

# Natural Language Processing in Finance: Shakespeare Without the Monkeys

For institutional investor, qualified investors, and investment professionals only. Not for retail public distribution.

Slavi Marinov, Co-Head of Machine Learning, Man AHL

July 2019

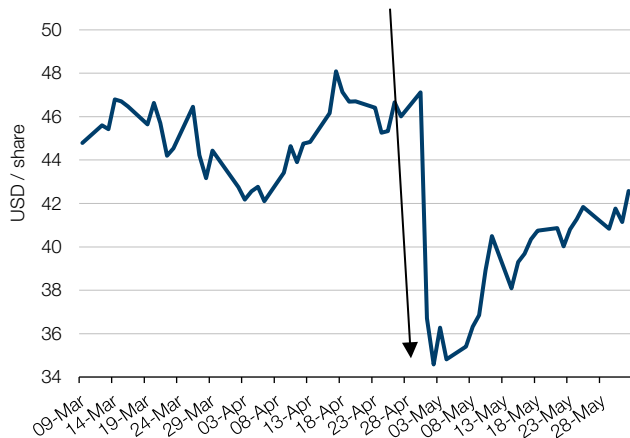
## 1. Introduction

"We want Facebook to be somewhere where you can start meaningful relationships," Mark Zuckerberg said on 1 May, 2018.

The announcement sparked gasps – not just from the crowd in front of whom Zuckerberg was talking – but also in financial markets. The share price of Match Group (the company that owns Match.com, Tinder and other dating websites) plunged by more than 20%.

**Figure 1. Facebook's Announcement Creates Ripples in Match.com**

Share price of Match.com plunged on 1 May, 2018, as Facebook announced that it would integrate a feature for online dating directly onto its app.



Source: Bloomberg; between 9 March, 2018 and 1 June, 2018.

Why is this example significant? The answer is simple: Financial markets were being swayed by a sentence made up of just a few words. There was not a single number in the announcement. More interestingly, Zuckerberg's comment did not impact Facebook's share price – the biggest effect was felt by a company that until that moment may not have even been considered a competitor to Facebook. The move was large, and almost instantaneous.

This behaviour – a few words causing strong reactions rippling through markets – happens all the time, albeit usually more subtly. The focus of this article is the automatic analysis of text by computers, also known as Natural Language Processing ('NLP'), which aims to extract meaning from words and predict the ripples even as they are happening.

## 2. What Is NLP?

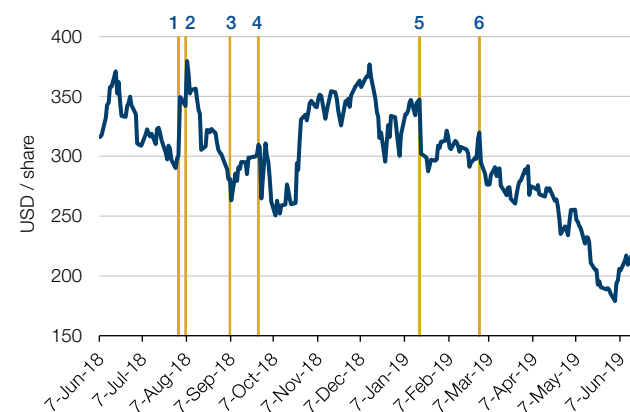
NLP is a sub-field of artificial intelligence ('AI'), which seeks to program computers to process, understand and analyse large amounts of human (or 'natural') language.

How is this useful in finance?

### 2.1. Detecting Material Events

As we saw in the Facebook example, it's useful in uncovering market-moving events. Facebook unveiled a new product – like Apple unveiling the iPhone – and that resulted in a very strong market move. Numerous such events happen in financial markets all the time. Indeed, for a lot of them, text, or even the spoken word, is the primary source. As such, methods from NLP can be used to automate this process: monitoring many text data streams and automatically issuing notifications upon the emergence of market-moving events.

**Figure 2. An Example of How Events Affected Tesla**



Source: Bloomberg; as of 21 June, 2019.

- 2 August, 2018: Shares soar as Tesla says production of its lower-cost Model 3 sedan is growing and CEO Elon Musk says the company does not expect to need to raise more money from investors.
- 7 August, 2018: Musk announces on Twitter that he wants to take Tesla private in a deal that would value the company at USD70 billion.
- 8 September, 2018: Just hours after Musk finishes smoking marijuana in a more than 2 1/2-hour podcast with comedian Joe Rogan, it is confirmed that both his chief accounting officer and head of human resources are leaving. Shares plunge.
- 27 September, 2018: Shares fall as the SEC accuses Musk of misleading investors with his 7 August tweet about taking Tesla private, raising questions about Musk's future.
- 18 January, 2019: Shares fall sharply as Musk warns that Tesla could struggle to make a profit in the first quarter and as he cuts more than 3,000 jobs from the electric carmaker.
- 1 March, 2019: Shares slide as Musk confirms that Tesla will not be profitable in the first quarter.

There are, however, many other ways in which machines can help.

2.2. Understanding Document Tone

Perhaps one of the most common applications of NLP in finance is measuring document tone, also known as sentiment. The idea is simple: get the machine to ‘read’ a document and assign it a score from -10 (very negative) to +10 (very positive).

Take the sentence below:

*French Cosmetics giant L’Oreal said strong demand for luxury skin creams helped it beat fourth-quarter sales forecasts - another company reporting better-than-feared demand from China after LVMH last week.*<sup>1</sup>

This would maybe get a score of 8.

Now take this sentence:

*Construction was a weak spot with Denmark’s Rockwool sinking 13% after full-year earnings missed expectations, and Sweden’s Skanska losing 7.8% after it cut its dividend and lagged profit estimates.*

This may get a score of -9 for Rockwool and Skanska.

While the two examples above are company-specific, sentiment analysis can also be done with respect to the economy in general, or even toward specific topics such as inflation or interest rates.

However, there are many challenges to correctly identifying sentiment, one being associating the tone with the correct entity. Consider this sentence:

*European stocks faltered on Friday after their worst day in six weeks as downgrades to growth forecasts weighed, while weak numbers from Umicore, Skanska, and Rockwool outweighed a sales beat from L’Oreal.*

A computer processing it cannot just assign a single sentiment score, as the sentence is negative for Umicore, Skanska, and Rockwool, but positive for L’Oreal.

2.3. Modelling Document Topics

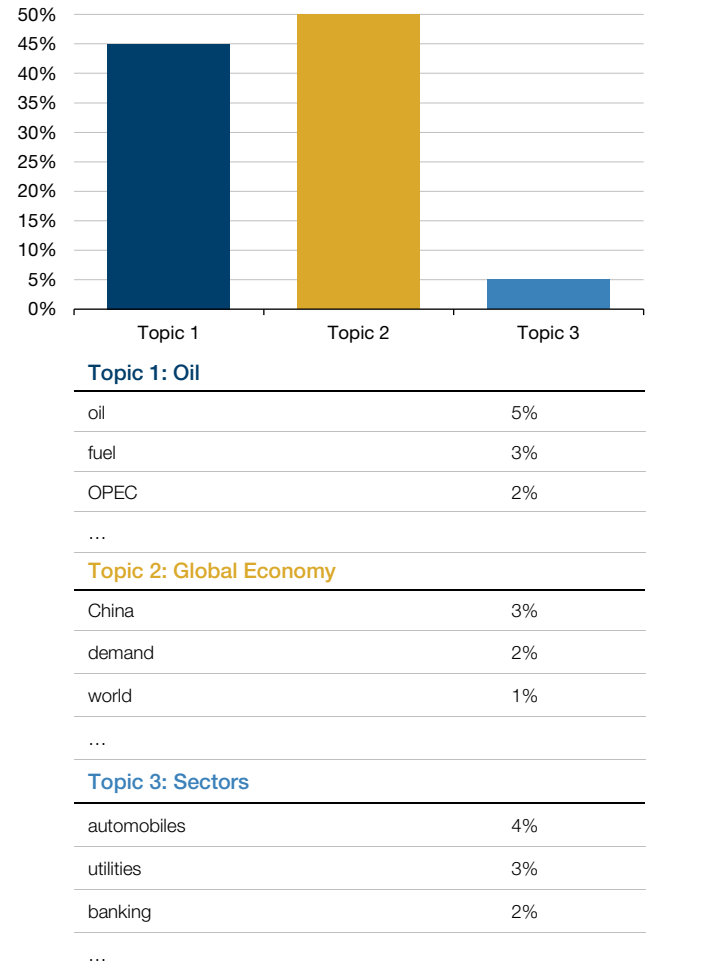
To be successful, NLP systems in finance often need to automatically extract a document’s topic structure. Consider this snippet from a news article:

*Oil prices fell on Monday after climbing to their highest this year earlier in the session as China reported automobile sales in January fell for a seventh month, raising concerns about fuel demand in the world’s second-largest oil user.*<sup>2</sup>

Often, the important information in a document is not just the tone, but its focus. In this example, there are two key topics: the first is oil, with words such as “oil”, “prices”, “fuel”, “fell” and “climb”; the second is the global economy, with words such as “world”, “China”, “demand” and “sales”. Understanding the topic structure of a document helps identifying events, informs the correct attribution of sentiment and allows to assess document similarity on a semantic level.

Figure 3. An Example of Machine Learning Models Inferring Topic Structure From a Document

Oil prices fell on Monday after climbing to their highest this year earlier in the session as China reported automobile sales in January fell for a seventh month, raising concerns about fuel demand in the world’s second-largest oil user



Source: Man Group; for illustration purposes only. The model has determined that the sentence is 45% about Topic 1, 50% about Topic 2, and 5% about other topics. We have explicitly labelled Topic 1 as Oil and Topic 2 as the Global Economy based on the most probable words associated with each of the topics.

The above example also highlights another subtle, but important, aspect of quantifying text data: timeliness. Even if we correctly identified the document’s topics, there are two timeframes mentioned: “fell on Monday” and “climbing to their highest this year earlier in the session”. Clearly, these two moves were attributed to a single entity – oil prices – yet they have opposite directions. Correctly identifying the evolution of events can be a crucial task for computer algorithms.

2.4. Detecting Subtle Change

This theme of subtlety is quite prevalent in NLP research. The information contained in text data is sometimes very obvious to the human eye (a new product launch in the news; lots of positive words by a company executive), but can just as often be buried. One example application of NLP is measuring textual change: comparing the same documents over time, and finding subtle differences.

For example, in IBM's 2016 annual report, the company had a snippet related to its brand risks under a risk factor called "Failure of Innovation Initiatives". In the following year's annual report, IBM decided to extract it as a separate risk factor called "Damage to IBM's Reputation", and explicitly listed eight broad categories of example sources of reputation risk.

Figure 4. Comparing IBM's Annual Reports

IBM's 2016 annual report <sup>3</sup>	IBM's 2017 annual report <sup>4</sup>
[...] IBM has one of the strongest brand names in the world, and its brand and overall reputation could be negatively impacted by many factors, including if the company does not continue to be recognized for its industry-leading technology and solutions and as a cognitive leader. If the company's brand image is tarnished by negative perceptions, its ability to attract and retain customers could be impacted.	<b>Damage to IBM's Reputation Could Impact the Company's Business:</b> IBM has one of the strongest brand names in the world, and its brand and overall reputation could be negatively impacted by many factors, including if the company does not continue to be recognized for its industry-leading technology and solutions and as a cognitive leader. <b>IBM's reputation is potentially susceptible to damage by events such as significant disputes with clients, product defects, internal control deficiencies, delivery failures, cybersecurity incidents, government investigations or legal proceedings or actions of current or former clients, directors, employees, competitors, vendors, alliance partners or joint venture partners.</b> If the company's brand image is tarnished by negative perceptions, its ability to attract and retain customers could be impacted.

Such subtle changes can be tricky and painstaking for a human to identify, especially given the typical length of annual reports and an investible universe of thousands of companies. Yet, to a machine, these changes are obvious: an algorithm can automatically scan through millions of documents and identify the added, deleted, or modified risk factors, classify them according to their topic, and even check which other companies have modified their risk factors in similar ways. Another example is the transcripts from the Federal Open Market Committee ('FOMC') on US interest rate policy, where the market typically reacts not to the current transcript, but rather to slight changes in wording between the current and previous ones.

### 2.5. Working Across Multiple Languages

All of the above examples are in English. While documents in English are convenient to consider because there is a vast amount of academic research in the area, it clearly isn't the case that all market-moving information originates in English. Consider the volume of important documents in any other language – be it Chinese or Russian, Japanese or Portuguese. To be able to leverage text from different languages and sources, one has to either develop models specific to that language (e.g. a Chinese sentiment model), or translate documents into English and then apply an English model. Indeed, both applications are currently a heavy focus of NLP.

### 2.6. Going Beyond Written Text

Finally, all examples so far assume that the text we are interested in already exists in written form. That is not always the case. For example, every quarter, many global public companies host earnings conference calls – the timeliest source of financial results.<sup>5</sup> Techniques from speech recognition research can be used to automatically transcribe

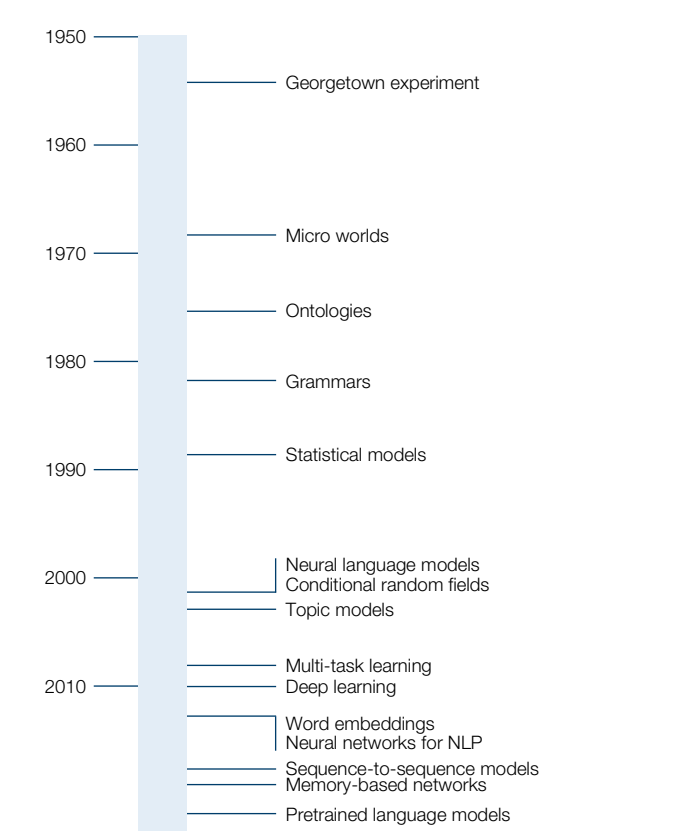
documents as the call is progressing, or even analyse the subtle nuances of management tone to measure emotions.<sup>6</sup>

All of the above – and many others – are central research topics within NLP. Indeed, various methods – many of which may provide potential sources of quant alpha – have been developed over the last seven decades. In addition, these methods can be used to quickly uncover emerging risk factors. The ball doesn't just stop there, however. The same ideas can be used to scan for investment ideas by discretionary managers or give early warnings of key developments to companies in their investment portfolio.

## 3. Why Should We Care About NLP Now?

From an academic perspective, NLP is a vast research field. And like any branch of AI, it can trace its modern roots back to the 1950s.

Figure 5. A Timeline of NLP



Source: IBM, Man Group.

For the first 30 years of their history, most NLP systems were based on large sets of carefully hand-crafted rules. Successful as these early programs were, they quickly became impossible to maintain and extend due to the huge amount of complexity.

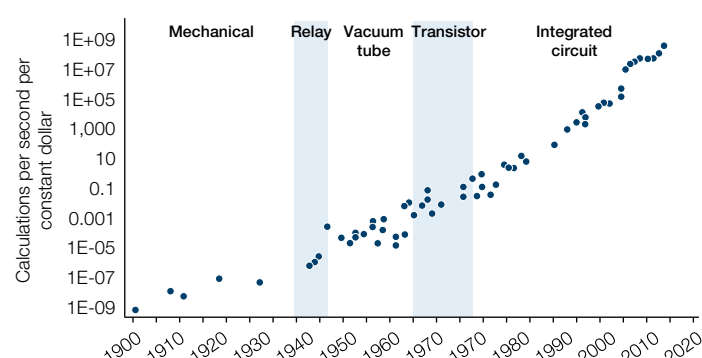
Starting in the 1980s, the field transitioned to statistical learning methods. Instead of explicitly hand-coding thousands and thousands of rules into the machine, what if the machine could automatically learn statistical regularities by observing large amounts of text? There would be no need to teach the machine the rules of grammar – it would automatically infer patterns by painstakingly going through bodies of text. Researchers would spend their time developing useful representations of text (also

known as features) that could be fed into the machine. This was the major idea behind second-generation NLP of the 30 years that followed, and resulted in a wealth of exciting innovations.

However, there was still one problem. While researchers were no longer hand-crafting rules, they were now hand-crafting features for their statistical machine learning models to work. This was better, but still not ideal. The transition from hand-crafted rules to statistical learning seemed transformational; could another paradigm shift be seen by going from hand-crafted features to learning them directly from the data?

In the last 10 years, we witnessed the third major wave of scientific breakthroughs. These innovations come from the field of neural networks – also known as deep learning. Many of the basic ideas were not new, dating back to the 1950s, though they had largely gone out of favour. What was new was the vast amounts of computing power that was available, and a fresh look at making these powerful methods practical.

**Figure 6. 120 Years of Moore's Law**



Moore's Law is the observation made in 1965 by Gordon Moore, co-founder of Intel, that the number of transistors per square inch on integrated circuits had doubled every year since the integrated circuit was invented. Moore predicted that this trend would continue for the foreseeable future. More transistors means faster chips, with the performance gains compounding over time. As a result of this exponential growth in technological capabilities, many computationally demanding methods – such as deep learning – have become practical.

These neural network models get their inspiration from the human brain. Building blocks, called artificial neurons, are connected together, in code, to form larger networks. These neurons take some raw input data, fire up and transfer their impulses forward, ultimately resulting in a prediction. A researcher can define the 'shape' of the network: the connectivity pattern between the neurons. By designing different layouts and stacking them on top of one another (hence the name, 'deep' learning), researchers can impose their prior knowledge of the world. Given sufficiently large and complex datasets and compute resources, the strength of the connections between the artificial neurons can be learned. The researchers can create the blueprint (called the 'architecture'), supply the data and guide the learning process; the neural networks adjust the neuron connection strengths to make the most accurate predictions.

Neural network models have successfully modelled problems ranging from how to represent the meaning of words in a computer (word embeddings<sup>7,8,9</sup>), through capturing the meaning of chunks of words (convolutional neural networks<sup>10,11,12</sup>), to modelling the sequential (recurrent neural networks<sup>13,14,15</sup>), and compositional (recursive neural

networks<sup>16</sup>) nature of phrases. Indeed, these ideas have been the foundation of many of the recent state-of-the-art results in modern NLP.

As one example (of many), in September 2016, Google transitioned into using neural models, announcing its Google Neural Machine Translation ('GNMT') system.<sup>17</sup> The system "utilizes state-of-the-art training techniques to achieve the largest improvements to date for machine translation quality: GNMT reduces translation errors by more than 55%-85% on several major language pairs measured on sampled sentences from Wikipedia and news websites with the help of bilingual human raters".

## 4. Challenges When Using NLP

Working with language can inherently be more difficult than working with well-structured numerical data.

### 4.1. Scale and Technology

The first obvious challenge is scale. Unlike many numerical datasets, text data can be very large and thus requires significant investments in data storage and computation capacities. To be effective, large-scale distributed computation resources (hardware) are typically required, along with enough storage for all raw and intermediate data (with data stored efficiently) so that ideas can be iterated quickly. In addition, specialist software may need to be developed, to help visualise the complexities of the NLP research stages and aid research.

At Man AHL, we have made significant investments in all of the above: from advanced hardware storage<sup>18</sup> and parallel processing systems, to internally developing and then open-sourcing our fast time series store<sup>19,20</sup>, and NLP visualisation frameworks.<sup>21</sup>

### 4.2. Grammar, Semantics and Context

The next challenge is that 'natural' language often doesn't do a particularly good job of conforming to cleanly defined grammatical rules. Some datasets you may want to look at in finance – such as annual reports or press releases – are carefully written and reviewed, and are largely grammatically correct. They are thus relatively easy for a computer to analyse. But how about tweets, product reviews or online forum comments? These tend to be full of abbreviations, slang, incomplete sentences, emoticons, etc – all of which make it quite tricky for a machine to decipher. On top of this, many of the documents of interest to finance come in fairly messy formats such as PDF or HTML, requiring careful processing before you can even get to the information of interest.

Understanding semantics – what the document is about – is even more challenging. Human language is full of ambiguity. Issues such as synonymy, polysemy (the coexistence of many possible meanings for a word or phrase), presupposition (the tacit assumption at the beginning of a sentence i.e. "we have settled the lawsuit" pre-supposes a lawsuit has taken place), sarcasm and other linguistic quirks make NLP a highly challenging discipline.

The difficulty is compounded by the existence of domain-specific language within different contexts. For instance, "fine" is a very positive answer to the question: "How are you feeling



today?”, but may not be quite as positive when it relates to the outcome of a legal proceeding against a company.

Context is perhaps the most challenging aspect to get right. As humans, we have vast amounts of context and common sense accumulated over years of experience. Even within the same document, we need to specifically set up machines so that they carry over and ‘remember’ concepts across sentences. It gets much more difficult when the context is not even present in the body of documents a machine is processing.

### 4.3. The Talent War

Perhaps the ultimate challenge is talent. To make sense of text data, experts from the fields of linguistics, machine learning and computer science need to be hired. In today’s highly competitive market, one needs to compete in the talent war for the best and brightest.

## 5. Case Study: Sentiment Analysis

As an illustration of how we may think about applying NLP to finance, consider the following intuitive alpha idea: when people talk about companies using more positive language, these companies may be more likely to generate higher stock-market returns than companies with more negative language. People in this setting will be kept deliberately generic: these may be journalists, company executives, analysts, individual investors, clients, or political figures.

To test this idea, we would first need to collect a large dataset of documents associated with our tradable universe of stocks. These may be news, annual reports, earnings calls, blog posts, product reviews, social media posts, etc. We would then need to analyse all text documents related to each company and rank all stocks based on how positive its associated documents are. Finally, we are going to test going long the companies with the most positive language, and short the companies with the most negative language, continually rebalancing our portfolio as new documents appear.

The core task of such a trading system is measuring document tone: we need a mechanism to convert each document into a number. The more positive the document, we would expect a higher number assigned by our system. As mentioned above, this is a central problem of NLP and is known as sentiment analysis.

How do we design such a system? We will go through a series of approaches, each one building upon the previous, to illustrate one potential path of the core ideas.

### 5.1. A Naïve Sentiment Model

To begin with, let’s look at the following two news headlines:

*Disney CEO is upbeat about deal with Fox.*<sup>22</sup>

*Germany's tightening labor market might spell more trade trouble.*

As humans, why do we think the first one is positive, and the second one is negative? At a basic level, it is because of words such as “upbeat” and “trouble”. Throughout our lives, education and work, we have assembled very large internal

databases of word meaning, and when we see a sentence we apply these instantaneously and automatically.

Our first sentiment model is then simply going to replicate this human intuition at scale: count positive and negative words. What if we had a long list of all the positive and negative words that could appear in our documents? We could then simply get a computer to count these words and compute an aggregate score:

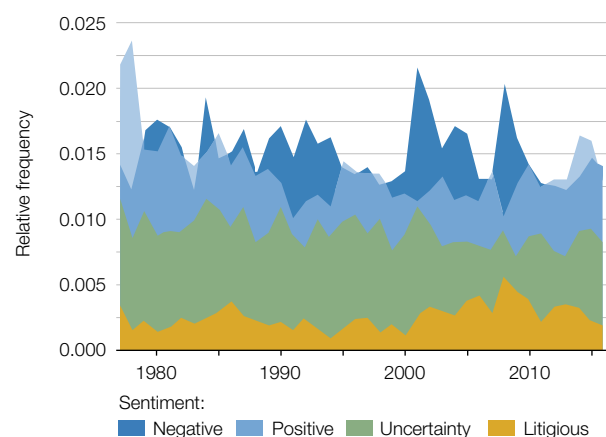
$$\text{sentiment} = \frac{\text{num}_{\text{positive}} - \text{num}_{\text{negative}}}{\text{num}_{\text{total}}}$$

That is, we simply subtracted the number of negative words from the number of positive words, and normalised this score by the total number of words in a document.

In fact, this is exactly the approach taken by one of the first highly influential academic papers in the fields of text analysis in finance. In 2007, Professor Paul Tetlock showed that counting negative words in a particular column of the Wall Street Journal had predictive power over the future price moves of the Dow Jones Industrial Average Index and the daily volume traded on the New York Stock Exchange.<sup>23</sup>

### 5.2. Finance-Specific Dictionaries

**Figure 7. Sentiment Analysis of Warren Buffett's Shareholder Letters (Using the Loughran-McDonald Lexicon)**



Source: <https://juliasilge.com/blog/tidytext-0-1-3/>; as of 18 June, 2017.

The core challenge of any word-counting method is coming up with the ‘right’ long lists of words to count. The more thorough and accurate the word lists are, the higher is the quality of our sentiment measure, and thus the more profitable our trading strategy.

Tetlock’s paper used general-purpose word lists – General Inquirer’s Harvard IV-4 psychosocial dictionary. The dictionary is not specific to finance and contains many words that – while positive or negative in the field of psychology – have little sentiment meaning in the domain of finance. This led, in fact, to a highly influential 2011 paper by Professors Tim Loughran and Bill McDonald (‘LM’)<sup>24</sup> showed that almost 75% of the negative words in the Harvard GI dictionaries were irrelevant in finance!<sup>25</sup>

These professors and their students then set off on a mission to build a finance-specific dictionary, one that would fit the bill of being comprehensive, domain-specific and accurate. What

they published in 2011 quickly became the de-facto standard in academic finance. Their word lists, containing about 2,300 negative words and more than 350 positive words, lead to the publication of academic papers applying them to financial datasets ranging from news through press releases to blogs, earnings calls, and even mutual fund letters to shareholders.

This core observation – that domain adaptation is important – is the centre of the **second** sentiment model, where, instead of counting words from a general-purpose dictionary, finance-specific words from the LM dictionaries are counted. We know these are comprehensive, accurate, and widely accepted in academic finance as state of the art: a solid foundation for the word counting strategy!

Are we done? Not quite.

### 5.3. A Problem With Counting Words

To see why, let's look at the following sentence:

*Our Board of Directors has approved a new \$100 billion share repurchase authorization as well as a 16% increase in our quarterly dividend.*<sup>26</sup>

If you run this – and many other similar sentences – through a computer program that counts words from the LM dictionary, you will get a sentiment score of exactly zero. This is unfortunate, because we, as humans, see sentences like these as highly positive. Is that an isolated example or a more fundamental flaw? Is it simply the case the LM word lists have a few missing words we can add and move on?

If we look at why we as humans think the above sentence is positive, we quickly notice that none of the words above are positive or negative in isolation. A “dividend” or an “increase” are on their own neutral<sup>27</sup>; it is the combination “increase in dividend” that makes us think the sentence is positive. Consequentially, even if we had the most comprehensive and accurate word lists in our domain, any word-counting method will be unable to capture such sentences.

To solve this problem, we leave the world of academic finance and enter the world of machine learning, where sentiment classification has been studied for a long time, and where there have been some exciting recent advancements.<sup>28</sup>

### 5.3. Machine Learning and Targeted Sentiment

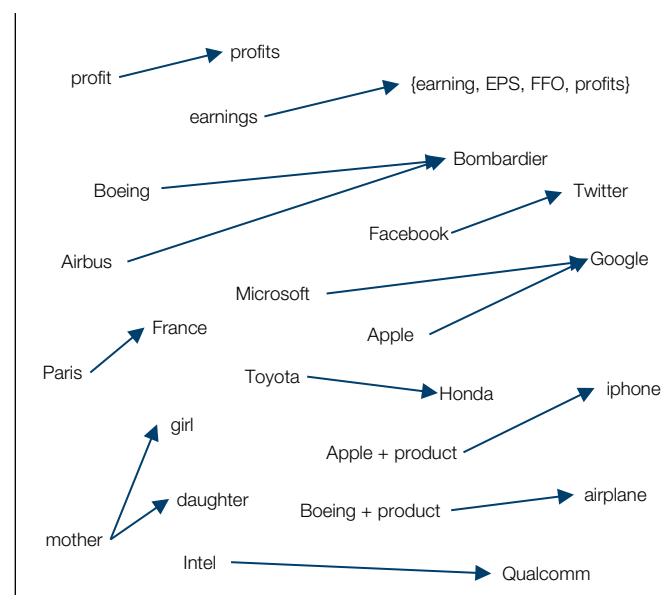
To tackle sentences like the one above, ‘targeted sentiment’ methods are used, i.e. given a target (like “dividend”), models are built that will tell us whether a sentence is positive or negative with respect to this target. In the example above, our model should tell us that the sentence is positive with respect to both “repurchase” and “dividend”.

To do this, three problems have to be solved: how to compile the list of interesting targets, how to find all words in a sentence related to a target, and how to come up with the words that are positive or negative given a target. It turns out that recent advances in machine learning are very helpful in solving these problems.

The first and last tasks – coming up with lists of targets of interest, and positive/negative word lists for each target – look remarkably similar to what Loughran and McDonald did in their

2011 work. In their case, their research group manually and painstakingly went through tens of thousands of words, reviewing each one manually and deciding whether each word was positive, negative or neutral. While that can be done, it would take vast amounts of time. Instead, a recent technique in machine learning called word embeddings can be used to automatically generate similar words given a set of seed words.

Figure 8. Word Vectors



Source: Man Group; for illustrative purposes only.

One of the most popular word embedding models is called word2vec because it can learn to convert any given word to a sequence of numbers. To do that, word2vec – a simple neural network model published in 2013 by a research group at Google – plays a game with itself. It goes through a body of text (e.g. all financial news) looking at all possible sequences of, say, seven words. It then ‘hides’ the middle word from itself, and tries to guess what it is based on its context (or vice versa – tries to predict the context from the middle word). By doing this painstakingly hundreds of millions of times (computers are good at this: they don’t get bored easily), it ends up implicitly learning word similarity. This is because words that appear in similar contexts end up represented as similar vectors of numbers. For example, “EBIT” is the most similar to “earnings” and “EBITDA” as all of these get used in fairly similar contexts.

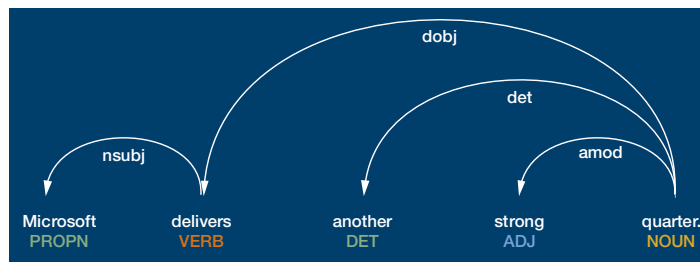
The best part? These similarities are learned completely independently, in a dataset-specific way, without the need of any human supervision.

How can word2vec help? We can simply provide a set of seed target words (e.g. “EBIT”) – and then query the word embedding models for all words that are similar to our seeds (e.g. “EBITDA”, “earnings”). We can then greatly expand our list of seed targets with the ones suggested by word2vec. Likewise, we can start with a set of sentiment-bearing seed words (e.g. “increase”) and use word embeddings to expand them (e.g. to “improve”, “up”, “skyrocket”).

There is, however, one final problem to be solved – given a target, how to find all words in a sentence that are related to this target? In NLP, this problem is known as dependency

parsing, and again, the state-of-the-art models are neural-network based.

Figure 9. Dependency Parsing



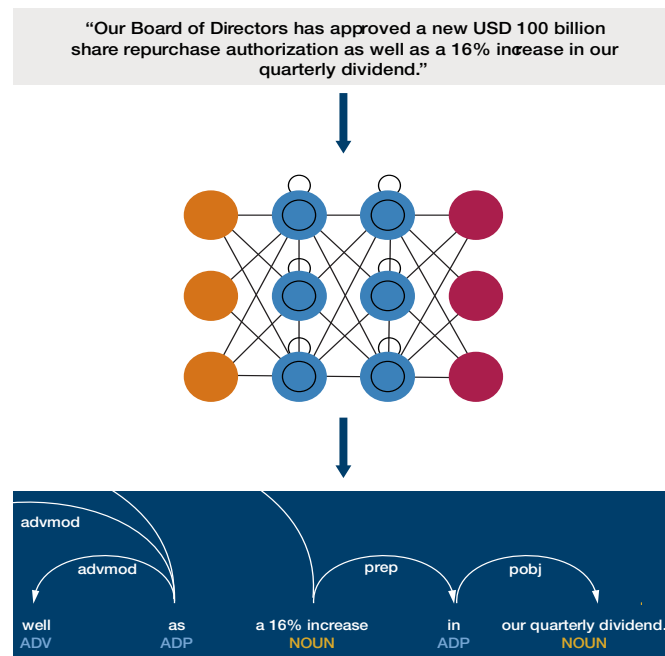
Quote source: M.W. De Oliveira (2018/05/01). Microsoft Does It Again (SeekingAlpha). Image generated with spacy: <https://spacy.io/>

Given a sentence, a dependency parser would automatically identify the relationships between the words.

In Figure 9, the dependency structure is illustrated by the arrows. They indicate that “quarter” is the direct object of the verb “delivers”, and that “Microsoft” is its subject. They tell us that “strong” is an adjective modifying the noun “quarter”. The colourful words in uppercase are known as part of speech tags; these speech tags show that “delivers” is a verb, “strong” is an adjective and “quarter” is a noun.

Note that the annotations in the above figure were not generated by a human – they were generated by a neural network. These models are nowadays trained on huge amounts of data and are surprisingly accurate.

Figure 10. Dependency Parsing and POS Tagging



Quote source: Apple second-quarter results - earnings call transcript. Image sources: Man Group (visualisation using <https://spacy.io/>) and <http://www.asimovinstitute.org/neural-network-zoo/>

All we need to do now to create our third sentiment model, one that can correctly capture targeted sentiment, is to put all these bits together.

We start with a set of seed targets (“EBITDA”, “repurchase”, “dividend”) and use word embeddings to generate expanded

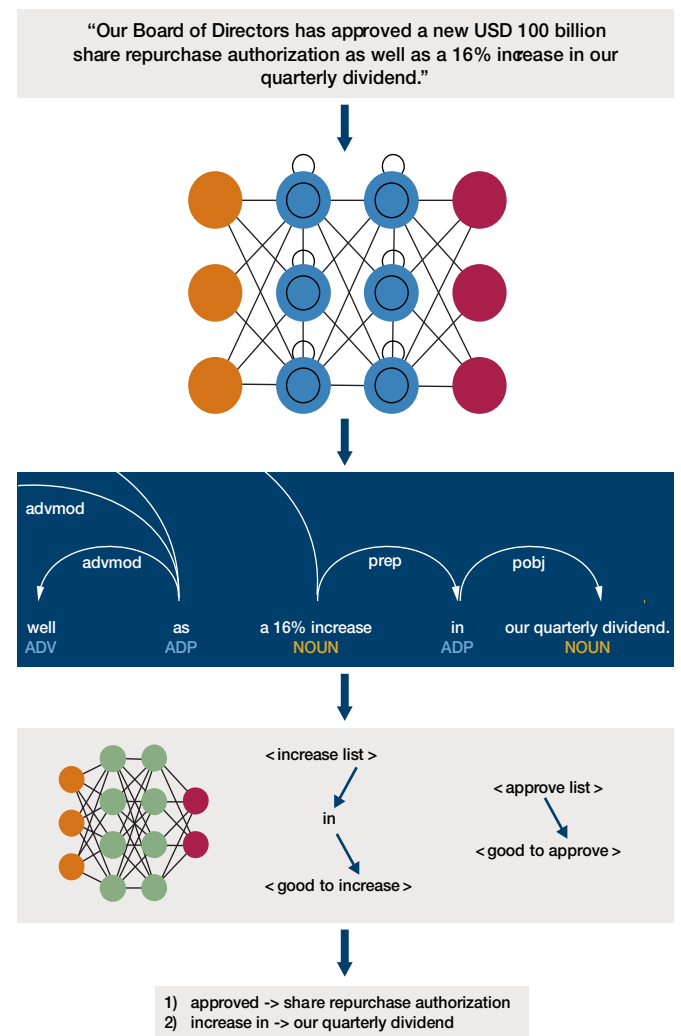
lists of targets of interest. We then scan each sentence and check if any of the targets of interest is in it. If so, we use a neural network to identify the dependency structure of the sentence and find all words related to our target. In this neighbourhood, we count the target-dependent positive or negative words (again, constructed by taking a set of seed sentiment words and expand them using our word embeddings).

Our final sentiment score is going to be computed as before:

$$Sentiment = \frac{num_{positive} - num_{negative}}{num_{total}}$$

However, there is one minor modification: instead of simply counting positive and negative words, we count positive and negative targeted phrases.

Figure 11. Targeted Phrases



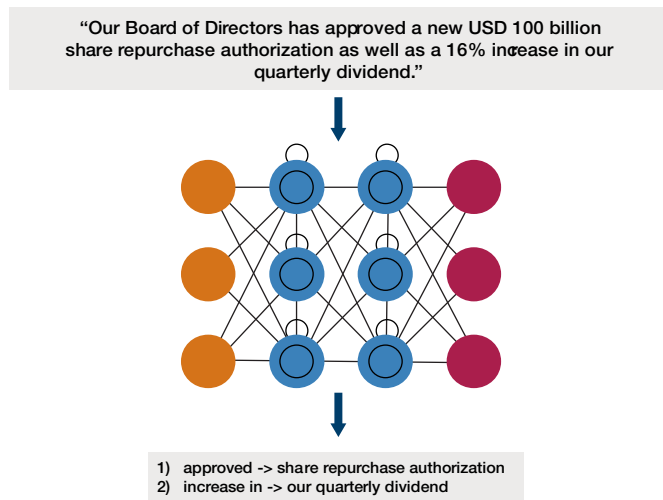
Quote source: Apple second-quarter results - earnings call transcript. Image sources: Man Group (visualisation using <https://spacy.io/>) and <http://www.asimovinstitute.org/neural-network-zoo/>

## 5.4. End-to-end sentiment models

While we could choose to stop here, we will mention one final extension. Note that all the work so far requires various independent steps, some of them fairly manual: pick seed targets; pick seed sentiment words; train a word2vec model; expand the targets. Most notably, the method assumes

sentences are fairly well-formed so that our dependency parser model can do a good job. What if we could combine all of these sub-models into one full end-to-end model?

**Figure 12. End-to-End Models**



Quote source: Apple second-quarter results - earnings call transcript. Image source: and <http://www.asimovinstitute.org/neural-network-zoo/>

Nowadays, end-to-end neural network-based models have been developed to start with raw sentences and directly learn to classify them into positive and negative. These methods do not rely on any intermediate steps and instead leverage large labelled datasets and learn intermediate representations and sentiment scores directly. These models are particularly useful

in areas such as social media analysis, where dependency parsing is tricky. An end-to-end neural network is the fourth and (perhaps) final iteration of our sentiment model.

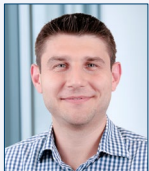
We stress that there are many other ways in which sentiment models can be improved, but hopefully this provides a basic illustration of the evolution of the field and how recent advancements in NLP can be used to construct ever-more-accurate models.

We can then use the results from our sentiment model to add sentiment signals to our quant portfolios, amend our discretionary stock selection process, or identify emerging risk factors.

## 6. Conclusion

Regardless of the methods used, we believe NLP is an extremely exciting research area in finance due to the vast range of problems it can tackle for both quant and discretionary fund managers. In particular, firms with strong investments in technology infrastructure and machine learning talent have positioned themselves to potentially capitalise on successfully applying these methods to finance.

On top of that, more recently, neural models have opened up a new frontier in NLP: one where machines can learn end-to-end models and improve state of the art results on a wide range of problems. Combined with the availability of more data than ever, vast amounts of available compute and improved tools<sup>29,30,31,32</sup>, these exciting recent research advances may create a rich and fruitful alpha opportunity.



**Slavi Marinov**

Co-Head of Machine Learning at Man AHL

Slavi Marinov is Co-Head of Machine Learning at Man AHL, responsible for everything from hardware to research frameworks to live trading strategies. Prior to joining Man AHL in 2014, he was Co-founder and Vice President of Engineering for a natural language processing fintech start-up. Before that, he performed various senior software engineering and product management roles. Slavi holds a BSc in Computer Science and a Silver Medal from the Balkan Olympiad in Informatics.



## Footnotes

1. <https://uk.reuters.com/article/europe-stocks/european-stocks-falter-as-investors-digest-weak-earnings-loreal-impresses-idUKL5N2031VI>
2. Source: Reuters.
3. <https://www.sec.gov/Archives/edgar/data/51143/000104746917001061/a2230222z10-k.htm>
4. <https://www.sec.gov/Archives/edgar/data/51143/000104746918001117/a2233835z10-k.htm>
5. <https://seekingalpha.com/article/4241565-deere-and-company-de-q1-2019-results-earnings-call-transcript>
6. William Mayhew and Mohan Venkatachalam (2012), The Power of Voice: Managerial Affective States and Future Firm Performance, Journal of Finance.
7. Tomas Mikolov, Kai Chen, Greg Corrado, Jeffrey Dean (2013); Efficient Estimation of Word Representations in Vector Space.
8. Jeffrey Pennington, Richard Socher, Christopher D. Manning, GloVe: Global Vectors for Word Representation.
9. Armand Joulin, Edouard Grave, Piotr Bojanowski, Tomas Mikolov (2016); Bag of Tricks for Efficient Text Classification.
10. Ronan Collobert, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, Pavel Kuksa (2011); Natural Language Processing (Almost) from Scratch.
11. Nal Kalchbrenner, Edward Grefenstette, Phil Blunsom (2014); A Convolutional Neural Network for Modelling Sentences.
12. Yoon Kim (2014); Convolutional Neural Networks for Sentence Classification.
13. Tomáš Mikolov, Martin Karafiát, Lukáš Burget, Jan "Honza" Cernocký, Sanjeev Khudanpur (2010); Recurrent neural network based language model.
14. Shujie Liu , Nan Yang , Mu Li and Ming Zhou (2014); A Recursive Recurrent Neural Network for Statistical Machine Translation.
15. Tony Robinson, Mike Hochberg and Steve Renals (1996); The Use of Recurrent Neural Networks in Continuous Speech Recognition.
16. Richard Socher, Alex Perelygin, Jean Y. Wu, Jason Chuang, Christopher D. Manning, Andrew Y. Ng and Christopher Potts; Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank.
17. Quoc V. Le and Mike Schuster (2016); A Neural Network for Machine Translation, at Production Scale.
18. [https://www.purestorage.com/content/dam/purestorage/pdf/Case%20Studies/Pure\\_Storage\\_Case\\_Study\\_Man\\_AHL.pdf](https://www.purestorage.com/content/dam/purestorage/pdf/Case%20Studies/Pure_Storage_Case_Study_Man_AHL.pdf)
19. <https://github.com/manahl/arctic>
20. <https://www.mongodb.com/press/man-ahl-arctic-open-source>
21. <https://github.com/manahl/pynorama>
22. Source: WSJ <https://www.wsj.com/articles/strong-box-office-propels-disney-to-profit-growth-1525813208>  
Source: Bloomberg <https://www.bloomberg.com/news/articles/2018-05-09/germany-s-tightening-labor-market-might-spell-more-trade-trouble>
23. Paul Tetlock (2007); Giving Content to Investor Sentiment: The Role of Media in the Stock Market.
24. Tim Loughran and Bill McDonald (2009); When is a Liability not a Liability? Textual Analysis, Dictionaries, and 10-Ks.
25. These seven words – tax, costs, loss, capital, cost, expense and expenses – account for more than one-fourth of the total count of 'negative' words. Yet in the financial world, firm costs, sources of capital or the amount of tax paid are neutral in nature; managers using this language are merely describing their operations. In some non-business situations, 'foreign' or 'vice' might appear as negative words. In 10-K text, however, it is far more likely that 'foreign' is used in the context of international operations or 'vice' is used to refer to vice-presidents of the firm.
26. Source: Apple (AAPL) Q2 2018 Results - Earnings Call Transcript.
27. Just think of a "decrease in dividend" and "an increase in costs": putting dividend and increase in different contexts make us flip our assessment of the polarity of the sentence.
28. We should point out that a simple word-counting approach has many more gaps: we will try to focus on just this one and try to show how using modern advancements in machine learning can help solve this problem.
29. <https://www.tensorflow.org/>
30. <https://spacy.io/>
31. <https://spark.apache.org/>
32. <https://github.com/manahl/pynorama>

## Important Information

This information is communicated and/or distributed by the relevant AHL or Man entity identified below (collectively the "Company") subject to the following conditions and restriction in their respective jurisdictions.

Opinions expressed are those of the author and may not be shared by all personnel of Man Group plc ('Man'). These opinions are subject to change without notice, are for information purposes only and do not constitute an offer or invitation to make an investment in any financial instrument or in any product to which the Company and/or its affiliates provides investment advisory or any other financial services. Any organisations, financial instrument or products described in this material are mentioned for reference purposes only which should not be considered a recommendation for their purchase or sale. Neither the Company nor the authors shall be liable to any person for any action taken on the basis of the information provided. Some statements contained in this material concerning goals, strategies, outlook or other non-historical matters may be forward-looking statements and are based on current indicators and expectations. These forward-looking statements speak only as of the date on which they are made, and the Company undertakes no obligation to update or revise any forward-looking statements. These forward-looking statements are subject to risks and uncertainties that may cause actual results to differ materially from those contained in the statements. The Company and/or its affiliates may or may not have a position in any financial instrument mentioned and may or may not be actively trading in any such securities. This material is proprietary information of the Company and its affiliates and may not be reproduced or otherwise disseminated in whole or in part without prior written consent from the Company. The Company believes the content to be accurate. However accuracy is not warranted or guaranteed. The Company does not assume any liability in the case of incorrectly reported or incomplete information. Unless stated otherwise all information is provided by the Company. Past performance is not indicative of future results.

Unless stated otherwise this information is communicated by AHL Partners LLP which is registered in England and Wales at Riverbank House, 2 Swan Lane, London, EC4R 3AD. Authorised and regulated in the UK by the Financial Conduct Authority.

**Australia:** To the extent this material is distributed in Australia it is communicated by Man Investments Australia Limited ABN 47 002 747 480 AFSL 240581, which is regulated by the Australian Securities & Investments Commission (ASIC). This information has been prepared without taking into account anyone's objectives, financial situation or needs.

**Austria/Germany/Liechtenstein:** To the extent this material is distributed in Austria, Germany and/or Liechtenstein it is communicated by Man (Europe) AG, which is authorised and regulated by the Liechtenstein Financial Market Authority (FMA). Man (Europe) AG is registered in the Principality of Liechtenstein no. FL-0002.420.371-2. Man (Europe) AG is an associated participant in the investor compensation scheme, which is operated by the Deposit Guarantee and Investor Compensation Foundation PCC (FL-0002.039.614-1) and corresponds with EU law. Further information is available on the Foundation's website under [www.eas-liechtenstein.li](http://www.eas-liechtenstein.li). This material is of a promotional nature

**European Economic Area:** Unless indicated otherwise this material is communicated in the European Economic Area by Man Asset Management (Ireland) Limited ('MAMIL') which is registered in Ireland under company number 250493 and has its registered office at 70 Sir John Rogerson's Quay, Grand Canal Dock, Dublin 2, Ireland. MAMIL is authorised and regulated by the Central Bank of Ireland under number C22513.

**Hong Kong:** To the extent this material is distributed in Hong Kong, this material is communicated by Man Investments (Hong Kong) Limited and has not been reviewed by the Securities and Futures Commission in Hong Kong. This material can only be communicated to intermediaries, and professional clients who are within one of the professional investor exemptions contained in the Securities and Futures Ordinance and must not be relied upon by any other person(s).

**Liechtenstein:** To the extent the material is used in Liechtenstein, the communicating entity is Man (Europe) AG, which is regulated by the Financial Market Authority Liechtenstein (FMA). Man (Europe) AG is registered in the Principality of Liechtenstein no. FL-0002.420.371-2. Man (Europe) AG is an associated participant in the investor compensation scheme, which is operated by the Deposit Guarantee and Investor Compensation Foundation PCC (FL-0002.039.614-1) and corresponds with EU law. Further information is available on the Foundation's website under [www.eas-liechtenstein.li](http://www.eas-liechtenstein.li).

**Switzerland:** To the extent this material is distributed in Switzerland, this material is communicated by Man Investments AG, which is regulated by the Swiss Financial Market Authority FINMA.

**United Kingdom:** Unless indicated otherwise this material is communicated in the United Kingdom by Man Solutions Limited ('MSL') which is an investment company as defined in section 833 of the Companies Act 2006. MSL is registered in England and Wales under number 3385362 and has its registered office at Riverbank House, 2 Swan Lane, London, EC4R 3AD, United Kingdom. MSL is authorised and regulated by the UK Financial Conduct Authority (the 'FCA') under number 185637.

**United States:** This material was prepared by AHL Partners LLP ("AHL Partners") and is presented by Man Investments Inc. ("Man Investments"). Man Investments is registered as a broker-dealer with the U.S. Securities and Exchange Commission ("SEC") and is a member of the Financial Industry Regulatory Authority ("FINRA"). Man Investments is also a member of Securities Investor Protection Corporation ("SIPC"). Man Investments is a wholly owned subsidiary of Man Group plc. 'Man Group' refers to the group of entities affiliated with Man Group plc. AHL Partners is registered as an investment advisor with the SEC. AHL Partners is also registered as a commodity pool operator and a commodity trading adviser with the National Futures Association ("NFA") as authorized by the US Commodity Futures Trading Commission ("CFTC"). The registrations and memberships in no way imply a certain level of skill or expertise or that the SEC, FINRA, SIPC, CFTC or NFA have endorsed the entities mentioned above.

This material is proprietary information and may not be reproduced or otherwise disseminated in whole or in part without prior written consent. Any data services and information available from public sources used in the creation of this material are believed to be reliable. However accuracy is not warranted or guaranteed. © Man 2019

2019/1665088/S/US/GL/VA