Natural Language Processing:
Unlocking New Frontiers in Corporate Earnings Sentiment Analysis

February 22, 2018

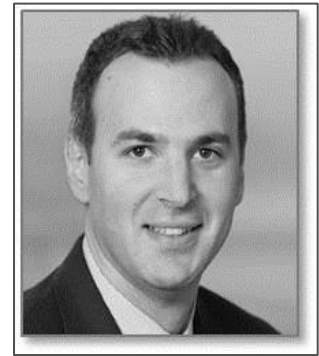**S&P Global**
Market Intelligence

Today's Speakers

Moderator:
Daniel J. Sandberg, PhD, CFA
Director, Quantamental Research
S&P Global Market Intelligence

Frank Zhao
Director, Quantamental Research
S&P Global Market Intelligence

Michael Patton
Senior Director, Data Management Solutions
S&P Global Market Intelligence

S&P Global
Market Intelligence

Agenda

1. Motivation – What is NLP and why is it important?

2. General NLP steps

3. Actionable ideas – How can NLP help me?
   - Sentiment analysis: stock- and industry-level
   - Earnings call segmentations: prepared remarks, Q&A and speakers
   - Behavioral finance: softening of bad news may manifest via language complexity and analyst selectivity

4. Implementation – Where do I get started?

5. Loughran and McDonald (2011)

6. Transcripts data

S&P Global
Market Intelligence

Motivation – What is NLP?

What is **Natural Language Processing (NLP)?**

- Leveraging computers and statistics to process and make sense of language in a systematic and sensible way without human intervention besides coding.

What are **Machine Learning Algorithms (ML)**?

- Tools to do NLP
- Statistics - what data scientists call applied statistics

Two major categories

- **Supervised**
  - Researchers provide rules
  - Empirical researchers have been using tools in this category for decades
    - E.g., regression, Bayesian statistics
- **Unsupervised**
  - Let the algorithms find patterns (more black-box)
    - E.g., clustering, principal component analysis

- See **NLP primer** for details.

S&P Global
Market Intelligence

Motivation – Why is NLP Important?

- **Forty zettabytes (10^21 bytes) of data** are projected to be on the internet by 2020, out of which more than eighty percent of the data are unstructured in nature, requiring NLP to process and understand[1]

- "In the past two days, the amount of unstructured data that was created (on the internet) is equivalent to the the same amount that was created **from the beginning of human civilization through the end of 2003**"[2] – Eric Schmidt, Executive Chairman of Alphabet/Google 2001 - 2017

- NLP can read through documents at a rate far more rapidly than humans. This allows textual data to be **more efficiently** processed, consumed and made sense at a **much larger scale**.

[1] Mearian, L. (2012, Dec. 11). By 2020, there will be 5,200 GB of data for every person on Earth. Retrieved from http://www.computerworld.com.
[2] DeAngelis, S. F. (2014, Feb.). The Growing Importance of Natural Language Processing [Web log post]. Retrieved Aug 8, 2017, from http://www.wired.com/

S&P Global
Market Intelligence

Agenda

1. Motivation – What is NLP and why is it Important?

2. General NLP Steps

3. Actionable ideas – How can NLP help me?

   • Sentiment analysis: stock and industry-level

   • Earnings call segmentations: prepared remarks, Q&A and speakers

   • Behavioral finance: softening of bad news may manifest via language complexity and analyst selectivity

4. Implementation – Where do I get started?

5. Loughran and McDonald (2011)

6. Transcripts data

S&P Global
Market Intelligence

NLP General Steps

- **Step 1: Text Preprocessing**
  - Noise Removal
  - Lexicon Normalization (e.g., plays, played => play)
  - Object Standardization (e.g., Luv => love)

- **Step 2: Text to Features**
  - N-grams
    - Bag-of-words (i.e., unigram)
  - Syntactical parsing
    - Grammar tree or part-of-speech tagging
  - Statistical features
    - E.g., # of tokens (words, numbers , abbreviations), # of sentences

- **Step 3: Testing, Refinement and Efficacy**
  - Where model is calibrated, refined and tested.

- See **NLP primer** for details

S&P Global
Market Intelligence

Agenda

1. Motivation – What is NLP and why is it Important?

2. General NLP Steps

3. Actionable ideas – How can NLP help me?

   - Sentiment analysis: stock and industry-level

   - Earnings call segmentations: prepared remarks, Q&A and speakers

   - Behavioral finance: softening of bad news may manifest via language complexity and analyst selectivity

4. Implementation – Where do I get started?
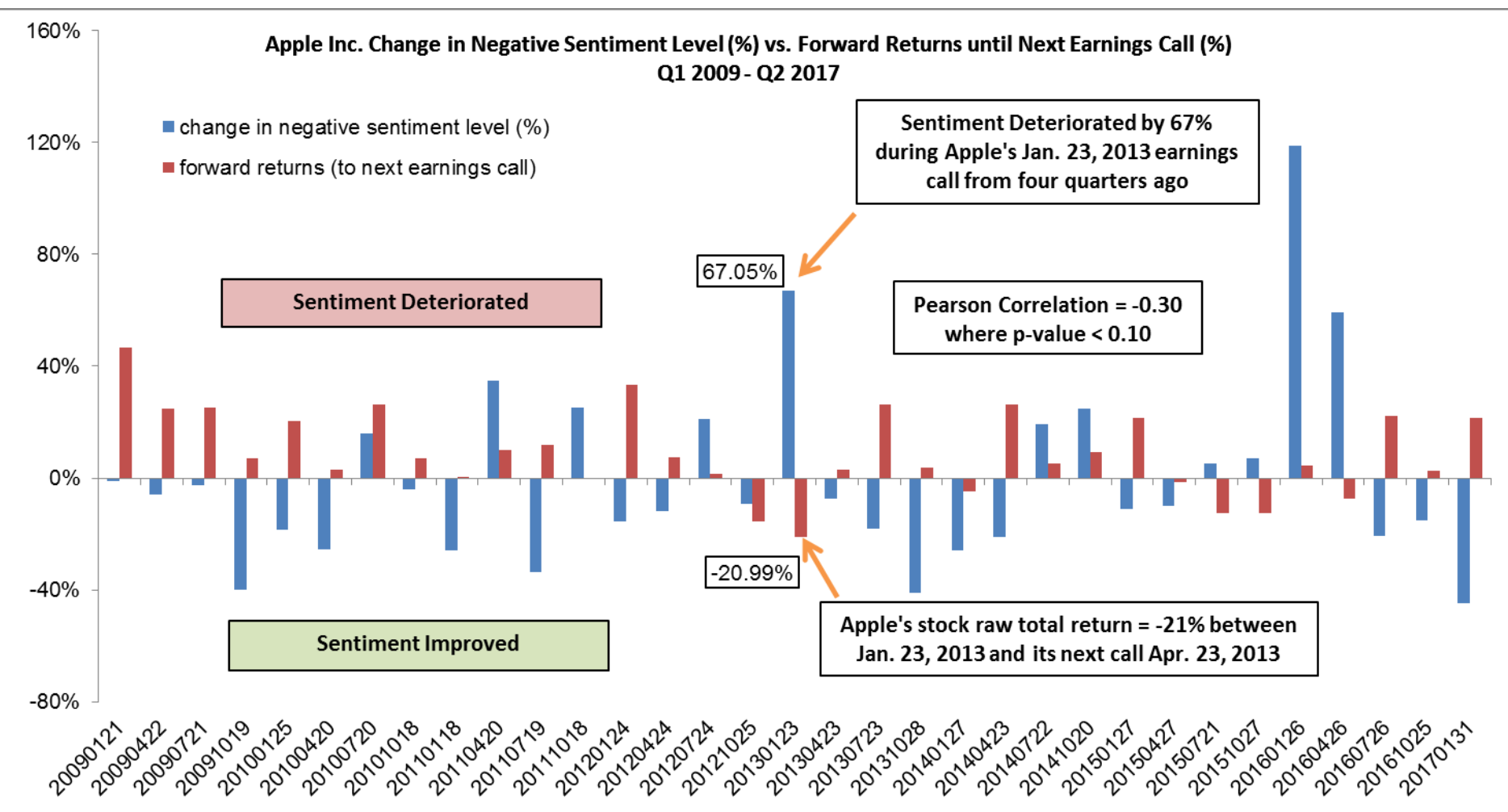
5. Loughran and McDonald (2011)

6. Transcripts data

S&P Global
Market Intelligence

Actionable Ideas

Sentiment analysis:

Stock- and industry-level

Actionable Ideas: Stock-Level Sentiment Trends vs. Forward Returns

Apple Inc. Change in Negative Sentiment Level (%) vs. Forward Returns until Next Earnings Call (%)
Q1 2009 - Q2 2017

- change in negative sentiment level (%)
- forward returns (to next earnings call)

Sentiment Deteriorated by 67% during Apple's Jan. 23, 2013 earnings call from four quarters ago

67.05%

Sentiment Deteriorated

Pearson Correlation = -0.30 where p-value < 0.10

-20.99%

Sentiment Improved

Apple's stock raw total return = -21% between Jan. 23, 2013 and its next call Apr. 23, 2013
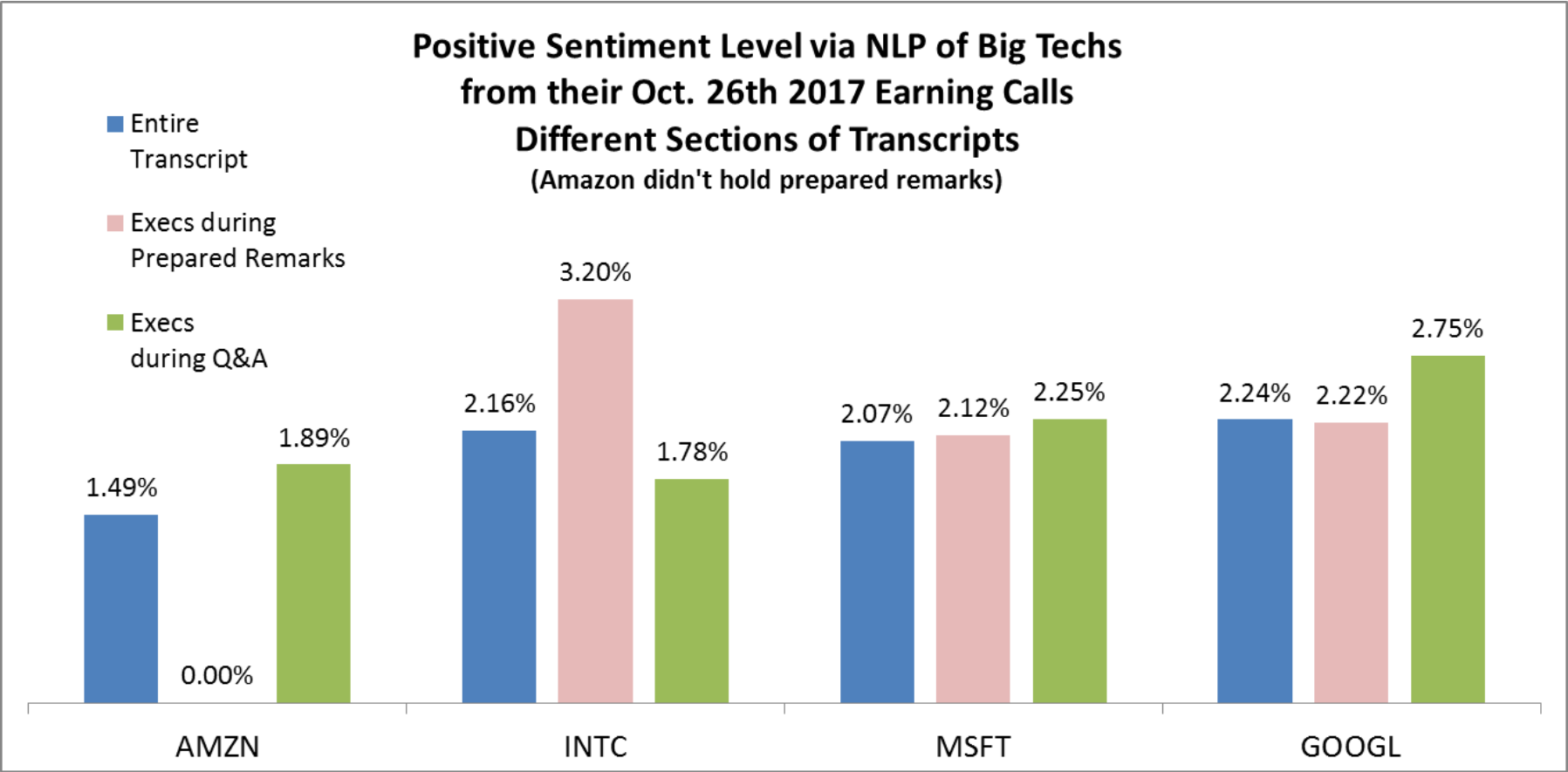
Source: S&P Global Market Intelligence Quantamental Research as of 8/8/17 (charts are for illustrative purposes only)

S&P Global
Market Intelligence

Actionable Ideas - S&P 500 Earnings Call Sentiment Trends

GICS Industry Group Level Sentiment Trends

GICS Industry Groups	Calendar Quarters			
	Q3 2016	Q4 2016	Q1 2017	Q2 2017
Consumer Services	-4.1%	-4.0%	16.9%	20.1%
Software & Services	-0.2%	4.1%	-3.9%	19.2%
Transportation	-3.6%	-2.8%	14.6%	19.1%
Diversified Financials	5.4%	13.8%	16.5%	18.0%
Technology Hardware & Equipment	4.8%	7.1%	18.1%	17.5%
Banks	-3.7%	-0.4%	18.1%	16.8%
Capital Goods	-4.9%	9.1%	16.9%	15.5%
Commercial & Professional Services	7.6%	-4.4%	-6.2%	14.6%
Utilities	-4.6%	7.2%	7.0%	14.1%
Insurance	0.0%	14.4%	10.2%	11.9%
Energy	0.1%	13.6%	25.8%	10.3%
Real Estate	-11.5%	-2.4%	0.5%	5.5%
Media	-11.6%	-12.6%	5.0%	4.2%
Materials	5.7%	-1.8%	12.2%	1.4%
Semiconductors & Semiconductor Equipment	6.3%	-3.7%	14.5%	0.5%
Retailing	-18.4%	-0.3%	2.4%	-1.6%
Consumer Durables & Apparel	10.7%	0.2%	-4.7%	-2.6%
Pharmaceuticals, Biotechnology & Life Sciences	-5.7%	-3.6%	-1.3%	-2.9%
Household & Personal Products	-3.3%	-2.9%	-7.8%	-3.9%
Food & Staples Retailing	17.7%	-0.8%	-16.5%	-4.0%
Food, Beverage & Tobacco	2.2%	0.4%	-3.6%	-5.0%
Health Care Equipment & Services	-8.2%	-2.7%	-1.0%	-7.5%
Autos & Components	-14.6%	-4.3%	34.2%	-8.5%
Telecommunication Services	9.1%	-13.6%	-36.9%	-29.4%

Sentiment Improved

Sentiment Deteriorated

Source: S&P Global Market Intelligence Quantamental Research as of 8/8/17 (charts are for illustrative purposes only)

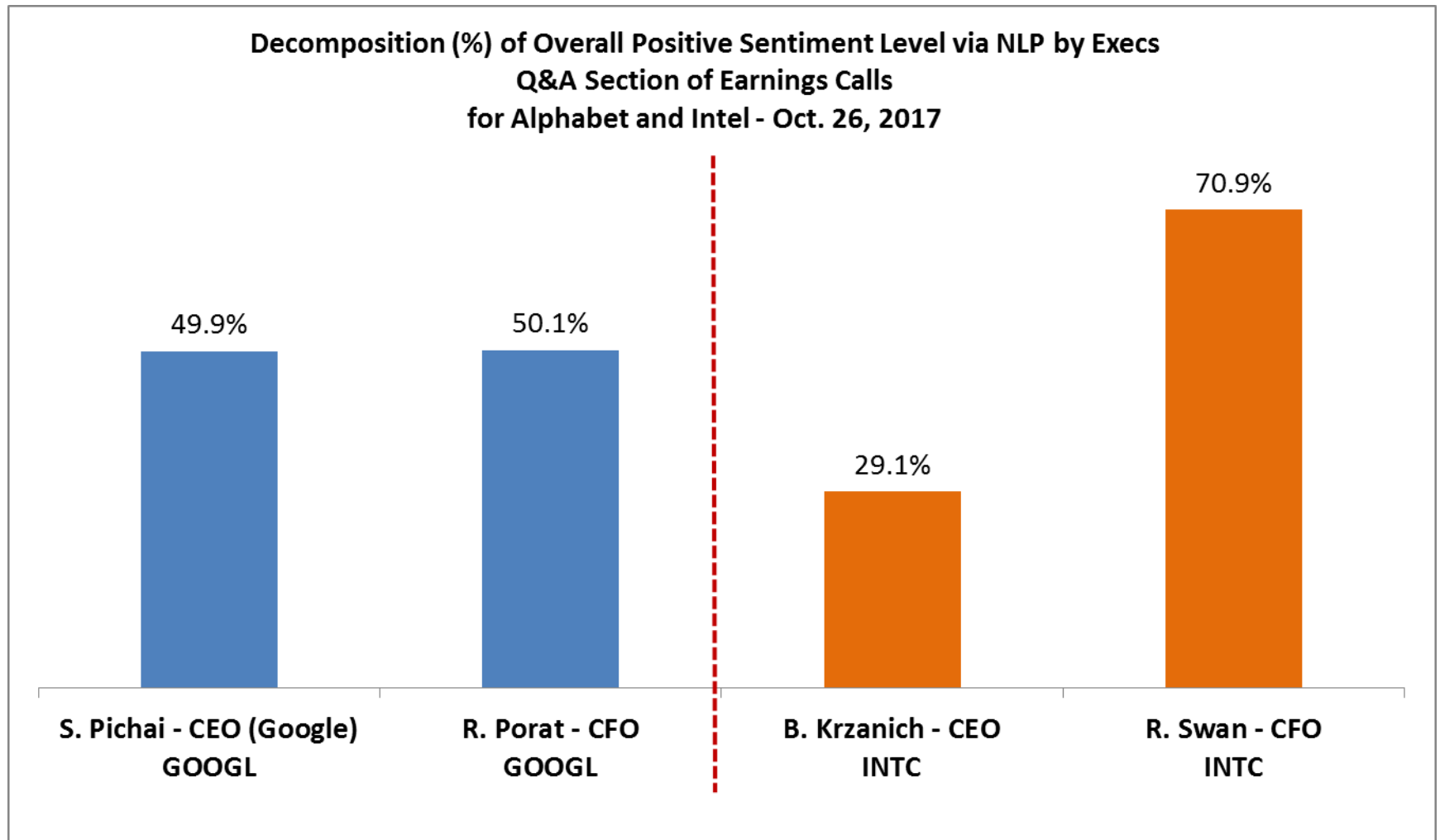**S&P Global**
Market Intelligence
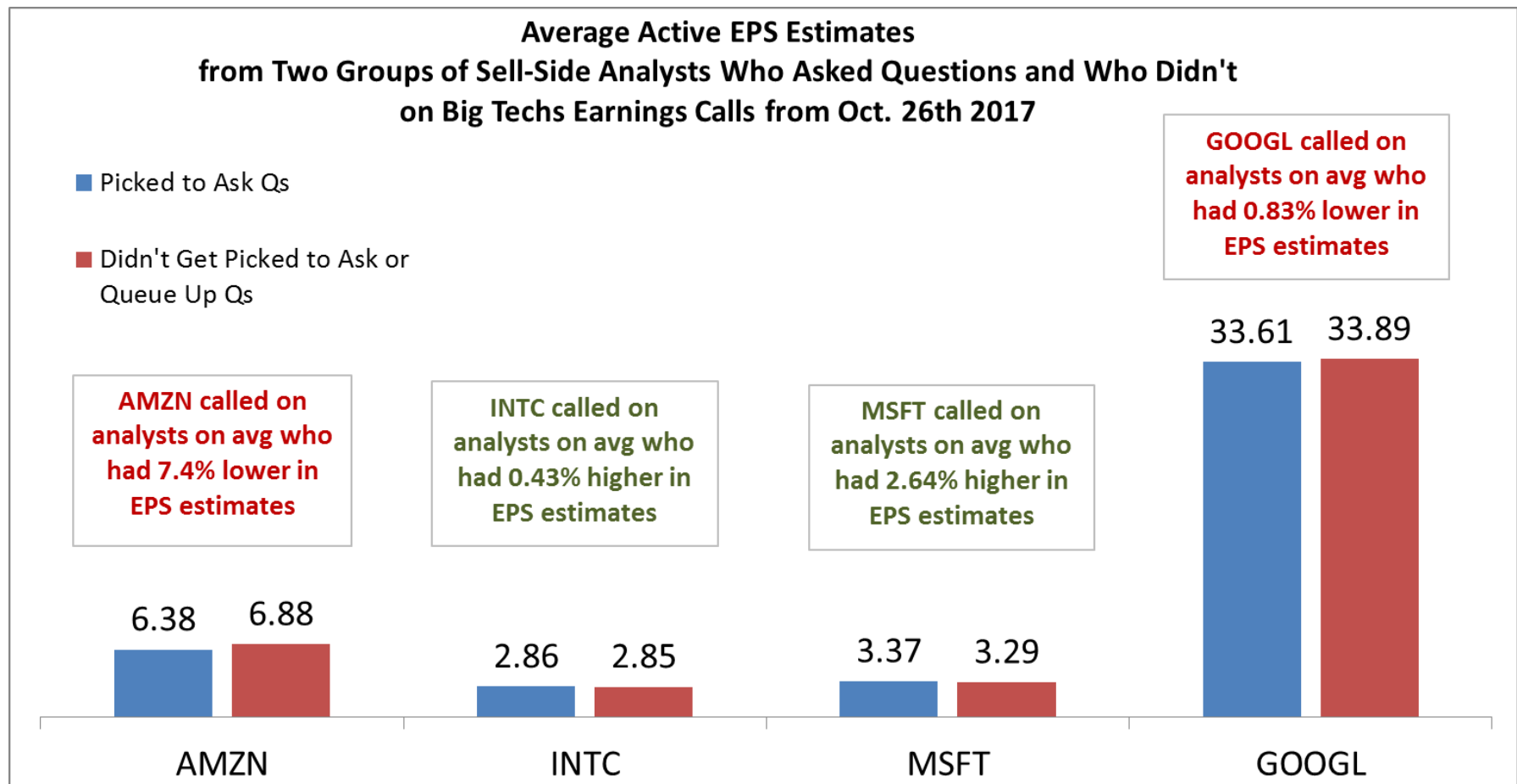
Actionable Ideas

Earnings call segmentations:

Prepared remarks, Q&A and speakers

Actionable Ideas - NLP and Transcripts Segmentation Example

Positive Sentiment Level via NLP of Big Techs from their Oct. 26th 2017 Earning Calls Different Sections of Transcripts (Amazon didn't hold prepared remarks)

Legend:
- Entire Transcript
- Execs during Prepared Remarks
- Execs during Q&A

AMZN: 1.49%, 0.00%, 1.89%
INTC: 2.16%, 3.20%, 1.78%
MSFT: 2.07%, 2.12%, 2.25%
GOOGL: 2.24%, 2.22%, 2.75%

S&P Global
Market Intelligence

Actionable Ideas – NLP and Transcripts Meta-Tagging Example

Elon Musk's Sentiment Past Nine Qtrs During Q&A
sentiment = (pos words - neg words) / total words
(higher means more positive; most recent one highlighted in red)

Date	Value
20151103	-0.43%
20160210	-0.03%
20160504	0.25%
20160803	0.70%
20161026	0.24%
20170222	-0.35%
20170503	0.60%
20170802	0.26%
20171101	0.14%
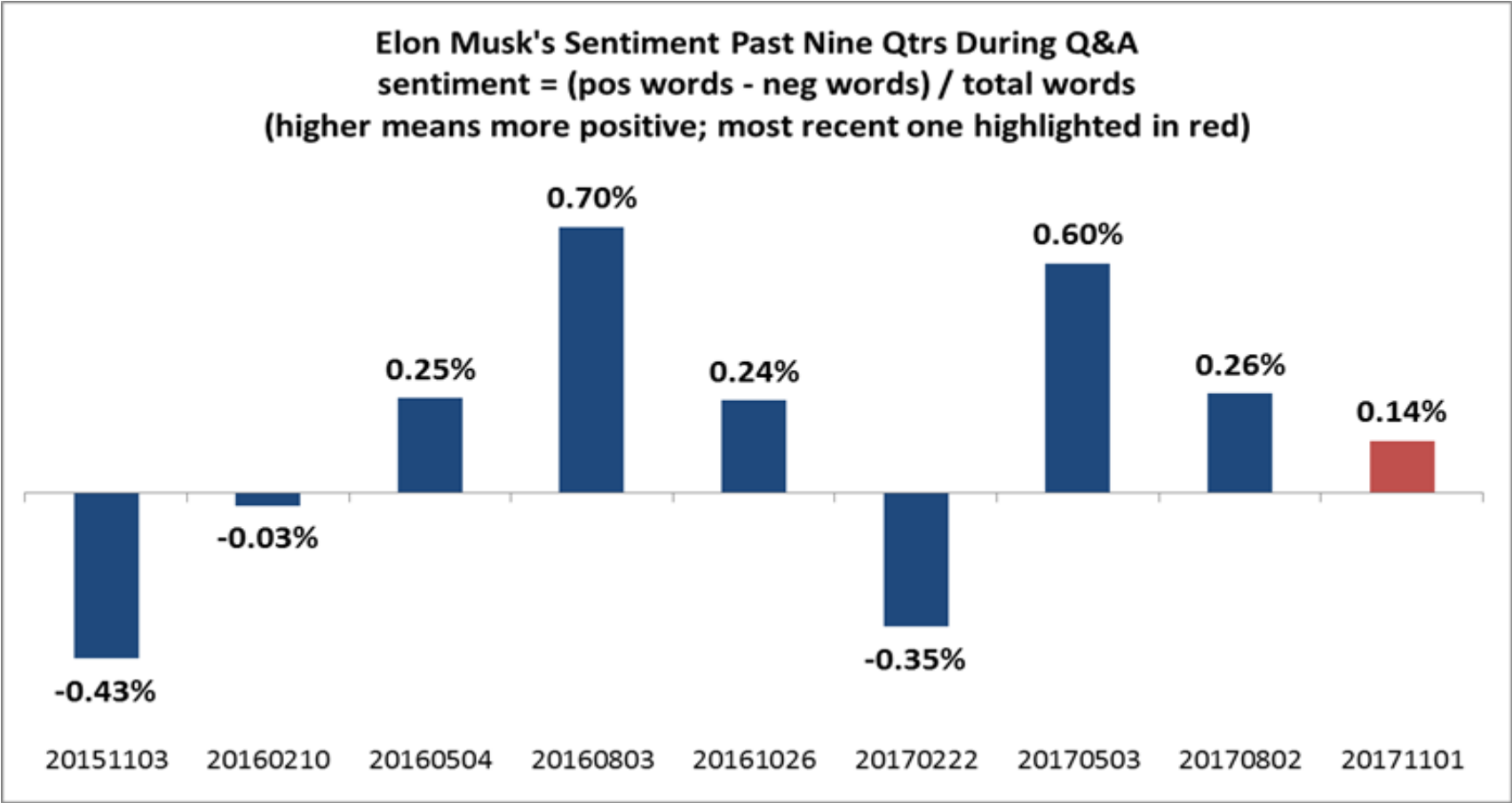
Source: S&P Global Market Intelligence Quantamental Research as of 11/3/17 (charts are for illustrative purposes only)

S&P Global
Market Intelligence

Actionable Ideas: NLP and Transcripts Meta-Tagging Example – Executives Tagging

Breakdown of sentiment level by executives who spoke

Decomposition (%) of Overall Positive Sentiment Level via NLP by Execs
Q&A Section of Earnings Calls
for Alphabet and Intel - Oct. 26, 2017

- S. Pichai - CEO (Google) GOOGL: 49.9%
- R. Porat - CFO GOOGL: 50.1%
- B. Krzanich - CEO INTC: 29.1%
- R. Swan - CFO INTC: 70.9%

S&P Global
Market Intelligence

Actionable Ideas – NLP and Transcripts Sell-Side Estimates Example

**Average Active EPS Estimates
from Two Groups of Sell-Side Analysts Who Asked Questions and Who Didn't
on Big Techs Earnings Calls from Oct. 26th 2017**

- ■ Picked to Ask Qs
- ■ Didn't Get Picked to Ask or Queue Up Qs

GOOGL called on analysts on avg who had 0.83% lower in EPS estimates

AMZN called on analysts on avg who had 7.4% lower in EPS estimates

INTC called on analysts on avg who had 0.43% higher in EPS estimates

MSFT called on analysts on avg who had 2.64% higher in EPS estimates

	Picked to Ask Qs	Didn't Get Picked
AMZN	6.38	6.88
INTC	2.86	2.85
MSFT	3.37	3.29
GOOGL	33.61	33.89

Source: S&P Global Market Intelligence Quantamental Research as of 10/31/17 (charts are for illustrative purposes only)

S&P Global
Market Intelligence

16

Actionable Ideas

Behavioral finance:

Softening of bad news may manifest via language complexity and analyst selectivity

Language Complexity and Analyst Engagement S&P500 Q2 Conference Calls
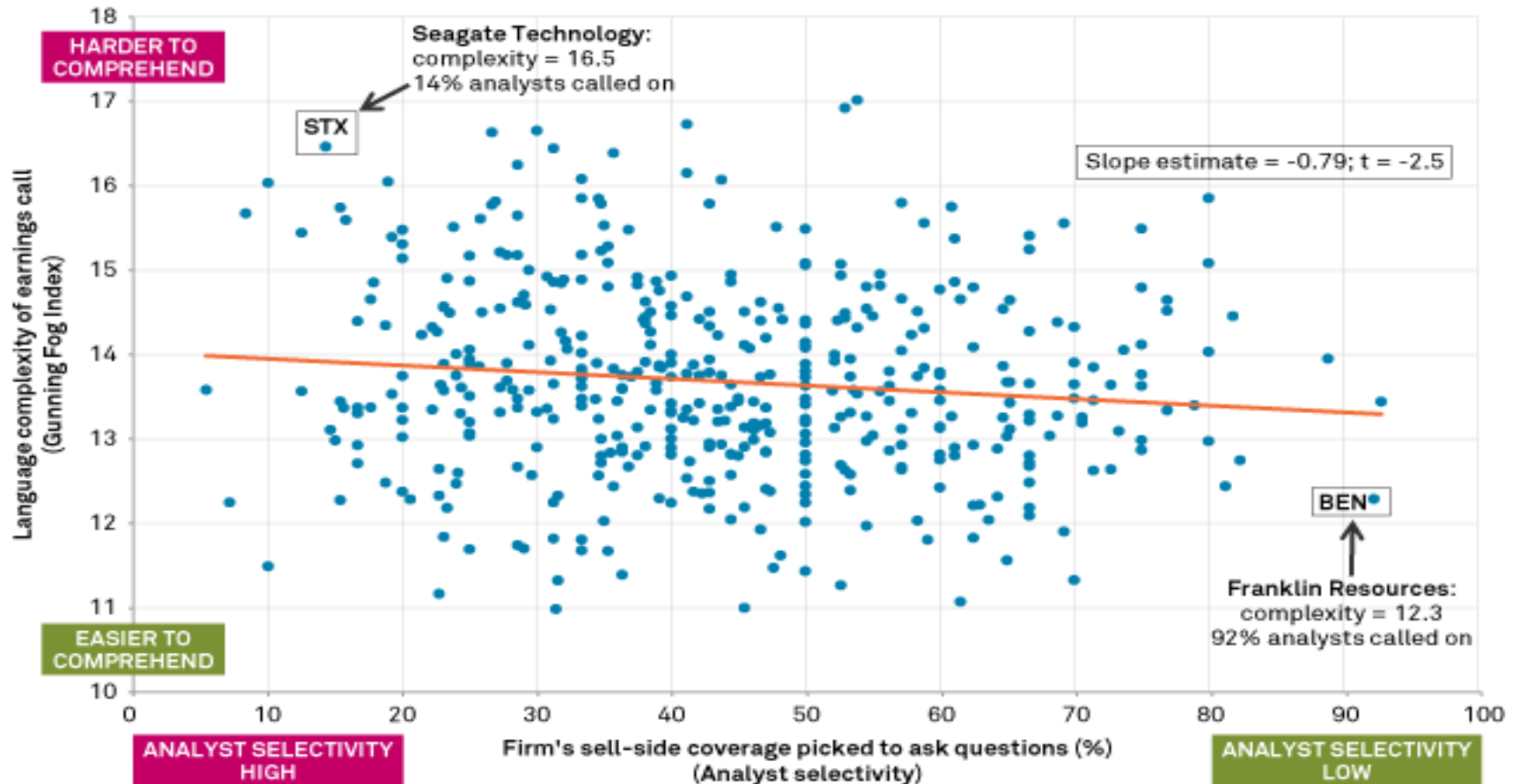
Source: S&P Global Market Intelligence Quantamental Researchas of 8/8/17 (charts are for illustrative purposes only)

Analyst Engagement and Stock Performance – S&P 500 Q2 Earnings Calls
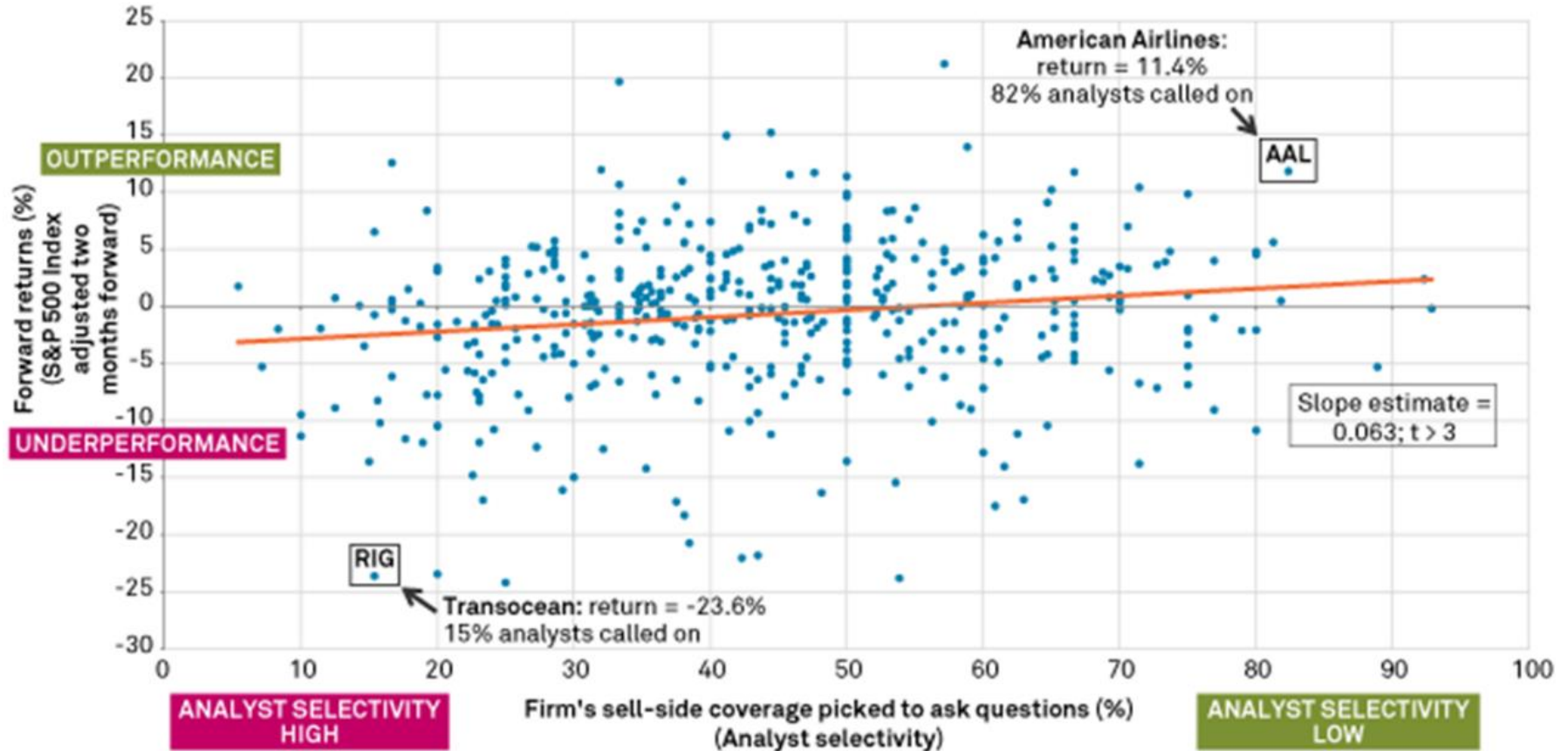
Source: S&P Global Market Intelligence Quantamental Research as of 8/8/17 (charts are for illustrative purposes only)

S&P Global
Market Intelligence

Agenda

1. Motivation – What is NLP and why is it Important?

2. General NLP Steps

3. Actionable ideas – How can NLP help me?

   • Sentiment analysis: stock and industry-level

   • Earnings call segmentations: prepared remarks, Q&A and speakers

   • Behavioral finance: softening of bad news may manifest via language complexity and analyst selectivity

4. Implementation – Where do I get started?

5. Loughran and McDonald (2011)

6. Transcripts data

S&P Global
Market Intelligence

Implementation - Where do I get started?

- We made about 400+ rows of Python <u>code</u> available in the public domain to do basic NLP analysis using earnings call transcripts

- It generates the following analytics historically for an investable universe such as S&P 500
    - Number of tokens (words, #s, etc.)
    - Sentiment scores (positive, negative, etc.)
    - Language complexity using Gunning fog index (# of sentences, syllables for each word)

- Access to the <u>code</u> is available via the primer

S&P Global
Market Intelligence

Agenda

1. Motivation – What is NLP and why is it Important?

2. General NLP Steps

3. Actionable ideas – How can NLP help me?
   - Sentiment analysis: stock and industry-level
   - Earnings call segmentations: prepared remarks, Q&A and speakers
   - Behavioral finance: softening of bad news manifests via language complexity and analyst selectivity

4. Implementation – Where do I get started?

5. Loughran and McDonald (2011)

6. Transcripts data

S&P Global
Market Intelligence

Dictionary Selection – Loughran and McDonald

- **Utilized the Loughran and McDonald (2011) business dictionary <u>https://www3.nd.edu/~mcdonald/Word_Lists.html</u>)**

- **Advantages**

    - **Accessibility** – Word lists are readily accessible because they are freely posted online.

    - **Comprehensiveness** – The dictionary is comprehensive such that it is difficult for some managers to game the system (i.e., circumvent certain words that have empirically been shown to lead to future stock underperformance) because they start with every conceivable English word with all inflections of a word, totaling 80+K distinct words in the master word list.

    - **Financial-specific context** – The initial master word is filtered down to their sentiment word lists by examining 10-K filings between 1994 and 2008 inclusively.

    - **Permanence of words** – The master and sentiment word lists are less transitory because they start with the most comprehensive list of English words possible and, more importantly, the master word list doesn't rely on transitory terms such as iphone.

    - **Unambiguous and Singularly Connoted Words** – Sentiment word lists are unambiguous and singularly connoted words by looking at the most frequently occurring words in the 10-Ks from the master word list. From there, they went word-by-word and assessed each of the word's meaning in a business context. At the end of their process, the words that ended up in their word lists are less ambiguous in their meaning with singular connotation

- **See <u>primer</u> for details**

Event Study Results – Loughran and McDonald

Signal: Positive Sentiment Level
Event Study [t+2, t+23]
S&P 500 2008 - 2017

- Annualized Median Market-Adjusted Returns (lhs)
- Stock-Level Hit Rate: % of stocks have positive returns (rhs)

Most Positive: 2.39%
2: 1.03%
3: 0.71%
4: -0.58%
Least Positive: -3.83%
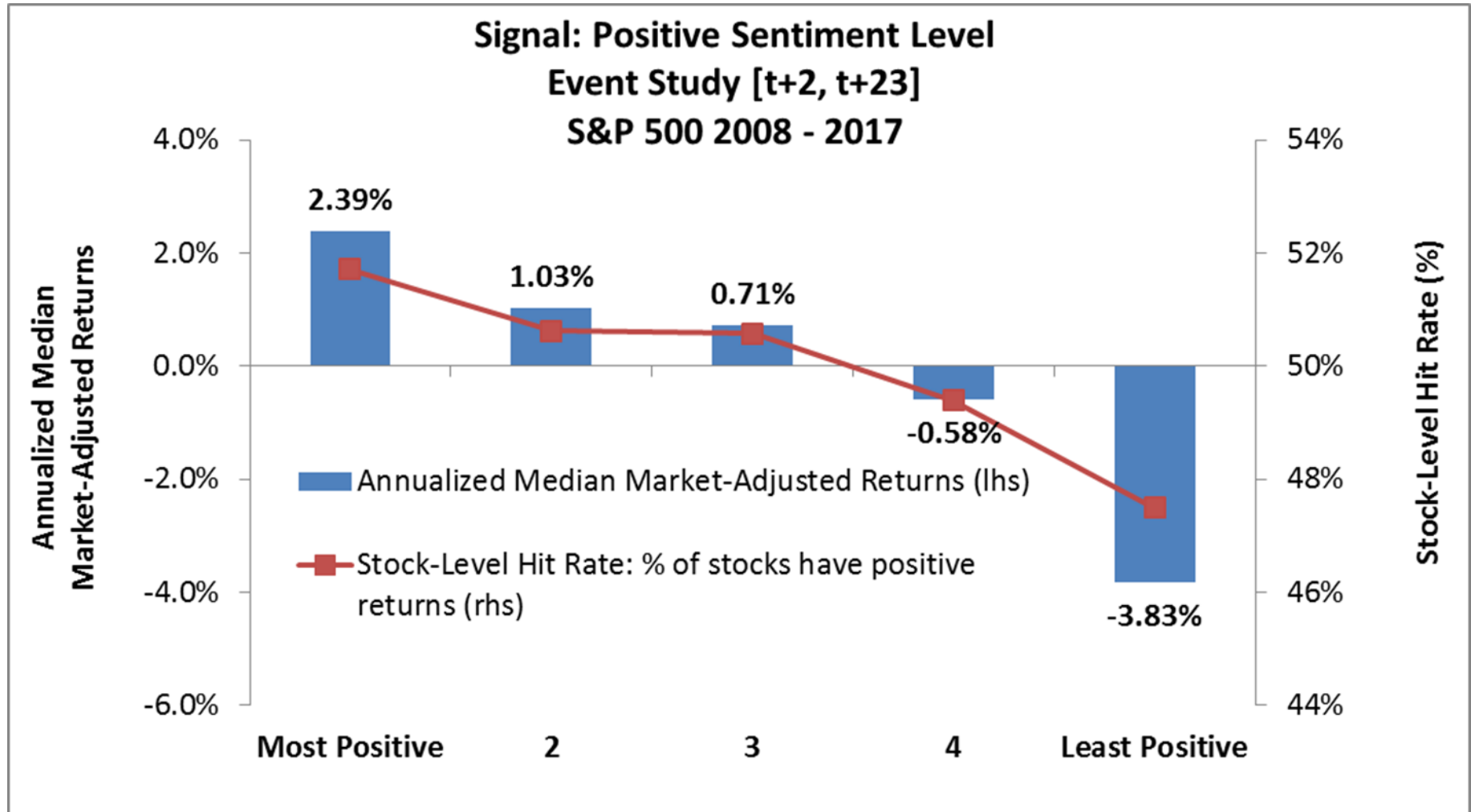
Source: S&P Global Market Intelligence Quantamental Research as of 02/01/18 (charts are for illustrative purposes only)

Agenda

1. Motivation – What is NLP and why is it Important?

2. General NLP Steps

3. Actionable ideas – How can NLP help me?

   - Sentiment analysis: stock and industry-level
   - Earnings call segmentations: prepared remarks, Q&A and speakers
   - Behavioral finance: softening of bad news may manifest via language complexity and analyst selectivity

4. Implementation – Where do I get started?

5. Loughran and McDonald (2011)

6. Transcripts data

S&P Global
Market Intelligence

Unstructured/Alternative Data
METADATA Pieces It All Together

Gregory R. Friedman Vice President of Investor Relations

"As you know, Trian Fund Management has invested in DuPont stock and proposed a plan to break up the company, a spin upon a spin. Our Board of Directors and management are unanimous in their belief that the plan we are pursuing will continue to deliver superior value for all DuPont shareholders."

26222634

768440

538648891

26858581

Jeffrey John Zekauskas JP Morgan Chase & Co, Research Division

"I was looking at the funds flow statement, and your change in operating assets and liabilities was a use of $1.3 billion, and the previous year was $1.4 billion. What's behind that change? In that your receivables and inventories don't really change that much, but there's something that's using up cash. And do you have a forecast for 2015 for that number? "

658776

Ellen J. Kullman Former Chairwoman and Chief Executive Officer

"Yes, Jeff. So you have a knack for going to a very specific place, and Nick has just now told me he's figured it out. So Nick? "

25051680

Nicholas C. Fanandakis Chief Financial Officer and Executive Vice President

"Jeff, I think the biggest thing there is the absence of the tax payment for the coatings sale. We had about $700 million of tax payment that shows as part of that free cash flow, but the inflow of cash is below the free cash flow line in the investments area. So that's the biggest reason for the delta on a year-over-year basis"

25912184
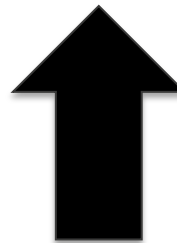
Estimates	Ownership	Cap Structure	Cross Reference	Professional Data
Financials	Events	Key Devs	People	Transcripts

S&P Global
Market Intelligence

Transcripts – Coverage Stats

- **Event Types**

  - Earnings Calls
  - Guidance/Update Calls
  - M&A Calls
  - Sales/Trading Statement Calls
  - Special Calls
  - Shareholder/Analysts Calls

  - Company Conference Presentations
  - Analyst/Investor Day
  - Operating Results Calls
  - Interim Management Statement Calls
  - Fixed Income Calls

- **Current Coverage –** 8000+ global public companies
- **History** – back to 2004
- **Timeliness** – Significant number of all calls are transcribed in real-time. All other calls are completed within 24 hours
- **Instances of Calls –** Preliminary, Edited, Proofed, Audited
- **Timestamps –** Currently available

Source: S&P Global Market Intelligence as of 2/1/18

S&P Global
Market Intelligence

Transcripts – Xpressfeed Packages

Delivery

- Package delivery frequency consists of Full Files and Change Files:
  - **Full Files**: Delivered weekly and applied once to initialize the database
  - **Change Files:** Real time updates are generated throughout the day within our span technology
- Future events file inclusive within the service
- Note: Additional delivery available for real-time consumption and XML files

S&P Global
Market Intelligence

Q&A

Daniel J. Sandberg, PhD, CFA
Director, Quantamental Research
S&P Global Market Intelligence

Frank Zhao
Director, Quantamental Research
S&P Global Market Intelligence

Michael Patton
Senior Director, Data Management Solutions
S&P Global Market Intelligence

S&P Global
Market Intelligence

S&P Global
Market Intelligence