How Does TED Talk? A preliminary analysis

Andrew Tsou, Indiana University
Bradford Demarest, Indiana University
Cassidy R Sugimoto, Indiana University

Abstract
TED Talks is one of the leading science communication initiatives in the digital age. Although previous work has analyzed the demographics of speakers & audience reaction to TED Talks, there is a dearth of research into the actual content of these talks. The transcripts for TED videos were downloaded from the official TED website and analyzed as to word use by different speaker classes (male academics, female academics, male non-academics, and female non-academics). The two subpopulations (males vs. females; academics vs. non-academics) exhibited marked differences in the words that they used during their talks, which may indicate different sentiments, topical preoccupations, and goals for the presentation. Gender was an important variable throughout the study, indicating an issue worthy of further investigation.
Keywords: TED Talks, science communication, CMC, linguistics
Citation: Tsou, A., Demarest, B., Sugimoto, C.R. (2015). How Does TED Talk? A Preliminary Analysis. In *iConference 2015 Proceedings*.

Acknowledgements:
Research Data: In case you want to publish research data please contact the editor.
Contact: atsou@umail.iu.edu, bdemares@indiana.edu, sugimoto@indiana.edu

1 Introduction

TED Talks (Technology, Entertainment, Design) is one of the leading science communication initiatives in the digital age. The organization reports that videos on its website have been viewed over a billion times, with millions more views on other video sharing platforms across the Internet (TED Blog website, 2012).

Although previous analyses have focused on aspects of the TED experience such as presenter demographics (Sugimoto et al., 2013) and the content of comments left on the TED site and YouTube mirrors of TED videos (Tsou, Thelwall, Mongeon, & Sugimoto, 2014), a linguistic analysis of TED Talks themselves has yet to be conducted. Accordingly, little is known about what, precisely, these popular science communicators are talking about (the TED site tags talks with keywords such as “entertainment” and “science,” but the content of the talks within these categories can, of course, vary wildly); it follows, then, that little is known about the different topical foci and considerations adopted by presenters of different genders and academic backgrounds.

The following analysis is an exploratory study of TED presentations, with an eye towards analyzing content differences based on presenters’ genders and academic affiliations (i.e., academics vs. non-academics). Specifically, the study aims to answer the following research questions:

RQ1: Do the contents of TED Talks differ according to the gender of the presenter, and if so, how are these differences manifested?

RQ2: Do the contents of TED Talks differ according to the presenter’s academic status, and if so, how are these differences manifested?

The TED website makes available transcripts for most of the videos on its website (most videos missing transcripts are those, such as musical performances, for which transcripts would not capture the essence of the presentation). A PHP script was used to automatically grab these transcripts, from which 188 were sampled for the purpose of this analysis. Four distinct groups of presenters – male academics, female academics, male non-academics, and female-non academics – were identified, and each were represented equally in the sample.

KWIC Concordance software was used to obtain word counts for the four different sets of transcripts, which then formed the basis for the de facto analysis. Chi-square tests were performed on a variety of words and population variables, thus allowing the transcript sets to be accurately compared despite their differing word counts.

A look at the use of pronouns by the various speakers is quite revealing. There were significant differences between the two genders, with men preferring the word "you," while women were more likely to use first-person addresses. This suggests that men are perhaps more likely to "preach" (given their preference for the word "you"), while women are more likely to give personal anecdotes (first-person singular) or tell stories (third-person). Crucially, women were more likely to use words such as "we," "our," and "us," suggesting that TED speakers are not equally concerned with the idea of audience inclusion.

Similarly, there were significant differences in pronoun use between academics and non-academics, with the scholars using more pronouns than the non-scholars in all cases except one: the "I" case ("me," "my," etc.). This suggests that scholars make a concentrated effort to engage with their audience directly, while non-academics are more likely to talk about themselves (perhaps because they were invited to give a TED talk on the basis of celebrity, whereas scholars were invited on the basis of academic merit?). This is supported by the observation that academics are far more likely to use "we" words. A summary of these results can be seen in Table 1 (all differences are significant at $p < 0.0001$).

	First	Second	Third	"We"
Males	3966	4157	2970	4624
Female	4884	3300	3259	4865
Academics	4178	4549	3854	5830
Non-academics	4672	2908	2375	3659

1.1 **Table 1. Use of pronouns according to different populations (all p-values are < 0.0001)**

Apart from these broad classes, the use of gendered pronouns by the speakers reveals significant differences in usage. Speaking strictly in terms of raw counts, the word "he" was used 984 times, while "she" was used 486 times (less than half as often!). In fact, male pronouns were used almost twice as frequently as female pronouns (n=846 and 1,634, respectively). As Table 2 reveals, there were significant differences in uses of gendered pronouns by all sub-groups.

	Female pronouns	Male pronouns
Male	303	899
Female	543	735
p-value	<.0001	0.0009
Academic	251	740
Non-academic	595	894

p-value	<.0001	<.0001

1.2 Table 2. Use of gendered pronouns according to subpopulation.

Given that women are more likely to use female pronouns than men are, one possible explanation is that women are more likely to discuss gender differences or inequalities even when discussing topics that are similar to those discussed by men. However, the fact that men used male pronouns significantly more than women did suggests that perhaps the reason for the difference is due to the use of personal anecdotes, which would possibly be more likely to involve people who were of the same gender as the speaker (due to the nature of friendships, etc.). Whereas the two genders tended to favor gendered pronouns that correlated with their own status, non-academics were simply more likely to use any gendered pronouns than their academic counterparts were. This may be due to non-academics having a preference for storytelling, which would naturally involve more gendered pronouns than more "factual" speeches would.

In addition to the pronoun differences, women used the word "women" more than 8 times as frequently as men did. A cursory analysis of the usage of the word "women" by male non-academics reveals that the word was used primarily in the context of story-telling (i.e., "went in, paid my ten pesos, walked up – 10 women, two guys"). Conversely, females (both academics and non-academics) discussed perceptions and situations relating to women's roles in matters as diverse as economics, verbal ability, sex and romance, education, voting, and the job market. Male academics occasionally mentioned "serious" women's issues, but again, many uses of female pronouns were in the context of anecdotes.

The use of "hedge words" was also analyzed. Interestingly, while there was no significant difference in hedge word usage by the two genders (p=0.418), academics used hedge words nearly twice as much as the non-academics (n = 5,076 and n = 2,886, respectively), for a p-value of less than 0.0001. One possible explanation is that the academics have the ability to see both sides of an issue, and thus are more interested in presenting a neutral overview of the situation (or as neutral as is possible) than in taking a specific stand. It could also be that academics chose to discuss topics that do not lend themselves to grandstanding or sweeping generalizations. Taking a more cynical view, one could also postulate that academics are simply trained to avoid committing themselves to a given stance, which would quite naturally involve the use of hedge words.

Further investigation was done on "popular and interesting" words; to qualify for this list, a word had to have been used at least 100 times by one of the subpopulations, and it also had to contain some semantic value beyond acting as a transition marker (e.g., "on" was excluded, while "can" was included, as the latter contained a sentiment that could be of potential interest). Chi-square tests were then done on each of these words between the two genders, as well as between the academics and non-academics. The results can be found in Table 3, which only contains words that exhibited statistically significant differences in frequency of use by the two groups in question.

Word	Male	Female	p-value	Academic	Non academic	p-value
can	1276	935	< .0001	1477	734	< .0001
not	880	906	0.16	999	787	0.04
very	720	598	0.007	817	501	0.0068
people	689	679	0.675	846	522	0.0082
think	480	577	0.0003	685	372	< .0001
really	495	598	0.0002	686	407	0.0028

actually	465	518	0.023	648	335	< .0001
world	416	385	0.57	386	415	< .0001
time	423	398	0.735	450	371	0.041
brain	208	176	0.204	311	73	< .0001
no	298	329	0.08	310	317	< .0001
fact	193	162	0.195	244	111	0.0001
power	174	44	< .0001	157	61	< .0001
china	124	13	< .0001	127	10	< .0001
idea	152	122	0.132	184	90	0.0027
data	124	84	0.012	181	27	< .0001
work	193	246	0.0035	245	194	0.284

1.3 **Table 3. "Popular and interesting" words that were used differently by the different sub-populations. P-values are presented to four decimal places.**

2 Conclusion

The two subpopulations (males vs. females; academics vs. non-academics) exhibited marked differences in the words that they used during their talks, which may indicate different sentiments, topical preoccupations, and goals for the presentation. It appears as if academics are more concerned with discussing facts, ideas, and "serious" issues than their non-scholarly counterparts, although the academics seemed less capable of engaging their audience, as laughter and applause were less frequent than would be expected during the scholars' talks. Gender was an important variable throughout the study, indicating an issue worthy of further investigation.

Further analysis would be required to determine the sentiments of academics' talks vs. non-academics' talks, which could be quite an interesting exercise, particularly given the academics' proclivity for including hedge words in their talks. Indeed, future studies could take as their base point any number of interesting variations within the data, as detailed content analysis would be required to fully take advantage of the richness of these data. Word use by gender is particularly telling, and it would be interesting to see if the gender differences seen in the use of the word "power" would carry over to other "loaded" terms that have historically been shown to be disproportionately favored by one gender.

References

Sugimoto, C.R., Thelwall, M., Lariviere, V., Tsou, A., Mongeon, P., & Macaluso, B. (2013). Scientists popularizing science: Characteristics and impact of TED Talk presenters. *PLoS ONE, 8*(4), e62403. doi:10.1371/journal.pone.0062403.

TED Blog website (2012) What talks have resonated most with you? Tweet TED's billionth video view. Retrieved from http://blog.ted.com/2012/11/13/what-talks-have-resonated-most-with-you-tweet-teds-billionth-video-view/

Tsou, A., Thelwall, M., Mongeon, P., & Sugimoto, C.R. (2014). A community of curious souls: An analysis of commenting behaviour on TED Talks videos. *PLoS ONE*.

Table of Figures

No table of figures entries found.

In your document, select the words to include in the table of contents, and then on the Home tab, under Styles, click a heading style. Repeat for each heading that you want to include, and then insert the table of contents in your document. To manually create a table of contents, on the Document Elements tab, under Table of Contents, point to a style and then click the down arrow button. Click one of the styles under Manual Table of Contents, and then type the entries manually.

Table of Tables

No table of figures entries found.

In your document, select the words to include in the table of contents, and then on the Home tab, under Styles, click a heading style. Repeat for each heading that you want to include, and then insert the table of contents in your document. To manually create a table of contents, on the Document Elements tab, under Table of Contents, point to a style and then click the down arrow button. Click one of the styles under Manual Table of Contents, and then type the entries manually.