The 2016 Election:  
What Did Pollsters and Pundits Miss?

On the morning of 8 November, pollsters predicted Hillary Clinton would easily be elected the 45th President of the United States; by late that evening they were proven wrong in one of the most surprising electoral outcomes in US history. In an era of big data, powerful mass analytic tools, micro-polling, and the lessons of the 2012 election, how could pollsters have been so wrong? What might have improved their prognostications? What did they miss?

Polling was unable to take into account what social media was telling us about voters. Pollsters may also have fallen victim to the impact that communications technology changes have had on communications culture. Americans appear less and less likely to respond to telephone canvassing or direct questioning but many are eager to volunteer—and promote—their views online. With more and more people sharing views and obtaining news online, greater attention to online content is now warranted.

Social Media and Polling

Analyzing Internet content is clearly not identical to polling. The Internet makes it difficult to identify audience segments and demographic details.

- Unlike polling, social media users are not necessarily representative of the population.
- Pollsters target audiences to construct samples representative of a desired cohort whom they can question. Internet users of typically unidentifiable demographics volunteer content to advance a view.

Social media users have clout.

- We have found that Internet users often have an impact on society disproportionate to their numbers—they tend to be more proactive and activist, adept at sparking reactions, and they have the stage necessary to make a case and attempt to convince others.
The View from the Web

Babel Street’s review of social media (over 300 million documents) related to Trump and Clinton in the year prior to the election consistently yielded a different picture than that depicted by pollsters and pundits. We found that online users expressed considerably more positive than negative sentiment toward Trump than polls were suggesting, and this positive sentiment was largely consistent throughout the year, including after debates that pundits said Trump lost. In contrast, positive Internet content related to Clinton was well below that for Trump during the year.¹

Negative sentiment surrounding both candidates was fairly identical over the year.

This analysis is based on open source information reviewed between October 2015 and October 2016 from over 30 social media sites and digital news that referenced candidates Trump and Clinton or their social media handles. We minimized content not directly related to the election. Babel Street analytics assigned a sentiment score to each mention of the candidates. Babel Street measures sentiment in five degrees: very positive, positive, neutral, negative, and very negative. For the purposes of our analysis, we defined positive as a combination of “positive” and “very positive” and defined negative responses similarly. A more detailed explanation of Babel Street’s methodology, including sentiment, may be found at the end of this article.

¹ Mentions of Trump outpaced mentions of Clinton by two to one (204 million vs. 100 million). To account for this difference we looked at positive and negative mentions as a share of total mentions for each candidate.
Sustained Positive Sentiment Toward Trump

President-elect Trump received nearly double the attention in social media than did Secretary Clinton. Some of this difference can be explained by Trump’s more active use of social media.

Given this volume difference, it is not surprising that Trump had more social media mentions associated with positive and negative sentiment than Clinton.

A more revealing picture emerges when we looked at positive and negative mentions as a share of total mentions, controlling for the difference in volume. Positive sentiment attached to social media mentions of Trump as a share of all of his social media mentions were consistently higher than those for Clinton. The gap narrowed slightly as the election approached.

In sum, positive sentiment associated with social media mentions about Trump as a share of his total mentions consistently and significantly outweighed those of Clinton, while each candidates’ share of negative sentiment associated with their mentions were persistently similar.

“'I think that social media has more power than the money they spent, and I think maybe to a certain extent, I proved that.”

-- President-elect Trump during 60 Minutes interview (11/13/16)
Sentiment Toward Trump More Consistent

Babel Street also looked at daily changes in sentiment associated with social media mentions of the two candidates from October 2015 to October 2016. We found more volatility in the positive sentiment toward Clinton than positive sentiment for Trump. Variation in negative sentiment for both candidates was fairly consistent and daily variation in negative sentiment for both candidates decreased somewhat as the election approached.

In sum, social media users were more fickle in their positive sentiment toward mentions of Clinton during the last year.
We are not suggesting that analysis of social media by itself would have been sufficient to predict which candidate would win. We acknowledge that our insights are based only on those who expressed views on the Internet in the last year and not necessarily on the larger voting population.

But we believe an election where sophisticated polling techniques did such a poor job of predicting the winner speaks to the importance of Internet collection and analytics. Analyzing the massive amounts of information that Babel Street collected provided unique insights into how sentiment toward the candidates changed over time and which issues resonated most with audiences. We believe complementing large scale analysis of Internet content with traditional polling methods would likely have resulted in a more accurate assessment of the prospects for the two candidates.

BY: PATRICK O’NEIL (BABEL STREET CHIEF ANALYTIC OFFICER) AND ABBIGAIL DOBBERTIN (BABEL STREET LEAD ANALYST) WITH CONTRIBUTIONS FROM: CALEB LOOMIS (BABEL STREET ASSISTANT DIRECTOR) AND TUCKER HOLMES (BABEL STREET INTERN)

The Role of Internet Analytics
We are not suggesting that analysis of social media by itself would have been sufficient to predict which candidate would win. We acknowledge that our insights are based only on those who expressed views on the Internet in the last year and not necessarily on the larger voting population.

But we believe an election where sophisticated polling techniques did such a poor job of predicting the winner speaks to the importance of Internet collection and analytics. Analyzing the massive amounts of information that Babel Street collected provided unique insights into how sentiment toward the candidates changed over time and which issues resonated most with audiences. We believe complementing large scale analysis of Internet content with traditional polling methods would likely have resulted in a more accurate assessment of the prospects for the two candidates.

BY: PATRICK O’NEIL (BABEL STREET CHIEF ANALYTIC OFFICER) AND ABBIGAIL DOBBERTIN (BABEL STREET LEAD ANALYST) WITH CONTRIBUTIONS FROM: CALEB LOOMIS (BABEL STREET ASSISTANT DIRECTOR) AND TUCKER HOLMES (BABEL STREET INTERN)
Methodology Notes

Babel Street

Babel Street is a geo-enabled, multi-lingual text analytics platform capable of evaluating, interrogating, and analyzing very large amounts of diverse data. We calibrate filters to focus collection on areas of interest and minimize noise (content not directly associated with the focus of the query). Babel Street was able to review persistently massive amounts of online information related to the two candidates, quantify the volume, identify the key terms within the content, and attach sentiment values to those terms.

Sentiment Analysis

Babel Street measures sentiment by employing proprietary algorithms to the data and scoring them on a scale from -1000 to +1000 based on key terms and their proximate location. Sentiment is measured more accurately when the data set is large, as in the case of this exercise (over 300 million documents). In measuring sentiment, we have found changes in sentiment to be more reliable than point estimates of sentiment. For example, we have more confidence in assessing that an audience is becoming more positive about an issue than pinpointing a precise degree of positive sentiment.

Collection Methodology

We establish separate filters for mentions of the entities “Donald Trump” and “Hillary Clinton” and their social media handles. Babel Street’s search capabilities translate these terms into multiple variations including nicknames, slang terms, euphemisms, and emojis drawing from an ontology of over 1 billion terms.