,075

^aIndicates that the speaker is an academic by our criteria.

of positive emotion words was associated with popularity. Speakers who used more social words in their talks were also viewed more. Word count, clout, “we” and negative emotions were not associated with a talk’s popularity.

3.2 | Viewer ratings and language

The relationship between viewer ratings and a speaker’s language was examined using Spearman correlations (Table 3). The talks given positive ratings had stronger correlations with language than the negatively rated talks. Talks given the more educational ratings of fascinating, informative, ingenious and persuasive tended to be less linguistically authentic, less likely to use “I” and positive emotions, and more likely to use “we.” Whereas the emotion-based ratings of beautiful, courageous, funny and inspiring were associated with talks that were more authentic, characterized by a greater use of “I,” social and positive emotion words.

3.3 | Comparing talks by academics and non-academics

A comparison of the language used by academics and non-academics was made to investigate H4 and H5

(Figure 1). Academics used almost 13% more words in their talks (median = 2,321, IQR = 1,066.75) than non-academics (median = 1953.50, IQR = 1,344.50), $U = 282,047$, $p < .001$. As predicted (H5), non-academics were more authentic in their speaking style, and used “I” and positive emotion words more often. Contrary to our prediction, there were no differences between academics and non-academics in the analytic speaking style (H6) or in the use of social or negative emotion words but academics were higher in clout and also used “we” more often.

The median number of views for talks by academics was 1,166,564 (IQR = 993,617.75) while non-academics received 1,117,894 views (IQR = 899,735.25). To investigate Hypothesis H6, the popularity of talks by academics and non-academics was compared using a Mann–Whitney U test. There was no significant difference in popularity for talks by academics and non-academics, $U = 362,769$, $p = .120$. This did not support our hypothesis and suggests that talks by academics and non-academics receive a similar number of views.

How viewers rated talks by academics and to non-academics were also compared and Figure 2 summarizes the differences between these groups. Viewers rated talks by academics as being more fascinating,

TABLE 3 Correlations between the talk popularity score, viewer ratings, and LIWC variables for TED talks

	LIWC variable								
	Word count	Analytic	Clout	Authentic	I	We	Positive emotions	Negative emotions	Social processes
Popularity score	−0.01	−0.15	−0.00	0.11	0.13	−0.09	0.12	0.07	0.10
<i>Positive ratings</i>									
Beautiful	−0.17	−0.04	−0.12	0.24	0.41	−0.17	0.11	−0.04	0.11
Courageous	−0.02	−0.01	−0.04	0.12	0.35	−0.05	0.10	0.45	0.29
Fascinating	0.15	0.06	−0.04	−0.06	−0.28	0.05	−0.25	−0.39	0.33
Funny	−0.01	−0.33	−0.05	0.09	0.24	−0.21	0.20	−0.03	0.11
Informative	0.14	0.26	0.15	−0.28	−0.56	0.19	−0.24	0.01	−0.13
Ingenious	−0.06	−0.04	0.01	−0.13	−0.19	0.10	−0.13	−0.36	−0.26
Inspiring	0.00	−0.06	−0.06	0.17	0.36	0.02	0.23	0.14	0.29
Jaw-dropping	0.12	−0.02	−0.09	0.02	0.06	0.03	−0.31	−0.19	−0.25
Persuasive	0.21	0.11	0.28	−0.21	−0.26	0.21	0.08	0.33	0.19
<i>Negative ratings</i>									
Confusing	0.07	−0.01	−0.02	0.03	−0.10	−0.01	−0.02	−0.02	−0.06
Longwinded	0.32	0.02	−0.05	−0.01	−0.09	−0.01	0.03	−0.05	−0.09
Obnoxious	−0.01	−0.02	−0.03	0.03	0.06	−0.05	0.10	0.02	0.01
Okay	−0.24	0.05	0.00	0.01	−0.11	−0.04	0.06	−0.11	−0.08
Unconvincing	−0.07	0.06	0.03	0.03	−0.11	0.04	0.08	0.03	−0.01

Note: Bold indicates significance at $p < .001$.

informative and persuasive than talks by non-academics. Conversely, non-academic talks were considered to be more beautiful, courageous, and inspiring. There were no differences at the $p \leq .001$ level for any of the negative ratings or funny, ingenious and jaw-dropping.

4 | DISCUSSION

TED speakers who had a more authentic rather than analytic linguistic style, used the word “I” and had a greater use of social and positive emotion words had a greater number of views. However, there appears to be two

FIGURE 1 Comparison of language use in TED talks by academics and non-academics. Medians are reported. Error bars indicate IQR. *** Indicates groups significantly different at $p < .001$

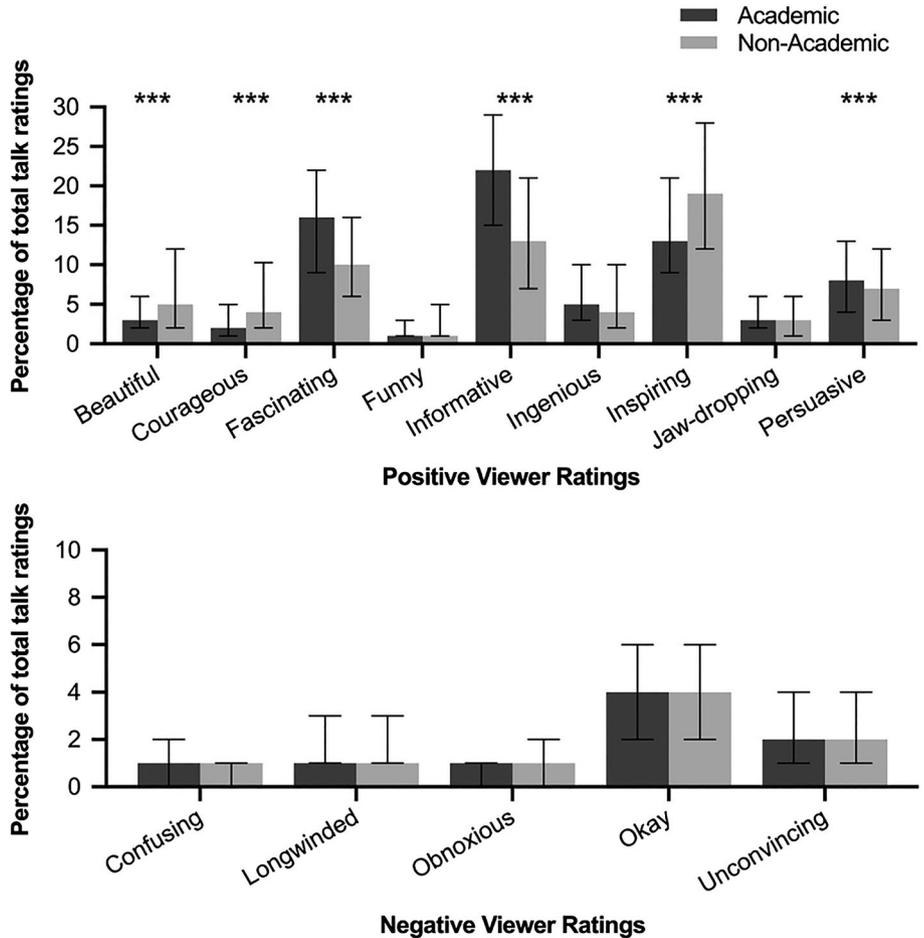
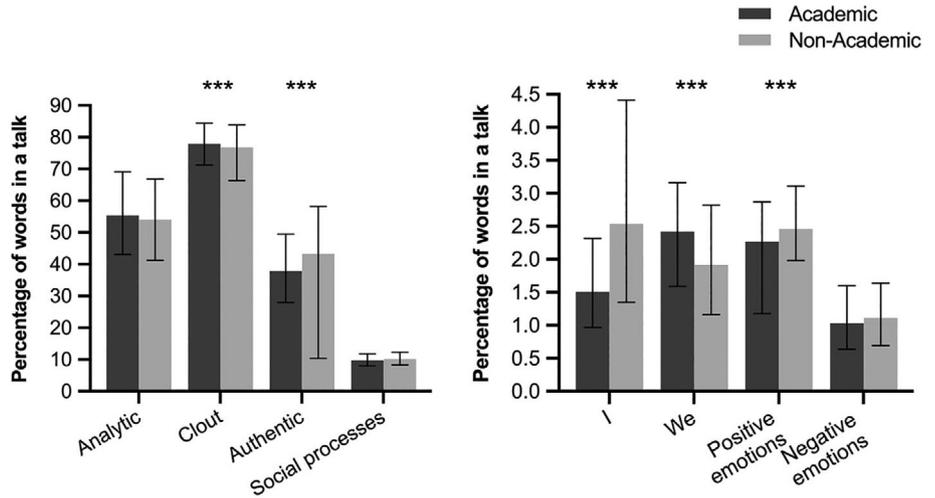


FIGURE 2 Viewer ratings and language use in TED talks by academics and non-academics. Medians are reported. Error bars indicate IQR. *** Indicates groups significantly different at $p < .001$

distinct linguistic profiles for talks given positive emotion-based ratings (beautiful, courageous, funny, inspiring) compared to educational ratings (fascinating, informative, ingenious, persuasive). Consistent with the pattern seen regarding the number of views, talks received emotional ratings when the speaker had a more authentic speaking style, used “I” more frequently, more social and positive words and fewer words associated with clout. The opposite linguistic profile was seen for talks given educational ratings, which were less authentic, less likely to use “I” and positive emotion words, and more likely to use “we.” No clear linguistic style was seen for the negative talk ratings or talks rated as jaw-dropping.

Comparing talks by academics and non-academics, viewers rated talks by academics as being significantly more fascinating, informative and persuasive, which appear to be the more educational descriptors. Non-academic talks showed the opposite pattern receiving the emotional ratings of beautiful, courageous, funny, and inspiring. The linguistic styles of academics versus non-academics were consistent with their viewer ratings, with non-academics being more authentic, lower in clout, and more likely to use “I” and positive emotions than academics. Interestingly, academics tended to use language associated with a greater sense of social status (i.e., greater clout, “we” and lower “I”) compared to non-academics, which may reflect the occurrence and importance placed on hierarchy in academia (Martin, 1998). While we found the speaking style characteristic of non-academics was associated with a more popular talk overall, talks by academics and non-academics received the same amount of views, indicating that they are equally popular. This suggests that talks by academics have other beneficial features that are not captured by linguistic analysis, which influence their popularity. For example, the general public may view academics as more reliable and have greater trust in the information they deliver (Tierney, 2006).

The current study moves beyond previous work on the relationship between popularity of the TED talk and demographic factors such as the speaker gender (Tsou et al., 2014) or age and also the prestige of university in the case of academics (Sugimoto et al., 2013), to examine more closely the language markers that are related to popularity and viewers' ratings. At a basic linguistic level, previous research has shown that viewer engagement with TED talks is increased through the use of personal pronouns such as “you” or “we” (Scotto di Carlo, 2014a), while the speaker's credibility is enhanced through “we” and “us” (Scotto di Carlo, 2014b). Another study that examined pronoun use in TED talks found that male speakers used “you” more often while women were more

likely to use “I” and “me” (Tsou et al., 2015). This research also examined academic word use and found a similar result to the current study in that academics were less likely to use “I” than non-academics. The authors proposed that this was due to academics largely being invited to give TED talks about research findings, whereas non-academics might be invited on the basis of celebrity, so likely to be talking about themselves or personal stories. Research has also shown that academics' language differs depending on their audience, with academics giving TED talks using “we” more frequently than academics delivering a university lecture. This may also reflect the fact much of the scientific enterprise is collective and involve groups of researchers (Compagnone, 2015; Nowotny et al., 2001).

However, the data in the current study extends work on linguistic analysis by indicating that the popularity of TED talks is driven both by the analysis and understanding that the talks provide as well as the emotional and authentic personal language used in the talk. The language differences between academics and non-academics are consistent with non-academics providing more personal stories, which viewers see as inspirational and courageous, while academics generally provide an analysis that describes complex phenomena for the lay audience (Tsou et al., 2015), which is viewed as being informative and fascinating. The fact that we found no difference in popularity between academic and non-academic TED talks is consistent with previous work by Sugimoto and Thelwall (2012) showing that the TED format did not disadvantage academics compared to non-academics and ran contrary to the popular belief that academics were poor communicators (Hoffman, 2016).

4.1 | Limitations

The current project is a real-world correlational study, which comes with both strengths and weaknesses. This study had a large sample size of talks in a similar format, an exceptionally large group of individuals who were rating each talk and also objective markers of language. Taken together, there is strong support that the patterns of effects found in this study are likely to be trustworthy, at least for TED talks.

However, the central limitation in this study is in assessing the causal mechanisms among language style, language content, and other features of human personality. Are audience members swayed by the ways speakers use of words or are they simply influenced by the speakers' stories, speaking styles (which includes tone of voice, nonverbal cues), attractiveness, or other dimensions? For example, we know that people applying for

college who use analytical language in their college admission essays (based on high rates of articles and prepositions and low rates of pronouns and auxiliary verbs) tend to have higher College Board scores and make better grades in their courses than people with lower analytic language (Pennebaker et al., 2014; Robinson et al., 2013). This language pattern is associated with more formal, logical, and hierarchical thinking. In the current study, analytic language was significantly correlated with talk ratings of “informativeness” but if the highly analytic speakers were given scripts with the same content so that their talks were low in analytic language, would audiences still have viewed them as informative?

Similarly, clout reflects where people stand in the social hierarchy. In one study, for example, people were randomly assigned a leadership position in small groups (Kacewicz et al., 2013). Those who were made leaders naturally used the language of higher clout (more use of “we” and “you” and less of “I”). In the current study, ratings of persuasiveness were significantly correlated with clout. Were people convinced of speakers’ clout based on their language or on their general bearing? An online survey found that participants gave similar ratings of charisma, credibility and intelligence for TED talks regardless of whether they saw and heard the speaker or viewed them without sound (Van Edwards & Vaughn, 2015).

It should also be acknowledged that the talks themselves are curated to fit into the TED format. So the style of speaking, word use and length can be modified to conform to this presentation structure. Furthermore, at any given time TED can change the information available for each talk. For example, since 2017 the viewer ratings have been removed and have been replaced with a “recommend” feature. This may change how viewers interact with TED talks and therefore the popularity metrics.

4.2 | Future research

The potential of applying computerized text analysis to the popularity of TED talks shows the potential that is now available in existing datasets to address questions of interest. The potential to use such linguistic analysis techniques to mine social media platforms and search engines is enormous (Boyd & Pennebaker, 2017). The LIWC has become an important a widely used tool for researchers as its psychometric properties have been validated over time and it is available for use in several languages (Boyd et al., 2020).

There is clearly more work to be done to look at how aspects of talks are related to ratings. A recent study found little evidence that people’s evaluations of TED talks are predicted by superficial characteristics of the

speakers (Gheorghiu et al., 2020). While the current results show that linguistic style is associated with talk popularity and viewer ratings, it is most likely that language interacts with other factors such as talk content and personality. The question still remains about what factor is most influential on viewers’ perceptions.

This study has demonstrated that talks delivered by academics compared to non-academics differ in their linguistic style. Future research may also investigate how speakers’ language styles are influenced by their nationalities and English varieties. While we have examined the linguistic aspects that make certain TED talks more popular, the characteristics of TED viewers may also influence why TED talks are perceived in certain ways.

5 | CONCLUSION

In this study we found that TED talks with a more authentic speaking style and a greater use of “I,” social and positive emotion words were viewed more. Talks with these linguistic characteristics were given more emotional descriptive ratings by viewers, while the opposite language style received educational ratings. While talks by academics and non-academics had different speaking styles as well as viewer ratings, they were equally popular. Future research must begin to carefully separate the various features of speaking style in order to examine the relative impact of the speakers’ personality and language use in influencing the judgment of listeners. It is most likely that there are several recursive relationships among personality, speaking style, and language content. Regardless of the outcomes of these studies, this research and similar studies that examine the popularity of TED talks will help to train people to be better speakers and communicators.

ORCID

Keith J. Petrie  <https://orcid.org/0000-0002-6337-2480>

REFERENCES

- Berry, D. S., Pennebaker, J. W., Mueller, J. S., & Hiller, W. S. (1997). Linguistic bases of social perception. *Personality and Social Psychology Bulletin*, 23, 526–537. <https://doi.org/10.1177/0146167297235008>
- Boyd, R. L., Blackburn, K., & Pennebaker, J. W. (2020). The narrative arc: Revealing core narrative structures through text analysis. *Science Advances*, 7, 2020. <https://doi.org/10.1126/sciadv.aba2196>
- Boyd, R. L., & Pennebaker, J. W. (2015). Did Shakespeare write double falsehood? Identifying individuals by creating psychological signatures with text analysis. *Psychological Science*, 26, 570–582. <https://doi.org/10.1177/0956797614566658>
- Boyd, R. L., & Pennebaker, J. W. (2017). Language-based personality: A new approach to personality in a digital world. *Current*

- Opinion in Behavioral Sciences*, 18, 63–68. <https://doi.org/10.1016/j.cobeha.2017.07.017>
- Chung, C. K., & Pennebaker, J. W. (2014). Using computerized text analysis to track social processes. In T. Holtgraves (Ed.), *The Oxford Handbook of Language and Social Psychology* (pp. 219–230). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199838639.013.037>
- Coxhead, A. J., & Walls, R. (2012). Ted talks, vocabulary, and listening for EAP. *TESOLANZ Journal*, 20, 55–67.
- Compagnone, A. (2015). The reconceptualization of academic discourse as a professional practice in the digital age: A critical genre analysis of TED talks. *HERMES—Journal of Language and Communication in Business*, 27, 49–69. <https://doi.org/10.7146/hjlc.v27i54.22947>
- Credé, M., & Phillips, L. A. (2017). Revisiting the power pose effect: How robust are the results reported by Carney, Cuddy, and Yap (2010) to data analytic decisions? *Social Psychological and Personality Science*, 8, 493–499. <https://doi.org/10.1177/1948550617714584>
- Cuddy, A. J., Schultz, S. J., & Fosse, N. E. (2018). P-curving a more comprehensive body of research on postural feedback reveals clear evidential value for power-posing effects: Reply to Simmons and Simonsohn (2017). *Psychological Science*, 29, 656–666. <https://doi.org/10.1177/0956797617746749>
- de Winter, J. C., Gosling, S. D., & Potter, J. (2016). Comparing the Pearson and Spearman correlation coefficients across distributions and sample sizes: A tutorial using simulations and empirical data. *Psychological Methods*, 21(3), 273–290. <https://doi.org/10.1037/met0000079>
- Dudo, A. (2012). Towards a model of scientists' public communication activity: The case of biomedical researchers. *Science Communication*, 35, 476–501. <https://doi.org/10.1177/1075547012460845>
- Fontaine, G., Maheu-Cadotte, M. A., Lavallée, A., Mailhot, T., Rouleau, G., Bouix-Picasso, J., & Bourbonnais, A. (2019). Communicating science in the digital and social media ecosystem: Scoping review and typology of strategies used by health scientists. *JMIR Public Health Surveillance*, 5(3), e14447. <https://doi.org/10.2196/14447>
- Gallo, C. (2017). TED talks star Hans Rosling taught us to 'edutain' our audiences. *Forbes*. <https://www.forbes.com/sites/carminegallo/2017/02/11/ted-talks-star-hans-rosling-taught-us-to-edutain-our-audiences/#1cd2b8c4568c>
- Gheorghiu, A. I., Callan, M. J., & Skylark, W. J. (2020). A thin slice of science communication: Are people's evaluations of TED talks predicted by superficial impressions of the speakers? *Social Psychological and Personality Science*, 11, 117–125. <https://doi.org/10.1177/1948550618810896>
- Hancock, J. T., Curry, L. E., Goorha, S., & Woodworth, M. (2008). On lying and being lied to: A linguistic analysis of deception in computer-mediated communication. *Discourse Processes*, 45, 1–23. <https://doi.org/10.1080/01638530701793181>
- Hoffman, A. J. (2016). *Why academics are losing relevance in society—and how to stop it*. The Conversation. <https://theconversation.com/why-academics-are-losing-relevance-in-society-and-how-to-stop-it-64579>
- Holtzman, N. S., Tackman, A. M., Carey, A. L., Brucks, M., Küfner, A. C. P., große Deters, F., Back, M. D., Donnellan, M. B., Pennebaker, J. W., Sherman, R. A., & Mehl, M. R. (2019). Linguistic markers of grandiose narcissism: A LIWC analysis of 15 samples. *Journal of Language and Social Psychology*, 38, 773–786. <https://doi.org/10.1177/0261927X19871084>
- Jordan, K. N., Pennebaker, J. W., Petrie, K. J., & Dalbeth, N. (2019). Googling gout: Understanding the perceptions and experience of gout through linguistic analysis of online search activities. *Arthritis Care & Research*, 71, 419–426. <https://doi.org/10.1002/acr.23598>
- Kacewicz, E., Pennebaker, J. W., Davis, M., Jeon, M., & Graesser, A. C. (2013). Pronoun use reflects standings in social hierarchies. *Journal of Language and Social Psychology*, 33, 125–143. <https://doi.org/10.1177/0261927X13502654>
- Khazaei, T., Lu, X., & Mercer, R. (2017). Writing to persuade: Analysis and detection of persuasive discourse. *iConference 2017 Proceedings*. <https://doi.org/10.9776/17077>
- Larrimore, L., Jiang, L., Larrimore, J., Markowitz, D., & Gorski, S. (2011). Peer to peer lending: The relationship between language features, trustworthiness, and persuasion success. *Journal of Applied Communication Research*, 39, 19–37. <https://doi.org/10.1080/00909882.2010.536844>
- Liang, X., Su, L. Y., Yeo, S. K., Scheufele, D. A., Brossard, D., Xenos, M., Nealey, P., & Corley, E. A. (2014). Building buzz: (scientists) communicating science in new media environments. *Journalism & Mass Communication Quarterly*, 91(4), 772–791. <https://doi.org/10.1177/1077699014550092>
- Ludewig, J. (2017). TED Talks as an emergent genre. *CLCWeb: Comparative Literature and Culture*, 19, 1–9. <https://doi.org/10.7771/1481-4374.2946>
- Martin, B. (1998). *Tied knowledge: Power in higher education*. University of Wollongong.
- Meza, R., & Trofin, C. (2015). Between science popularization and motivational infotainment: Visual production, discursive patterns and viewer perception of TED talks videos. *Studia Universitatis Babeş-Bolyai—Ephemerides*, 2, 41–60.
- Miller, G. A. (1995). *The science of words*. Scientific American Library.
- Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. (2003). Lying words: Predicting deception from linguistic style. *Personality and Social Psychology Bulletin*, 29, 665–675. <https://doi.org/10.1177/0146167203029005010>
- Nowotny, H., Scott, P., & Gibbons, M. (2001). *Re-thinking science: Knowledge and the public in an age of uncertainty*. Blackwell.
- Özmen, M. U., & Yucel, E. (2019). Handling of online information by users: Evidence from TED talks. *Behaviour & Information Technology*, 38, 1309–1323. <https://doi.org/10.1080/0144929X.2019.1584244>
- Pennebaker, J. W., Booth, R. J., Boyd, R. L., & Francis, M. E. (2015). *Linguistic inquiry and word count: LIWC 2015*. Pennebaker Conglomerates.
- Pennebaker, J. W., Chung, C. K., Frazee, J., Lavergne, G. M., & Beaver, D. I. (2014). When small words foretell academic success: The case of college admissions essays. *PLoS One*, 9(12), e115844. <https://doi.org/10.1371/journal.pone.0115844>
- Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). *Linguistic inquiry and word count: LIWC 2001 manual*. Lawrence Erlbaum Associates.
- Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology*, 54, 547–577.
- Pennebaker, J. W., Slatcher, R. B., & Chung, C. K. (2005). Linguistic markers of psychological state through media interviews: John

- Kerry and John Edwards in 2004, Al Gore in 2000. *Analyses of Social Issues and Public Policy*, 5, 197–204. <https://doi.org/10.1111/j.1530-2415.2005.00065.x>
- Petrie, K. J., Pennebaker, J. W., & Sivertsen, B. (2008). The things we said today: A linguistic analysis of the Beatles. *Psychology of Aesthetics, Creativity, and the Arts*, 2, 197–202. <https://doi.org/10.1037/a0013117>
- Robinson, R. L., Navea, R., & Ickes, W. (2013). Predicting final course performance from students' written self-introductions: A LIWC analysis. *Journal of Language and Social Psychology*, 32, 469–479. <https://doi.org/10.1177/0261927X13476869>
- Romanelli, F., Cain, J., & McNamara, P. J. (2014). Should TED talks be teaching us something? *American Journal of Pharmaceutical Education*, 78(6), 113. <https://doi.org/10.5688/ajpe786113>
- Rude, S., Gortner, E. M., & Pennebaker, J. W. (2004). Language use of depressed and depression-vulnerable college students. *Cognition & Emotion*, 18, 1121–1133. <https://doi.org/10.1080/02699930441000030>
- Schwemmer, C., & Jungkunz, S. (2019). Whose ideas are worth spreading? The representation of women and ethnic groups in TED talks. *Political Research Exchange*, 1(1), 1–23. <https://doi.org/10.1080/2474736X.2019.1646102>
- Scotto di Carlo, G. (2014a). The role of proximity in online popularizations: The case of TED talks. *Discourse Studies*, 16, 591–606. <https://doi.org/10.1177/1461445614538565>
- Scotto di Carlo, G. (2014b). Ethos in Ted talks: The role of credibility in popularised texts. *FACTA UNIVERSITATIS—Linguistics and Literature*, 12, 81–91.
- Simmons, J. P., & Simonsohn, U. (2017). Power posing: P-curving the evidence. *Psychological Science*, 28, 687–693. <https://doi.org/10.1177/0956797616658563>
- Stirman, S. W., & Pennebaker, J. W. (2001). Word use in the poetry of suicidal and nonsuicidal poets. *Psychosomatic Medicine*, 63(4), 517–522.
- Sugimoto, C. R., & Thelwall, M. (2012). Scholars on soap boxes: Science communication and dissemination in TED videos. *Journal of the American Society for Information Science and Technology*, 64, 663–674. <https://doi.org/10.1002/asi.22764>
- Sugimoto, C. R., Thelwall, M., Larivière, V., Tsou, A., Mongeon, P., & Macaluso, B. (2013). Scientists popularizing science: Characteristics and impact of TED talk presenters. *PLoS One*, 8, e62403. <https://doi.org/10.1371/journal.pone.0062403>
- Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29, 24–54. <https://doi.org/10.1177/0261927X09351676>
- TED (2012). TED reaches its billionth video view! In *TED Blog*. TED. www.blog.ted.com
- Tierney, W. G. (2006). *Trust and the public good: Examining the cultural traditions of academic work*. Peter Lang Inc.
- Tsou, A., Demarest, B., & Sugimoto, C. R. (2015). How does TED talk? A preliminary analysis. *iConference 2015 Proceedings*.
- Tsou, A., Thelwall, M., Mongeon, P., & Sugimoto, C. R. (2014). A community of curious souls: An analysis of commenting behavior on TED talk videos. *PLoS One*, 9, e93609. <https://doi.org/10.1371/journal.pone.0093609>
- Van Edwards, V., & Vaughn, B. (2015). 5 Secrets of a successful TED talk. *Science of People*. <https://www.scienceofpeople.com/secrets-of-a-successful-ted-talk/>
- Veletsianos, G., Kimmons, R., Larsen, R., Dousay, T. A., & Lowenthal, P. R. (2018). Public comment sentiment on educational videos: Understanding the effects of presenter gender, video format, threading, and moderation on YouTube TED talk comments. *PLoS One*, 13, e0197331. <https://doi.org/10.1371/journal.pone.0197331>
- Wuebben, D. (2016). Getting likes, going viral, and the intersections between popularity metrics and digital composition. *Computers and Composition*, 42, 66–79. <https://doi.org/10.1016/j.compcom.2016.08.004>
- Yarkoni, T. (2010). Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. *Journal of Research in Personality*, 44, 363–373. <https://doi.org/10.1016/j.jrp.2010.04.001>

How to cite this article: MacKrell K, Silvester C, Pennebaker JW, Petrie KJ. What makes an idea worth spreading? Language markers of popularity in TED talks by academics and other speakers. *J Assoc Inf Sci Technol*. 2021;1–11. <https://doi.org/10.1002/asi.24471>