

# **An exploratory study using Knowledge Organization Systems to automatically detect forward-looking sentiment in company reports to infer social phenomena and predict business performance.**

**Paul H Cleverley (Robert Gordon University) and Laura Muir (Edinburgh Napier University)**

## **ABSTRACT**

Knowledge Organization Systems (KOS) are a 'looking glass' by which we can render reality. Several KOS use modal verbs, adjectives and rules to indicate forward-looking assertiveness sentiment (opinions) within texts. However, no prior work has sought to assess the strengths and weaknesses of these KOS. Many existing KOS do not account for 'intensity' of forward-looking sentiment. This exploratory study aims to assess how a composite KOS could be an improvement on existing individual KOS when applied to company annual reports, in order to assess the extent to which cultural differences and rhetoric can be identified and future performance predicted.

A range of existing KOS were critiqued leading to the creation of a composite KOS representing the 'extreme edges' of strong assertive and hesitant forward-looking opinions. The composite KOS was applied to the genre of company annual reports for four large multinational Oil and Gas (O&G) companies between 2008 and 2015. Word frequency and biologically inspired diversity ratios for the categories were generated and examined.

A strong association was found between a decline in the use of assertive language for some companies over time, potentially indicating increasing business uncertainty. Sharp increases in mentions of 'future' and 'learnings' appeared to be linked to industrial disasters and crisis management in two of the companies studied. This may evidence aspects of organizational rhetoric and renewal. There was evidence to support the '*Pollyanna effect*' in one company: the social phenomena of over-positive business language. In the same company, a moderate association was found between increasing diversity of assertive language and declining financial performance the following year. This marker has not been reported before and provides an area for further research.

It is possible that while some companies render changing attitudes towards the future state of affairs through word usage in company reports, others may not and some may even deploy rhetoric. This may point to a more complex model than 'universal laws' when generalizing and interpreting organizational word usage in company reports.

Sentiment is typically applied with an *a priori* hypothesis in mind. Embedding these sentiment algorithms in standard enterprise search and discovery technology deployments may help generate new insights and knowledge in the most unexpected of places.

*Keywords: Knowledge management, sentiment analysis, enterprise search & discovery, text analytics, oil and gas, annual company reports, insight engines*

## **1. INTRODUCTION**

Much of the Library and Information Science (LIS) literature has been concerned with indexing and analysis of explicit document objects for their key characteristics so they can be stored in an information system for retrieval (Shiri, Revie and Chowdhury 2002, Tennis 2009, Zeng et al 2007).

Text and Data Mining (TDM) is the use of automated analytical techniques to analyse text and data for patterns, trends and other useful information. Cognitive Computing combines this with Natural Language Processing (meaning) and Machine Learning (prediction) to 'mimic' human thought processes to augment decision making. Some enterprise search and discovery implementations include the identification of insights hidden within texts for exploitation, which may include trends relating to sentiment (Kruschwitz and Hull 2017) making them examples of Cognitive Computing applications. However, sentiment engines are likely to need customizations in their underlying Knowledge Organization Systems (Van Boeyen 2014). For example, using an off-the-shelf commercial sentiment analysis tool, it was reported that the American Red Cross found that only 21% of positive comments and 53% of negative comments were successfully detected by the software (Grimes 2012).

Increasing volumes of information are leading to large aggregate 'information objects' that can be 'smashed apart' (Smiraglia and Van Den Heuvel 2011) where their sum is greater than the parts, leading to emergent properties (Aaltonen and Tempini 2014). Michel et al (2011) termed the study of word frequency patterns using computerized techniques on big data (4% of the world's books), 'culturomics'. This encompasses a wide range of disciplines including inferences of human thought and social phenomena events and trends through time. Using a geological analogy, these word frequency patterns may provide a time machine to examine 'trace fossils' of our social history, which may also provide a key to the present and future.

Machine generated computational linguistics techniques convert words into numbers to which mathematical operators can be applied. By linking or correlating these data with an external variable such as revenue, the potential exists to enhance business intelligence and potentially generate predictive models. A key decision to be made is what words to use and how they can be grouped into categories (Knowledge Organization Systems (KOS)) through common characteristics. Content analysis can "*stand and fall on its categories*" (Berelson 1952, pg. 92).

Cambria and Hussain (2015) suggest a shift has been taking place in recent years in sentiment analysis, from syntax to semantics and from statistics to linguistics. For example the phrase "*The iPhone6 is expensive but nice*" has the opposite polarity to the phrase "*The iPhone6 is nice but expensive*" in terms of consumer positive and negative sentiment, evidencing the significance of word order effects.

Studies have shown that various verbs (such as *shall*, *will*, *might*, *could*) and adjectives may provide an indication of assertiveness 'sentiment', a forward-looking opinion or hedge. An assumption made is that the use and distribution of particular verbs and adjectives in any manifesto is not random, but deliberate. Parts of speech such as modal verbs may indicate how definite or confident a company feels about a proposition; their attitude towards a state of affairs and possibility (certainty) of future events and outcomes. Forward looking sentiment in formal company annual reports may be one of the simpler facets of sentiment to detect, compared to for example, 'sarcasm' in the shorthand style typically used in Twitter (Pang and Lee 2008).

In addition to presenting indications of industry climate and company strategy, analysis of modal verbs used in company reports may also provide indications of organizational rhetoric concerned

with persuasion (Ulmer, Sellnow and Seeger 2011) and underlying beliefs and ideologies in the companies concerned. Usage of modal verbs has been reported as being higher in business language (Yasumasa 2008). In standard Information Retrieval (IR) it is even common for some of these words (such as 'might', 'could' and 'will') to be treated as stop words and removed altogether (Manning, Raghavan and Schutze 2008, Li 2010).

Interpreting texts is *"a process of discovery rather than a single stab at explanation"* (Kets de Vries & Miller, 1987, p. 238). There appears a growing realization that exploiting unstructured text can lead to potential insights on the future that cannot be gleaned from traditional numerical data and indices stored in structured databases. There may be value in surfacing hidden patterns in company annual reports using computerized text analysis techniques.

Several categorical dictionaries (KOS) exist in the literature for using modal verbs, adjectives and rules to indicate assertiveness sentiment within text, towards the future state of affairs. However, no prior work has sought to assess the strengths and weaknesses of these models. Specifically, how a composite model could compensate for those strengths and weaknesses when applied to company annual reports in order to assess the extent to which cultural differences and rhetoric can be identified and future performance predicted.

## **2. LITERATURE REVIEW**

Sentiment (tone) analysis has been concerned with automatically identifying subjectivity in text using algorithms. This includes opinion (versus factual), its polarity (positive or negative) and the strength or intensity of that opinion (Taboada 2015). Emotions such as anger, calmness, fear, happiness and surprise have also been detected from text using algorithms (Pulman 2014). Sentiment can be applied to the past (what was done), the present (what is being done) and future tense (what will, can or could be done) with general attitudes towards future likelihoods.

Annual company accounting reports have been extensively analysed through word frequencies of dictionary terms (Rutherford 2005), Natural Language Processing (Lexalytics 2017, El-Haj 2014) and collocation networks (Kloptchenko et al 2004) utilizing techniques such as neural networks (Hajek and Olej 2013).

In a review of the literature, Grimes (2010) suggested that the evidence for human agreement on sentiment is unlikely to better 80-90% (Wilson et al (2005) found 82% agreement between two annotators) and that this should be considered as a context when assessing the performance of automated machine classifiers. Accuracy for text classification appears to be typically in the range of 60-90% as a generalization (Jurka et al. 2013, Faith 2011, Sasaki 2008, Magnuson 2014, Miller 2014), although sentiment categorization may be particularly challenging due to the subtle and subjective nature of opinion (Pang and Lee 2008).

Findings in the literature include evidence in business communication for rhetoric and over-positive language '*the Pollyanna effect*' (Hildebrandt and Snyder 1981). In a study of accounting narratives, Rutherford (2005) proposed the effect as a form of stakeholder 'impression management'. Previous studies have addressed genres such as charged words (e.g. losses *versus* profits), temporal words (e.g. completed *versus* continued) and financial position (e.g. assets *versus* liabilities). There have been numerous manual studies (content analysis) of company accounting reports (Fisher et al 2008). The findings include evidence for smaller and less profitable firms disclosing less information; companies more likely to disclose positive rather than negative information (as noted above); company management tending to attribute negative outcomes to uncontrollable factors while taking credit for positive outcomes and companies disclosing more information during periods of increased earnings (Fisher et al 2008).

Different nationalities and cultures (including native and non-native English speakers) may use modal verbs (e.g. could, must, will, may) in different ways (Hinkel 1995). Studies have shown that articles written in English and Norwegian for example, contain more modal verbs than those written in French (Vold 2006). Research has also shown that non-native English speaking students used some modal verbs (such as can, will and may) with twice the frequency of their English speaking professional counterparts, but used others (such as might) with half the frequency (Hykes 2000). Other studies of marketing disciplines in business however, have shown no statistically significant differences in modal word usage between native and non-native English speakers (Nathan 2010).

Manual content analysis increases subjectivity and is costly to perform. Automated text analytics techniques including the use of concordances and word frequencies is well established. Links have been shown for increases in risk sentiment (words related to 'risk' and 'uncertainty') and lower future company earnings (Li 2006). In a study of financial news in the Wall Street Journal, researchers found high levels of pessimistic words in the column precede lower stock market returns the next day (Tetlock 2007).

Yuet-yung (2014) found that in a study of the worst performers and the top performers in the Fortune 500 in 2012 (including ExxonMobil, Chevron and Walmart), top-performers were more assertive in their presentation of future possibilities. Word use was dominated by modal verbs denoting obligations rather than possibilities to describe future events. Yuet-yung (2014) also found that bottom performers appeared to hedge more, with more of a reluctance to make strong assertions about future events. Conversely, top-performers used a greater frequency of bolder modal verbs which was suggested to relate to their leading market positions, effectively speaking with greater confidence from a position of strength.

Sentiment analysis has been applied to text such as movie reviews using SentiWordNet (2010), assigning a 0 (positive) or 1 (negative) score to words in the large lexical database WordNet (2010). However, this method is not suitable for finer grained categories related to possibility and certainty. Malhotra (2013) identified patterns and synonyms for detecting hypotheses in text using modal verbs and additional verbs and adjectives although no division was made on the strength or intensity of conviction. Bochkay and Dimitrov (2014) assessed the positive (optimism) and negative (pessimism) tone of sentences referencing the future in Management Discussion and Analysis (MD&A) sections of 10-K<sup>1</sup> reports filed with the SEC from 2000 to 2010. They found a systematic bias, when managers were more optimistic, future earnings were low and vice versa.

Finer grained continuous scales have been found to provide more accurate results than binary sentiment (Reagan et al 2015). Continuous scales have been developed for sentiment such as the Semantic Orientation Calculator (Taboada et al 2011) using a scale from +5 (strong) to -5 (weak). This method extracts sentiment-bearing words with a 'prior polarity' and considers 'valence shifters' (words that intensify, down-tone or negate the noun/verb/adjective in question). However, backward looking polarity (whether a sentence is negative or not) such as "*....the target was missed leading to disappointing results*" is not the same as opinions regarding future intent and possibilities "*We will take the necessary measures to improve*".

The use of strong and weak modal verbs as part of an ensemble machine approach using word context has been applied to risk (Wang et al 2013) and fraud detection in 10-K company filings (Humpherys et al 2011). In the latter study it was found that the ratio between hedge cues and total number of words in 10-K filings did not demarcate deceptive information. More frequent use of the hesitant modal verb 'could' and less frequent use of the strong modal verb 'will' did provide statistical significance for identifying fraud.

Loughran and McDonald (2011) found that companies with a higher proportion of weak or strong modal words, were more likely to have a material weakness in internal controls. It has been found that fraudulent reports tend to use more words (Bodnaruk, Loughran and McDonald 2015) supporting Management Obfuscation Theory (Bloomfield 2002). However, it has been reported that annual company reports have grown in size (number of words) by 50% between 2006-2015 due to the increasing complexities for regulation (Deloitte 2015). There is evidence that deception or attempts to conceal information, lead to higher lexical diversity (Siegel, Saukko and Houck 2013).

Minhas and Hussain (2016) found that nine constructs have been used successfully to identify deception using computational linguistics. These include word quantity (more words to obscure truth), cognitive complexity (fewer exclusion terms such as 'but', 'except', and fewer negation words such as 'not') and modal verbs (more hesitant words could lower commitment to facts). So it is possible that organizations using higher modal verb frequencies may have hidden motives.

---

<sup>1</sup> The 10-K form is an annual report required by the Securities and Exchange Committee (SEC) in the US

Most modal verbs are theoretically polysemic (have multiple meanings depending on the context). For example, 'could', expresses a realization/possibility of an event occurring "*Inflation could affect..*" as well as requesting permission "*Could I do this?*" and ability "*I could do this..*". However, in analysis of technical reports, the use of 'can' and 'could' in a permission (rather than possibility) sense, has been reported to be virtually absent, leading to effectively monosemic (single meaning) modal verb usage in certain contexts (Jaime and Perez-Guillot 2015). This was supported by Pique-Angordans, Posteguillo and Andreu-Beso (2002) who found a propensity for many modal verbs (such as can, could, may, might, will, would) within documentation of a technical nature, to display almost 100% epistemic (a hedge) meaning. In analysis of 80 years of the TIME magazine, Millar (2009) found that by the year 2006, the modal verb 'may' was associated with a hedge meaning 94% of the time, whilst 'must' and 'should' were deontic (about the future) 80% and 67% of the time respectively. The significant increase in frequency of 'may' and 'could' in the TIME corpus from 1923 to the year 2006, has been attributed to increased speculative reporting.

Unsupervised machine learning techniques using complex word co-occurrence are capable of surfacing intricate structure within texts (clustered topics) without any external information being input into the original text corpus (Yu et al 2014, Cleverley and Burnett 2015). Roesse and Sikström (2014) applied Latent Semantic Analysis (LSA) techniques to annual company reports of the same company between 2001 and 2010, finding that semantic similarity scores of entire reports varied only slightly. Whilst clustering can be useful, if testing for a specific hypothesis related to sentiment, some categorical information is typically required to be applied to the corpus.

Dictionary (rule based) approaches automatically count the number of times words appear in texts or sentences. Li (2010) found no association between sentiment dictionaries (such as Diction, General Inquirer and Linguistic Inquiry and Word Count (LIWC)) and financial performance in company reports, with the assertion made that the dictionaries did not work well for the financial domain. This is supported by Loughran and McDonald (2011) who concluded that, when financial text is analysed, traditional words described as 'negative' in dictionaries are not negative in a financial sense, such as 'liability', 'cost' and 'tax'. In general, dictionary approaches may not be transferable between domains, if they are domain specific.

Supervised machine learning statistical approaches use a training set of examples from various categories in order to build a statistical model to apply to a wider sample. Li (2010) classified 30,000 forward-looking statements in annual reports and used them to train a Naïve Bayes classifier. The derived model was subsequently applied to 140,000 annual reports. A positive tone was correlated positively with a 5% increase on return the following year. In general, machine learning approaches appear useful if existing dictionaries do not exist, or the domain scope is hard to pre-define.

Some methods include a hybrid approach. For example, Gupta and Liu (2017) inferred organizational culture towards risk of banks by analyzing annual reports of 578 banks using dictionaries of positive and negative words as well as words that represent risk categories. Instead of using raw word counts, assignment (for risk and sentiment) was performed at a paragraph level. An unsupervised clustering algorithm (k-means) was applied in order to group banks of similar types.

Many studies analysing word patterns in company reports seek to link statistically significant results to 'universal laws' that apply to all organizations (such as Bochkay and Dimitrov 2014, Humpherys et al 2011, Fisher et al 2008). Organizations are complex social systems. Rather than obeying 'laws' like the physical sciences, knowledge of organizational phenomena is likely to be more contextual and concerned with generative mechanisms and tendencies, rather than broad generalisations and absolute outcomes.

## 2.1 Lexical modality markers

Tausczik and Pennebaker (2010) suggest we are at the cusp of a technological revolution linking word usage to real-world intentions and behaviours. Word categories may reveal where individuals or entities are focusing their level of immersion and hidden motives. Their Linguistic Inquiry and Word Count (LIWC) dictionary contains over 5,000 words. Informal ‘spoken’ words in the LIWC dictionary (such as ‘*sure thing*’ and ‘*shoo-in*’) are however, unlikely to be common in formal corporate communications. The LIWC dictionary has two categories of interest to this study, ‘tentative’ (e.g. ‘maybe’, ‘perhaps’, ‘guess’) and ‘certainty’ (e.g. ‘always’, ‘never’). The dictionary is of a commercial nature so all the words are not available to the academic community without a financial subscription. These ‘tentative’ and ‘certainty’ category descriptions however, do not directly translate into forward-looking ‘tentative’ and ‘certainty’ opinion. For example, “*the work was done*” has high certainty but is backward not forward-looking.

Some studies have attempted to categorize lexical modality in Bio-medical texts (Thompson et al 2008), as shown in Table 1, presenting the words used in respective categories.

**Table 1** – Modality by lexical category defined from Biomedical texts (Thompson et al 2008)

	Category	Complete List of Words
<b>Knowledge type markers</b>	Speculative	<i>Assume, assumption, belief, believe, claim, conceivable, estimate, expect, expectation, hypothesize, hypothesis, hypothetical, in principle, in theory, judge, model, notion, predict, prediction, proposal, propose, speculate, suggest, suggestion, suppose, suspect, theory, think, to our knowledge, view</i>
	Deductive	<i>Argue, argument, deduce, imply, indicate, indication, infer, interpret, interpretation, suggest</i>
	Demon-strative	<i>Conclude, conclusion, confirm, confirmation, demonstrate, find, finding, proof, prove, report, reveal, show</i>
	Sensory	<i>Apparent, apparently, appear, observation, observe, evidence, evident, seem, see</i>
<b>Certainty markers</b>	Absolute	<i>Certainly, known</i>
	High	<i>Consistent with, clear, clearly, generally in agreement with, likelihood, likely, normally, obviously, probability, probable, strongly, support, would</i>
	Medium	<i>Can, could, feasible, may, might, perhaps, possibility, possible, potential, potentially</i>
	Low	<i>Unlikely, unknown</i>

The results indicate that the prediction of modality can be straight forward using lexical words with a small amount of contextual information. Critiquing Table 1, the use of just two words in the absolute and low categories for certainty markers may lead to sparse data and is unlikely to be comprehensive.

## 2.2 Modal verbs

There are three main groupings of modal verbs according to EOI (2012) as follows:

- Epistemic (A hedge, concerned with knowledge and beliefs. Assessing confidence in propositions, potential facts. Necessity i.e. deduction, possibility, speculation (e.g. this *will* happen, this *may* happen, this *might* happen)).

- Deontic (Generally about the future, how the world should be; Focus's on actions/intentions, obligations, compulsions and permission, trying to exert control over events. Refers to acts not propositions. Necessity (e.g. we *must* do this) or possibility (e.g. we *can* do this)).
- Dynamic (Potential for a situation to occur, ability & volition). Unlike epistemic and deontic modality it is not subjective (Dury 2000). For example, Paul *can* speak Welsh.

There are numerous interpretations of groupings for modal verbs. Table 2 (UNC Chapel Hill 2014) shows one such grouping by the dimensions of 'strong' and 'weak' and by their typical frequency of use. It has been suggested that modal verbs are used most frequently to indicate logical possibility.

**Table 2** - Modal verbs by strength and frequency (UNC Chapel Hill 2014)

	<b>Most frequent</b>	<b>&gt;</b>	<b>&gt;</b>	<b>Least Frequent</b>
	<i>Logical possibility</i>	<i>Ability</i>	<i>Necessity</i>	<i>Permission</i>
<b>Strongest</b>	Must	Can	Must	May
^	Will/would	could	Should	Could
^	Should			Can
^	May			
<b>Weakest</b>	Can/could/might			

Critiquing, Table 2 is missing 'ought' which is a modal verb and probably fits in between the strongest and weakest category. Piotti (2014) identified many devices indicative of 'hedging' with respect to a position on a future state of affairs (Table 3).

**Table 3** – Hedging devices (Piotti 2014)

<b>Categories</b>	<b>Words</b>
Modal auxiliaries	should, will, would, may, can, could, shall, might
Full verbs (reporting)	propose, imply, indicate, suggest
Full verbs (tentative cognition)	expect, assume, estimate, think, believe, evaluate, presume, allege
Adverbs of probability	likely, potentially, basically, possibly, reliably
Adverbs of infinite frequency	generally, regularly, usually, normally, typically, occasionally, rarely

Piotti (2014) complements modal verbs with additional verbs and adverbs that may be indicative of uncertainty. However, many of these words may be predominantly used in both a past (backward looking) or present tense (such as typically) rather than an opinion about the future. Modal verbs have also been grouped 'pragmatically' (TeachIT 2016) into three degrees of certainty (Table 4).

**Table 4** – Modal verbs grouped by degree of certainty (TeachIT 2016)

<b>Degrees of certainty</b>	<b>Modal verbs</b>
Strong	<i>will, shall, must</i>
Moderate	<i>should, would, can, ought</i>
Hesitant	<i>might, may, could</i>

From a possibility perspective, 'must' is very strong (forcing something to occur) whilst 'may', 'could' and 'might' are suggested as the weakest, showing low commitment or confidence.



In an analysis of 10-K company filings, Bodnaruk, Loughran and McDonald (2015) analysed all words that occur in at least 5% of the US Securities and Exchange Committee (SEC) EDGAR filings and extracted what they believed to be words indicative of strong and weak modality (shown in Table 5).

**Table 5** – Strong/weak modal words from SEC filings (Bodnaruk, Loughran and McDonald 2015)

<b>Certainty level</b>	<b>Modal verbs</b>
Strong	<i>will, shall, must, undoubtedly, never, lowest, is, highest, definitely, clearly, best, always</i>
Weak	<i>might, may, could, uncertainty, suggest, sometimes, seldom, possibly, possible, perhaps, occasionally, maybe, depends, depending, could, conceivable</i>

Compared to TeachIT (2016) shown in Table 4, this adds assertive words, some of which may not always relate to future outcomes (e.g. clearly). Cassidy (2013) organized hesitant words (also known as hedges) by function/categories to apply to a corpus (Table 6).

**Table 6** – Hedging lexicon with descriptions (Cassidy 2016).

<b>Category</b>	<b>Description</b>	<b>Examples</b>
Approximation	Indicates proposition is an estimate	About, almost, approximate, estimate, many, most, nearly, some
Degree	Indicates how well proposition fits into a category	Essentially, mostly, partially, quite, relatively, slightly, somewhat, virtually
Frequency	Indicates how often proposition occurs	Generally, normally, occasionally, often, rarely, usually
Intention	Indicates future plans	Intend, plan, propose, seek
Logic	Indicates proposition follows logically	Calculate, conclude, deductive, infer
Modality	Decreases a propositions certainty value	Could, may, might, ought, should, would
Objectivity	Extent to which data ‘speaks for itself’	Apparent, appear, imply, indicate, show, suggest
Prediction	Judgement about the future	Eventually, expect, forecast, maybe, perhaps, predict, project, reckon, somehow, soon, speculate
Probability	Propositions likelihood	Likely, possible, possibility, potential, probable, probably, probability, unlikely
Subjectivity	Proposition based on assumptions	Assumptive, belief, believe, connotative, feel, felt, guess, however, presumably, presumptive, think

In this hedging lexicon (Table 6) many of the words probably fall into the ‘moderate’ strength category which are neither strong (confident) or weak (hesitant) words. For example, ‘generally’, ‘reckon’ and ‘somewhat’. The moderate word ‘intend(s)’ is not a hesitant word (such as ‘may’) but it is not as strong as ‘will’ or ‘shall’. In everyday parlance, some words in Table 6 may be polysemic: used to discuss the past or present as well as future, such as ‘plan’, ‘project’ and ‘conclude’. Including moderate or ambiguous words may smooth or mask subtle opinions at the edges around certainty and hesitancy.

Baker et al. (2012) also highlighted phrases such as ‘have to’, ‘need to’, ‘has to’ and ‘had to’ informally termed ‘semi-modals’ because although they differ syntactically from modal verbs, they

share many of the same meaning characteristics. Whilst ‘had to’ does not convey an opinion regarding the future, ‘have to’, ‘need to’ and ‘has to’ may be useful as markers.

Muslu et al (2015) used three methods in combination to identify forward-looking sentences in financial reports and these are shown in Table 7.

**Table 7 – Words and rules used to identify forward-looking statements (Muslu et al 2015)**

<b>Word Rule</b>	<b>Description</b>
1a Keywords	Will, future
1b Keyword combination	Combining (next, subsequent, following, upcoming, incoming, coming) and (month, quarter, year, fiscal, period)
2. Verb (including lemma’s) conjugation	Combining (aim, anticipate, assume, commit, estimate, expect, forecast, foresee, hope, intend, plan, project, seek, target) with (we, and, but, do not, company, corporation, management, does not, is, are, not, normally, currently, also)
3. Mention of following year	For example, mentions of ‘2017’ in the 2016 annual report

The modal verbs (‘should’, ‘would’, ‘can’, ‘could’, ‘may’ or ‘might’) were not used and any sentences using these words were ignored as they were deemed to be of a ‘legal’ nature rather than forward-looking business opinions. The verb conjugation method (Table 7) was used to avoid false positives by picking up the noun versions of words such as plan (e.g. project plan). The method used by Muslu et al (2015) would identify forward-looking sentences, whereas a pure keyword method may assign multiple ‘hits’ for the same sentence. However just identifying forward-looking sentences only, may lose some measure of ‘intensity’ and strength (strong/hesitant) which was not identified in the study.

A review of the existing literature has led to the development of the following three research questions:

**Q1: Can a composite KOS be created for forward-looking assertive/hesitant sentiment which outperforms existing KOS models?**

**Q2: Do companies in the same industry exhibit different forward-looking word frequency and diversity patterns through time and what explanations could be postulated for those similarities and differences?**

**Q3: Is there an association between the use of strong and/or hesitant forward-looking word frequency and/or diversity and future financial performance?**

### 3. METHOD

A mixed methods critical realist philosophy was adopted for this exploratory study. This enabled both the identification of statistically significant demi-regularities along with explanations for why those may be occurring. The adoption of a stratified ontology enables the hypothesis of unseen hidden motives inferred through their manifest effects evidenced through word patterns. Explanations are therefore grounded in the data but are not constrained by empiricism.

The Oil and Gas (O&G) industry was chosen as it is commodity based (performance is linked to the oil price) and so nuances and differences between multinational companies are potentially more likely to be related to strategy and culture rather than markets. Four large multinationals from the same industry (O&G companies) were selected at random and their annual reports (including 20-F<sup>2</sup>) downloaded from their websites for the years 2008-2015. All data is in the public domain, however the four randomly chosen companies are coded 'Company A', 'Company B', 'Company C' and 'Company D' to focus the reader on the method and concepts, rather than specific company instances.

A dictionary (rule based) method was selected for this exploratory study for two reasons. Firstly, it is assumed that forward-looking strong and hesitant sentiment can be well defined and represented by a composite dictionary of terms from those that already exist in the literature. Secondly, for an exploratory study with a small dataset, a machine learning dataset may not be well suited to building a statistical model.

#### 3.1 Word categories and types

To avoid smoothing out small patterns, only the categories of 'strong' (certain) and 'hesitant' (weak) forward-looking sentiment are used in this study to pick up extreme (edge) forward-looking opinions. The selection of extreme edges enables a simple counting polarity based approach, rather than any gradational continuous scale. Using this lens, appropriate words (and stems) were extracted from Tables 1-7. All of the words were used from Table 5, augmented with selected words from the other tables and semi-modal concepts from Baker et al (2012). The composite set is shown in Table 8.

**Table 8 – Words and categories used in the analysis**

Category	Words/concepts
Strong	<i>will, won't, shall, must, certainly, known, definitely, always, is, undoubtedly, believe, has to, have to, need to, commit, aim, expect, anticipate, think, aspire, strive, optimistic, going to</i>
Hesitant	<i>might, may, could, unlikely, unknown, uncertain, suggest, sometimes, possibly, possible, perhaps, occasionally, depends, depending, seldom, conceivable, maybe, guess, speculate, hope, imaginably</i>

This was used to extract frequencies from the text in the study. Syntactic structure for readability was assessed using the Gunning Fog Index (Gunning 1969), incorporating the average number of words in a sentence and the number of 3+ syllable words. Typical scores for financial reports of publicly traded companies are 15-20, meaning they can be understood by people with 15-20 years of education, with evidence of fraud being associated with scores in the high thirties (Moran and Kral 2013). Sample sections of approximately 500 words were taken from each company's annual report related to the sections on outlook/Chairman/Chief Executive Officer's letter and the fog index calculated.

---

<sup>2</sup> A 20-F is an SEC filing made by certain foreign private issuers

### **3.2 Validation with human judges**

Some studies such as Li (2010) rely on the researcher to identify sentences that are indicative of the sentiment being analysed. However, as stated by Grimes (2010), relying on a single human assessment of sentiment is likely to lead to bias. Therefore, five human judges were purposefully recruited from the business social network site LinkedIn [www.linkedin.com](http://www.linkedin.com) and personal networks of the researchers. Each was an experienced business professional or academic in an information based discipline, the judges acting effectively as informants due to their knowledge. To ensure some level of diversity, the judges were from Europe and North America, two women and three men.

Each judge was given a random company annual report and asked to identify thirty sentences from any part of the report, which for them, represented a strong forward-looking sentiment, a hesitant forward-looking sentiment and neutral (neither) as a control. Each judge was therefore asked to identify ninety examples in total, to be cut and pasted to a notebook text file, labelled with the category and sent back to the researchers via email. Care was taken not to 'prime' the participants with trigger words to avoid biased data collection.

This exercise generated 450 test examples, with 150 in each of the three categories (strong, hesitant, neither). This was deemed sufficient based on practitioner based heuristics for machine learning classification, which indicate that 50-100 labelled training examples are typically required to give good results per category (Hedden 2013, Faith 2011).

The rules-based composite dictionary (Table 8) was applied to these training examples to identify recall (how many of the strong and hesitant sentiment examples would be identified) and precision (how many incorrect categorizations were made). An F1 score was calculated (weighted average of precision and recall) which takes into account both false positives and false negatives. An average from these judges scores was used. These data would support a judgement on whether the composite dictionary (KOS) was a reasonable surrogate for determining forward-looking sentiment within company annual reports.

### **3.3 Calculating word frequency and diversity data**

Using the words/concepts in Table 8 and OpenSource utilities including Python NLTK scripts applied to the text in company reports, word frequencies were calculated per category, per company and per year.

Mentions of the future included the explicit term 'future' and mentions of the following year in reports (for example, the mention of 2011 in the 2010 annual report) would be a successful match.

Supported by the literature (Jaime and Perez-Guillot 2015, Pique-Angordans, Posteguillo and Andreu-Beso 2002, Millar 2009), an assumption was made for this context that word usage is largely monosemic (single meaning). Analysis of concordance data however did parse for negation to remove these false positives from the respective categories. Negation is not straightforward as the phrase "not possible" changes the word 'possible' from hesitant/uncertain to strong/certain whilst the phrase "will not" does not change the certainty of the word 'will'. These natural language part of speech concept differences were included in the rule set as a crude form of Word Sense Disambiguation (WSD), as a generic Bag of Words (BoW) approach would ignore these contextual differences. Pang and Lee (2008) indicate that catering for negation can improve accuracy by as much as 3%.

The total frequency of strong words ( $S_f$ ) were divided by the total frequency of moderate and hesitant words ( $H_f$ ) to create a ratio  $S/H$ . Any corresponding increase in assertive words and decrease in hesitant words would therefore be amplified.

A measure of diversity for the use of strong and hesitant words for each company is inspired from biological ecology, assessed using an Equitability ( $E_D$ ) ratio which is the Simpson Diversity Index (Peet 1974) divided by the number of word types per category, given by the equation:

$$E_D = \frac{D}{D_{MAX}} = \frac{1}{\sum_{i=1}^s p_i^2} \times \frac{1}{s}$$

Where:

$D$  = Simpson diversity index

$S$  = Total number of species in the community (richness)

$p_i$  = Proportion of  $S$  made up by the  $i$ th species

$E_D$  = Equitability (evenness)

The  $E_D$  ratio is a range between 0 and 1, with 1 representing true evenness. For example, if there were ten types of words (or categories) where individual words occurred 1,000 times, an evenness of 1 would equate to a frequency of 100 for each of the ten word type/categories.

The resulting strong to hesitant ratio ( $S/H$ ) and evenness ratio ( $E_D$ ) were plotted for each of the four companies for the years (2008-2015).

Linear regression was performed to ascertain whether or not there were any strong associations. For small datasets, Moore, Notz and Flinger (2013) suggest an  $R^2 > 0.7$  provides a strong correlation. This was tested for frequency and diversity of word usage over time and with respect to financial performance.

Revenue figures (see Appendix I) were extracted from each report (including 2005-2007) in order to create a three-year moving average for each company, reflecting percentage revenue change. The close relationship with oil price is evident (Appendix II). The four companies were then compared to each other for each three-year window by calculating the percentage change compared to the mean of the four companies, given by the equations below:

$$AR_{cy} = \frac{\sum_{i=1}^n R_{cy}}{n}$$

$$ARPC_{cy} = \frac{AR_{cy}}{AR_{c(y-1)}} \times 100$$

$$ARPC\_DM_{cy} = \frac{\sum_{j=1}^c ARPC_{jy}}{c}$$

Where:

$R$  = Yearly Revenue (for company  $c$  in year  $y$ )

$n = 3$  (size of moving average window)

$c = 4$  (number of companies studied)

$AR_{cy}$  = Average revenue (3 year moving average for company  $c$  in year  $y$ )

$ARPC_{cy}$  = Percentage revenue change 3 year moving average ( $y$  to  $y-1$ )

$ARPC\_DM_{cy}$  = Percentage revenue change above or below the mean for  $c$

The  $ARPC\_DM_{cy}$  measure was used as a surrogate for relative business performance amongst the four peers.

### **3.4 Association between word frequency and revenue**

For the 2009 company annual reports onwards, new 20-F reporting disclosure rules (Milbank 2009) made it difficult to compare word frequency ratios before and after 2009. Therefore, only data from 2010-2015 were analysed.

Scatter plots and linear regression were used to explore relationships between strong/hesitant (S/H) word frequency and diversity ratio's compared to the following year's revenue between 2010-2015.

### **3.5 Study Limitations**

As raised by Hykkes (2000) modal verb usage can be influenced by nationality and culture rather than (or in addition to) corporate ideology, although differences in professional business contexts has been dismissed as negligible (Nathan 2010).

The entire public company annual report contents were used, as 6 of the 32 company report Adobe PDF file formats were locked and therefore frequencies had to be calculated using Adobe utilities. It was therefore not possible to isolate the business reporting sections. The additional sections including legal statements may therefore have influenced trends, although this was consistent for all companies.

A limitation of the method applied is generally assuming a Bag of Words (BOW) approach and monosemy (single meaning) for the modal verbs. However, natural language context negation was catered for in this study (section 3.3) and the inclusion of concepts (from Baker et al. 2012) rather than just single words moved the KOS closer to a Bag of Concepts (Cambria and Hussain 2015). Nevertheless, limitations of the BoW approach are well documented in the literature (Chan and Cong 2016). It is likely that in some contexts, the same modal verbs may be used with different meaning (polysemy) so would need to be disambiguated. However, supported by existing research, it is assumed the modal verbs analysed in this context have tendencies towards monosemic usage. The method is therefore deemed a valid approach for this exploratory study.

A sample of four companies means statistical generalisation of findings is not possible. However, generalisation of theoretical propositions can be proposed as areas for further research. In this exploratory study, the discovery of any potential association is intended to stimulate further large scale studies.

#### 4. RESULTS

Between 2008 and 2015, the total number of words used in annual company reports increased by 44% on average in the sample. Company C showed the greatest increase (71%) followed by Company D (56%), Company B (35%) and Company A (11%). The Gunning fog index scores were relatively equal, Company A=15.5, Company B=16.6, Company C=16.5 and Company D=16.3. These fall within the ranges of scores for reports of typical publicly traded companies.

In Company A the annual report contained a section on forward-looking statements which could be identified by words such as: ‘anticipate’, ‘believe’, ‘could’, ‘estimate’, ‘expect’, ‘goals’, ‘intend’, ‘may’, ‘objectives’, ‘outlook’, ‘plan’, ‘probably’, ‘project’, ‘risks’, ‘schedule’, ‘seek’, ‘should’, ‘target’ and ‘will’. Company C and D had a similar section with the same words, whilst Company B had no such statement.

##### 4.1 Evaluating the composite KOS to human judgements

The accuracy and completeness of the automatically applied rules based composite dictionary from Table 8, compared to the 450 human sentiment judgements is shown in Table 9.

**Table 9** – Recall, Precision and F1 scores for the Composite KOS applied to human judgements

	Strong Forward Looking Sentiment				Hesitant Forward Looking Sentiment		
	Recall	Precision	F1		Recall	Precision	F1
Judge 1	0.89	0.89	0.89		0.81	0.91	0.86
Judge 2	0.38	0.9	0.53		0.92	0.92	0.92
Judge 3	0.64	0.62	0.63		0.22	0.54	0.31
Judge 4	0.93	0.62	0.74		0.4	0.8	0.53
Judge 5	0.86	0.93	0.89		0.93	0.97	0.95
AVERAGE	0.74	0.792	0.77		0.66	0.83	0.73

Average accuracy F1 scores for the composite KOS (Table 8) as applied to the judges examples were 77% for strong forward-looking sentiment and 73% for hesitant forward-looking sentiment (Table 9).

##### 4.2 Comparing the composite KOS performance to existing models

Simple inference counting of ‘is’ for strong forward-looking sentiment was not found to predominantly represent a strong certainty sentiment about future possibilities. For example “*The office is located in xxx*” (Company B) and “*There is significant uncertainty*” (Company C). The word frequency of ‘is’ represented 68.4% of all strong words, heavily skewing the trend and so was therefore removed prior to context-free word frequency analysis due to its polysemy.

The words ‘won’t’, ‘undoubtedly’, ‘anticipate’, ‘think’, ‘aspire’, ‘strive’, ‘certainly’, ‘definitely’, ‘going to’, ‘seldom’, ‘hope’, ‘optimistic’, ‘commit’, ‘conceivable’, ‘maybe’, ‘guess’, ‘perhaps’, ‘unlikely’, ‘suggest’ and ‘speculate’ either did not occur in any company reports or occurred less than 10 times across the entire set. Table 10 shows the words/concepts used in this study (from the composite KOS in Table 8) that generated more than ten matches within the corpus of 32 reports.

**Table 10** - Word frequency counts (32 annual reports)

<b>Strong</b>		<b>Moderate</b>		<b>Hesitant</b>	
<i>Word</i>	<i>Frequency</i>	<i>Word</i>	<i>Frequency</i>	<i>Word</i>	<i>Frequency</i>
Will	9392	Should	1186	may	7270
Shall	1769	Would	2351	could	2894
Must	1599	Can	3705	Possible(ly)	1662
Believe	1471			Uncertain(ty)	1221
Expect	645			Depends(ing)	762
Known	532			might	262
Aim	504			sometimes	91
Have to	188			unknown	84
Need to	179			occasionally	21
Always	178				
Has to	85				
<b>TOTAL</b>	<b>16542</b>	<b>TOTAL</b>	<b>7242</b>	<b>TOTAL</b>	<b>14267</b>
<b>Evenness</b>	<b>0.26</b>	<b>Evenness</b>	<b>0.85</b>	<b>Evenness</b>	<b>0.34</b>

In the strong category, word frequency was dominated by 'will' (57%) and in the hesitant category word frequency was dominated by the word 'may' (51%).

Table 11 compares the performance of existing KOS's to the composite KOS created in this study as applied to the annual company reports corpus. Green cells indicate where a term used in the composite KOS is part of the respective model used by the author(s) in their KOS rules based dictionary.



**Table 11** – Comparison and performance of various KOS to the composite KOS (Table 8)

				Thompson et al (2008)	UNC Chapel Hill (2014)	Piotti 2014	TeachIT (2016)	Bodnaruk, Loughran and McDonald (2015)	Cassidy (2016)	Muslu et al (2015)	Baker et al (2012)
Strong	Count	%		1	2	3	4	5	6	7	8
will	9392	56.78									
shall	1769	10.69									
must	1599	9.67									
believe	1471	8.89									
expect	645	3.90									
known	532	3.22									
aim	504	3.05									
have to	188	1.14									
need to	179	1.08									
always	178	1.08									
has to	85	0.51									
<b>TOTAL</b>	<b>16542</b>	<b>100.00</b>	<b>%</b>	<b>3</b>	<b>57</b>	<b>72</b>	<b>78</b>	<b>87</b>		<b>64</b>	<b>81</b>
Hesitant	Count	%									
may	7270	50.96									
could	2894	20.28									
possible(ly)	1662	11.65									
uncertain(ty)	1221	8.56									
depends(ing)	762	5.34									
might	262	1.84									
sometimes	91	0.64									
unknown	84	0.59									
occasionally	21	0.15									
<b>TOTAL</b>	<b>14267</b>	<b>100.00</b>	<b>%</b>	<b>1</b>	<b>73</b>	<b>85</b>	<b>73</b>	<b>99</b>	<b>94</b>		

The total percentage for each column (the blue row at the base of each section) is relative to the composite KOS. For example, in column 7 for strong forward-looking sentiment, Muslu et al (2015)'s

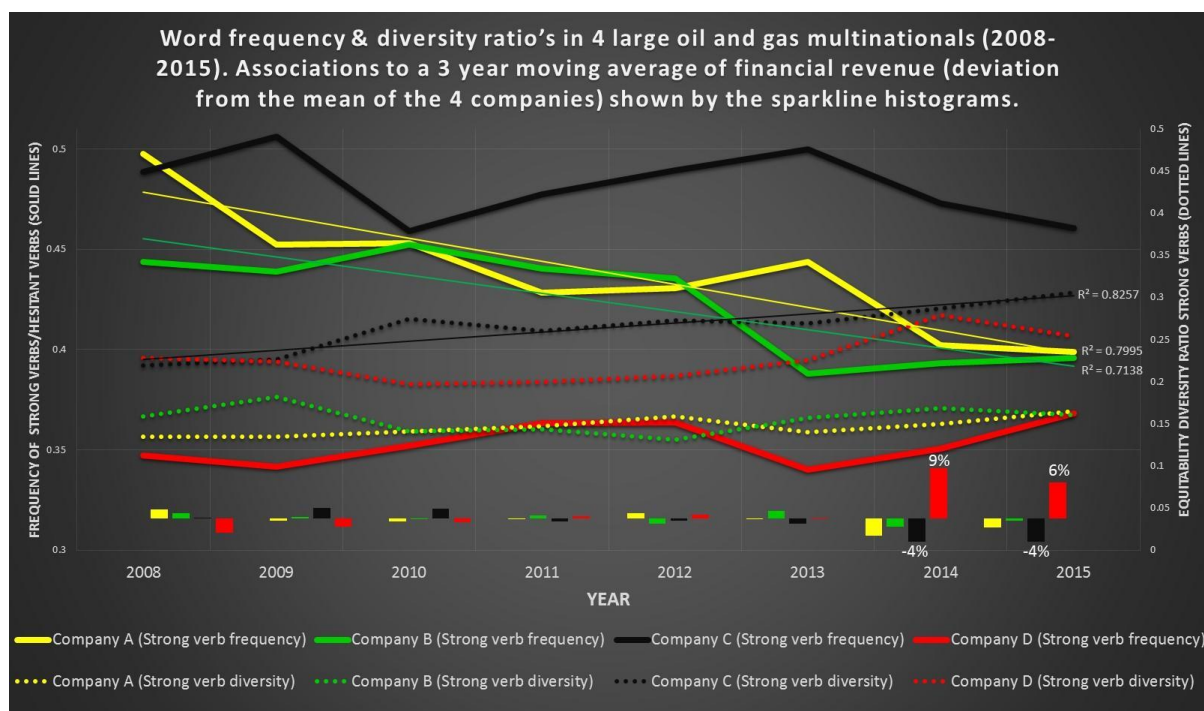
KOS model found 64% of the matches compared to the composite KOS used (Table 8). In other words, the composite KOS found 56% more strong forward-looking sentiment matches.

For the strong sentiment for forward-looking opinion, the composite performed significantly better than existing models, with Bodnaruk, Loughran and McDonald (2015) the closest model, finding 87% (Table 11) of words found by the composite. For the hesitant sentiment for forward-looking opinion, Bodnaruk, Loughran and McDonald (2015) performed very well at 99% (Table 10) of the composite.

A composite KOS therefore appears to have the potential to perform better than existing models in the literature, particularly in the area of strong forward-looking sentiment.

#### 4.3 Word frequency and diversity similarities and differences amongst companies

The strong/hesitant (S/H) word ratio's and diversity from Table 8 (split by company and year) are shown in Figure 1.



**Figure 1** – Changes in the use of word ratios (Strong/Hesitant word frequency - solid lines and diversity – dotted lines) for four O&G companies (2008-2015). Histogram sparkline shows deviation from the mean for the 4 companies using a three year moving average revenue percentage change.

A thick solid line was used for the S/H frequency ratio and dotted line for the S/H ( $E_D$ ) ratio. Strong linear regression trend lines are shown in a faint solid line with their  $R^2$  on the right hand side of Fig 1.

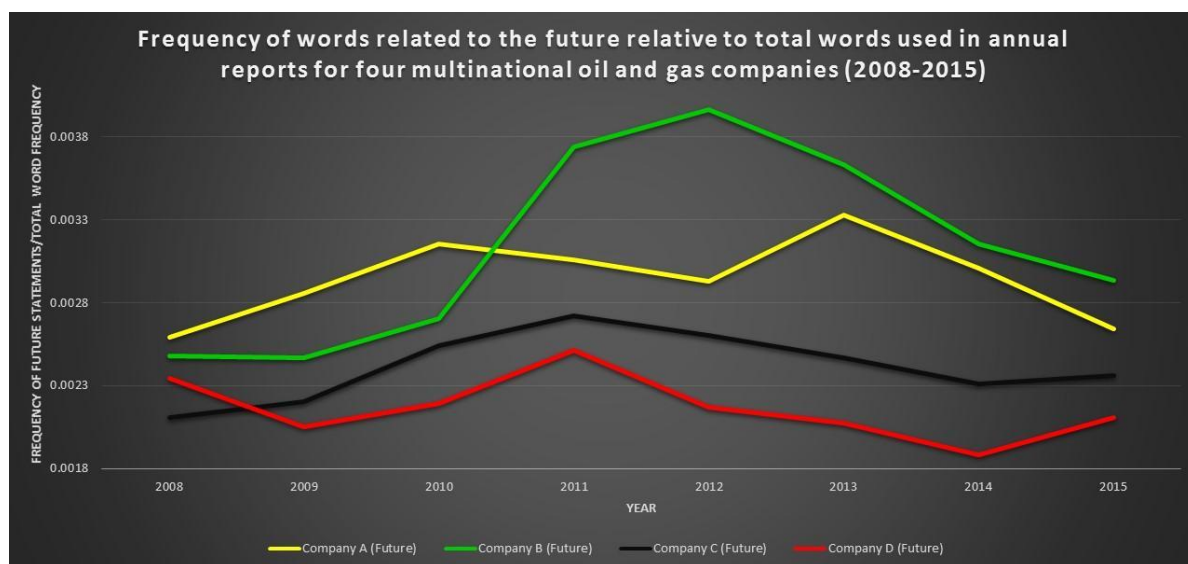
The Strong to hesitant (S/H) forward-looking word frequency ratio's (solid lines) for Company A (yellow) and Company B (green) showed a strong declining relationship with time (2008-2015) as shown in Figure 1. Assertive language reduced, with statistically significant correlations ( $R^2$  of 0.79 and 0.71 respectively). Company C (black) showed the highest S/H ratio's and Company D (red) showed the lowest S/H ratio's in the time period studied.

The equitability diversity (evenness) ratio ( $E_D$ ) for Company C (black dotted line) showed a strong increasing relationship with time, assertive language becoming more diverse ( $R^2=0.82$ ). The diversity

ratio's for Company A and Company B (yellow and green dotted lines respectively) were similar, lower and remained virtually the same over the entire time period studied. The diversity of strong words for Company D (red dotted line) dissected the other companies and yielded no strong associations through time (Figure 1).

At the base of Figure 1, the sparkline histogram ( $ARPC_{DM_{cy}}$ ) shows that in 2008, Company A (yellow) and Company B (green) performed above the average (of the four companies) with respect to revenue changes in a three-year moving average. Company C was neutral and Company D (red) was the worst performer. By 2014/2015, in relative terms, Company D (red) had become the top performer (9% and 6% above the average) with Company C the worst (4% below the average revenue percentage change). There was significantly more variation between 2014-2015 than in previous years.

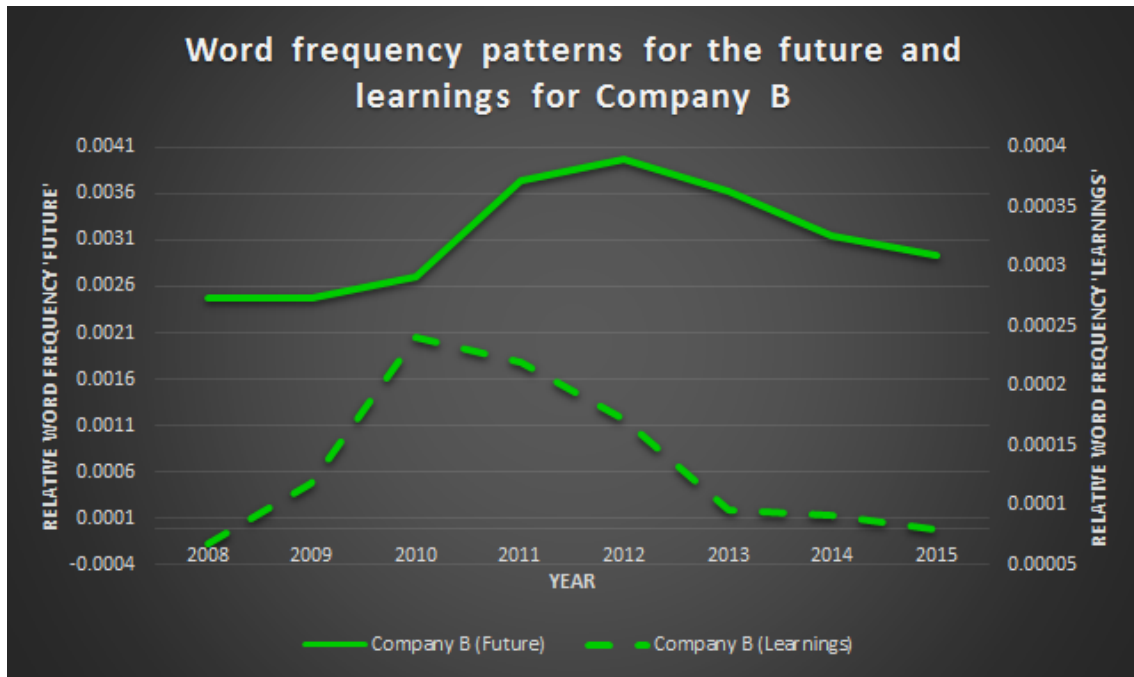
Figure 2 shows the results of counting the mentions of 'future' or the following year in the annual reports of the 4 companies between 2008 and 2015 (32 annual reports in total).



**Figure 2** – Relative frequency of words mentioning 'the future' for the 4 companies over time

Company C and D appeared to have relatively uniform changes (although nothing statistically significant), whilst Company A and B exhibited some sharp gradient changes from year to year. For example, from 2010 to 2011, Company B (green) showed a 38% increase in mentions of the future. Company A (yellow) showed the second highest 'spike' between 2012 and 2013 of 14%.

As part of the iterative process of discovery, further data were collected on word frequency for Company B to explore this gradient change around 2010-2011. It was found that the concept of learning (represented by the stems 'learn' and 'lesson') also showed steep gradient changes just before that time (Figure 3).

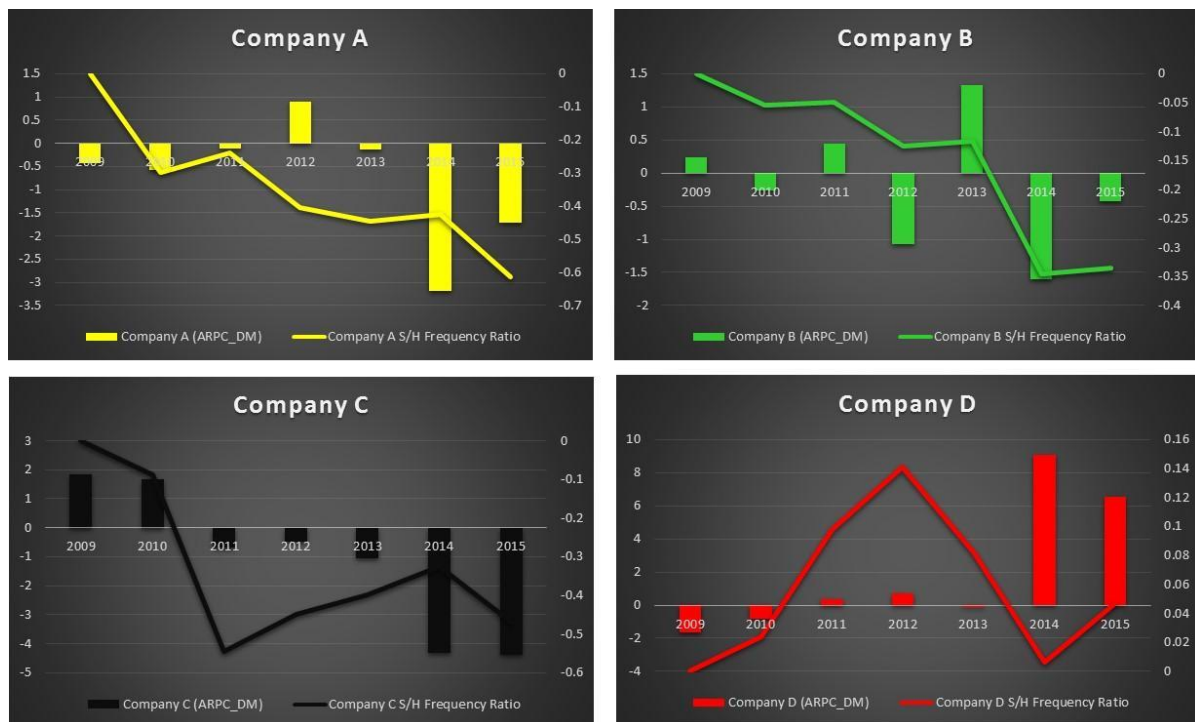


**Figure 3** – Relative word frequencies for ‘learning’ compared to ‘future’ (from Fig 2) for Company B

For the ‘learning’ concept, the major increase in word frequency gradient change (over 100%) occurred from 2009-2010 (dotted line Figure 3), prior to the increases in the ‘future’ concept between 2010-2011 (solid line Figure 3). Potential mechanisms are discussed in section 5.

#### 4.4 Correlating word patterns to future financial performance

Figure 4 shows a bar chart for moving average revenue change relative to the mean change for the four companies (*ARPC\_DM*) by individual company.



**Figure 4** –S/H frequency ratio for preceding year against relative business performance (*ARPC\_DM*)

A positive delta value (y-axis) means the company showed above average improvement (for the four companies) in its percentage revenue change (regardless of whether that was an absolute positive or negative revenue change). This is plotted against a line graph for the *previous years S/H* frequency ratio to visualize the predictive correlation between changes in strong and hesitant word frequency ratios the previous year, and subsequent business performance.

Performing linear regression, there were no strong relationships between the previous years S/H ratio and subsequent financial business performance.

Figure 5 shows the previous years S/H diversity ratio as a line graph for the four companies compared to relative business performance (ARPC\_DM) shown as a bar chart.



**Figure 5-** S/H diversity ratio for preceding year against relative business performance (ARPC\_DM)

There was a moderate/strong correlation ( $R^2 = 0.6365$ ) for Company C (in Black). As relative business performance (compared to the other three companies) decreased, the diversity of strong/hesitant language used within the previous years annual company report increased. An increase in diversity of language therefore for Company C may imply worsening performance (Figure 5).

## 5. DISCUSSION

The average increase in the total number of words in annual reports between 2008 and 2015 was 44%, roughly in line with the 50% figure stated by Deloitte (2015).

The method to analyze, select and combine words from Tables 1-6 into a composite model appeared justified due to the absence or low frequency of some words within existing models. For example, in Table 1 (Thompson et al 2008) there were only two words defining both the absolute and low certainty categories and one of these words in each category did not occur in the dataset (or had a frequency <5). Some words such as 'is' (Bodnaruk, Loughran and McDonald 2015) appeared to produce many false positives. None of the models identified (Tables 1-6) included appropriate semi-modal verbs (such as 'have to', 'need to', 'has to') so would have missed over 3% (Table 10) of occurrences of assertive strong forward-looking sentiment (Table 8).

Contrary to the assertion made by Muslu et al (2015) that modal verbs such as ‘could’ and ‘may’ were only used in a legal sense, many examples were provided by the human judges where ‘could’ and ‘may’ were deemed markers for forward-looking business sentiment.

Applying the KOS in a simple non-contextual, non-probabilistic way, generated F1 scores (75% average) that exceed the performance of some analogous classification studies (Jurka et al. 2013) and practitioner heuristics (Faith 2011) and fall within typical accuracy ranges (60-90%) as defined by Sasaki (2008), Magnuson (2014) and Miller (2014).

It is therefore proposed that the final composite list of words (Table 8) adequately represents forward-looking intensity and improves on existing models in the literature. This may provide a starting point towards a more compact and comprehensive representation of future certainty markers than existing models. Moving to a probabilistic Bayesian type and more sophisticated Natural Language Processing may enable additional rules to be used, context disambiguation to be applied and accuracy to be achieved.

Both Company A and Company B showed a strong association over the period 2008-2015 with a decline in the use of strong words/assertive language about the future (Figure 1). It is possible that this reflected increasing business uncertainty. This could be explained through a narrative which ties word frequencies to global events. The financial crisis had just occurred (2007/2008) and the oil price had fallen to its lowest level for four years by the end of 2008. Although rising again, the oil price dropped by over 50% from 2012 to 2015 (Appendix II). By 2015, Company A and Company B had reduced their use of assertive language regarding the future by 20% and 10% respectively from 2008.

Mentions of the ‘future’ relative to total words in annual reports showed some major deviations between companies through time (Figure 2). In particular, the 38% increase in mentions of the ‘future’ for Company A from 2010-2011 and 14% for Company B from 2012-2013. The annual reports and historical news archives provide evidence for two events (crises) that may have triggered these patterns. A crisis in this context is defined as, *“specific, unexpected, and non-routine events or series of events that created high levels of uncertainty and threat or perceived threat to an organization’s high-priority goals”* (Seeger, Sellnow and Ulmer 1998, pg. 231).

Firstly, in 2010 Company B had a major industrial accident (crisis) that received significant news coverage. This may have influenced company attitudes and rhetoric to focus on the ‘future’ as a vision to move forward. Secondly, in 2012, Company A suffered an accident (although nowhere near as serious as Company B) which received embarrassing news coverage. The small increase in mentions of the ‘future’ may have occurred for the same reasons. Both organizations may have been participating in a future-based ‘developmental conversation’ with stakeholders, as a form of impression management, *“While we can’t ignore the past, we also can’t change it. We can learn from it, but we shouldn’t dwell on it”* (Levin and Edwards 2007, pg. 155).

This proposition is supported by the increases in the word frequency of the ‘learnings’ concept (Figure 3). The word frequency patterns of the ‘future’ and ‘learnings’ concepts could support a narrative based on Discourse of Renewal Theory (DRT) as proposed by Ulmer, Sellnow and Seeger (2011). As part of an organizational rhetorical framework in a time of crisis, DRT focusses on renewal, growth and transformation. Reflecting on a crisis, it describes sense-making which contains a ‘learning’ component to gain confidence from stakeholders and providing a ‘future’ prospective vision for moving forward as a response to a crisis. The word patterns observed (Figure 2 and Figure 3) may support this theory.

The magnitude of the frequency ‘wavelength’ for increases in mentions of the ‘future’ may be related to the size of the crisis. For example, for the major disaster (Company B), 5 years later the level of word use had not dropped back to those prior to the accident (Figure 2). The ‘wavelength’ for the ‘learnings’ concept started earlier (in the immediacy of the accident) and had a smaller wavelength, returning to levels prior to the accident within 3 years (Figure 3). This may suggest a certain ‘sequencing’ with respect to word patterns in company reports responding to a crisis. First an initial focus on ‘learnings’ followed by a move to discuss the future. The ‘wavelengths’ for Company A (smaller accident) appear shorter. This may present an area for further research.

Company C showed a strong association between increasing diversity of strong assertive forward-looking language over time (Figure 1) and future decreasing revenues (Figure 5), despite being the worst performing company out of the four by 2013-2015. Its use of strong forward-looking assertive language regarding the future was also four times that of Company D which was the top performer in the sample by 2015. Nationality and cultural differences could be one explanation, however there is no evidence that a multinational company report written in English is influenced by the nationality of the country where its head office is based.

A competing explanation is that Company C deploys more optimistic rhetoric in its annual report, supporting the *“Pollyanna effect”* (Hildebrandt and Snyder 1981). This is proposed as the most plausible explanation based on the evidence collected in this study that links a number of markers for Company C to over-positive reporting. Four lines of evidence support this explanation. Despite being the worst performer of the four companies by 2013, Company C had:

- (i) the greatest increase in company report word length from 2008 to 2015 (71%), over 6 times the increase of Company A for the same period. It has been reported that there is a tendency for more words to be used in reports that are trying to conceal or obscure information (Bodnaruk, Loughran and McDonald 2015, Minhaus and Hussian 2016)
- (ii) the highest frequency (Figure 1) of strong forward-looking sentiment modal words – a potential sign of material weakness in controls (Loughran and McDonald 2011)
- (iii) the highest diversity (Figure 1) of forward-looking strong sentiment – a potential indicator of deception (Siegel, Saukko and Houck 2013)
- (iv) a moderate/strong correlation between increasing diversity of strong forward-looking sentiment and subsequently decreasing financial performance the following year (Figure 4).

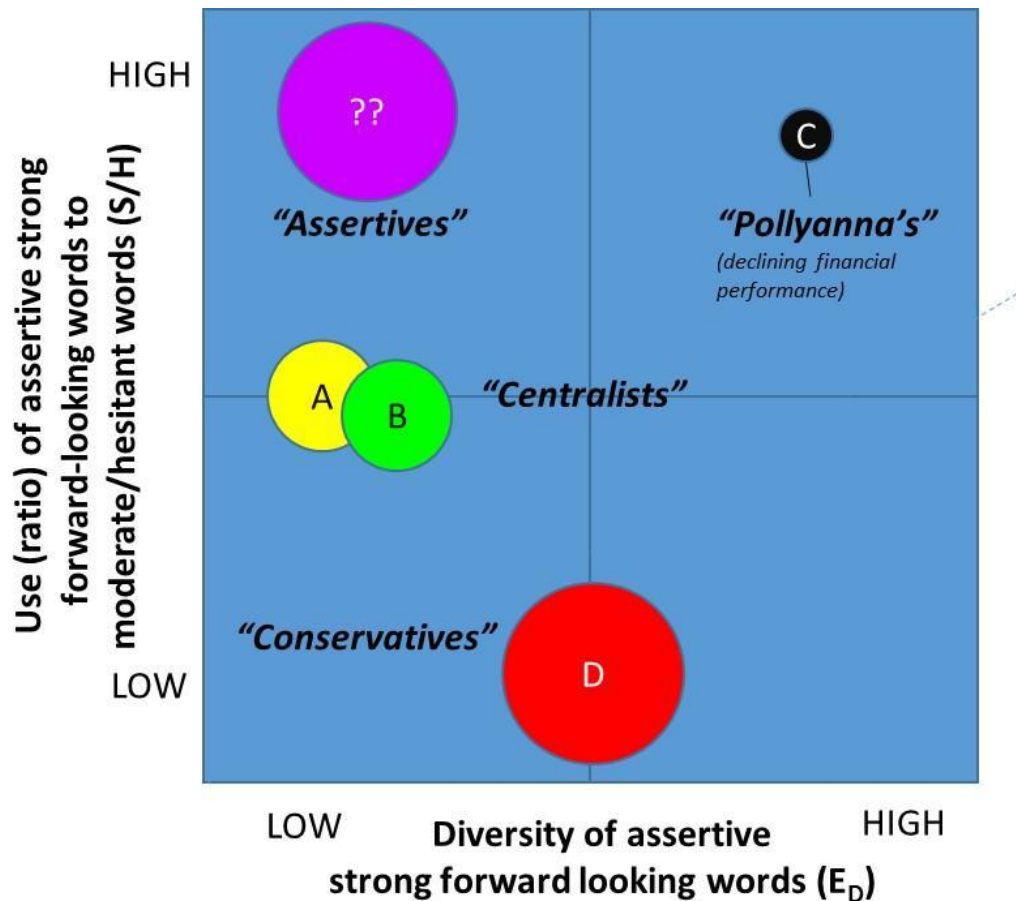
From Figure 1 (dotted lines) it can be seen that Company A and Company B showed a lower diversity of assertive language regarding the future than Company C and D. Company A and B may represent ‘middle of the road’ ideologies in their annual reports. They have neither the extreme use of strong words or diversity shown by Company C nor the minimal use shown by Company D. Their diversity of language is also low and consistent through time.

Company D went from being the worst performer in the group (2008-2010) to the best (2014-2015) as shown in the sparkline histogram in Figure 1, despite having the lowest ratio through time for assertive/hesitant language. No strong correlations existed for Company D between word frequency/diversity and financial performance. This finding contradicts Yuet-yung (2014) who suggested top performers would be more assertive in their presentation of future outcomes. One explanation could be the study sampling, which analysed four very large companies that occupy the top echelons of various financial market indices. Relative differences in word use between these companies may indicate other generative mechanisms, such as organizational cultural differences,



rather than outright fraud/deception extremes that previous studies have focused on using companies of markedly different sizes and profitability.

The findings from the study imply that some companies may behave in different ways (archetypes) rather than universal laws that apply to all. Figure 6 proposes a tentative model based on the study findings.



**Figure 6** – A proposed model for differing underlying company cultural ‘norms’ inferred by frequency/diversity ratio’s. The size of the circle is relative financial performance.

Company A and B are termed *"Centralists"* in that they occupy the middle ground compared to their peers for use of strong assertive forward-looking language and diversity. They appear to react (through their strong word frequency usage) to changing business uncertainty. Company D occupies a position termed *"Conservative"*, which is characterized by low use of assertive strong forward-looking words which has remained relatively unchanged, even when it performed well compared to its peers. This may evidence a more complex and perhaps ‘cautionary’ or ‘academic’ approach to communicating the future state of affairs than its peers.

A third suggested category (in purple) is termed *"Assertives"* which shows high frequency of strong word use and superior financial performance although there is no empirical evidence from this study for the existence of the group. It is theorized that some companies may exhibit these characteristics.

Finally, Company C is proposed as a *"Pollyanna"* where its high use of diverse assertive strong forward-looking language along with other markers, is not particularly justified by its performance



within the group. It may fit what Rutherford (2005) termed extreme *impression management*. Other companies may fit this archetype.

## **6. CONCLUSION**

The study has developed a composite rules-based dictionary for intensity of forward-looking sentiment. This is shown to be an improvement on current models described in the literature.

No previous studies have been found which examine the association between increasing diversity of strong forward-looking sentiment in company reports and corresponding decreasing future financial performance. The findings from this initial study may therefore act as a catalyst for further research.

Relative word frequency patterns in company reports may be useful to chart underlying trends in their attitudes towards the future. This can give voice to a narrative which may not necessarily be explicit reading company reports and may be increasingly useful in the 'post-truth' society.

Whilst some companies render changing attitudes towards the future state of affairs through word usage in their annual reports, others may not and some may even deploy rhetoric. These different behavioural archetypes may point to a more complex model than simple 'universal' laws when attempting to interpret text within company reports using automated techniques.

Use of words without their context and a small sample size is a limitation of this study. Further research could apply machine learning techniques such as training a Bayesian classifier and Natural Language Processing (applying the human judgement examples and composite clues identified by this study) to specific sections in reports from a larger sample of companies (of different sizes and industry sectors).

Sentiment is typically applied with an *a priori* hypothesis in mind. Embedding these sentiment algorithms in standard enterprise search and discovery technology deployments may help generate new insights and knowledge in the most unexpected of places.

## **REFERENCES**

Aaltonen, A. and Tempini, N., (2014). Everything counts in large amounts: A critical realist case study on data-based production. *Journal of Information Technology*, 29(1), pp. 97-110.

Baker, K. et al (2012). Modality and Negation in SIMT Use of Modality and Negation in Semantically-Informed Syntactic MT. *Computational Linguistics*, 38(2), pp. 411-438.

Berelson, B. R., (1952). Content Analysis in Communication Research. The Free Press, Glencoe, IL.

Bholat, D. et al., (2015). Text mining for central banks: handbook. Centre for Central Banking Studies (33). pp. 1-19. ISSN 1756-7270

Bloomfield, R.J. (2002). The 'incomplete revelation hypothesis' and financial reporting. *Accounting Horizons*, 16, pp. 233-243.

Bochkay, K. and Dimitrov, V. (2014). Qualitative Management Disclosures and Market Sentiment (December 15, 2014). Rutgers University. Online Article: <https://ssrn.com/abstract=2538812> or <http://dx.doi.org/10.2139/ssrn.2538812>

Bodnaruk, A., Loughran, T. and McDonald, B. (2015). Using 10-K Text to Gauge Financial Constraints. Online Article ([https://www2.warwick.ac.uk/fac/soc/wbs/subjects/finance/events/seminars/using\\_10-k\\_text\\_to\\_gauge\\_financial\\_constraints.pdf](https://www2.warwick.ac.uk/fac/soc/wbs/subjects/finance/events/seminars/using_10-k_text_to_gauge_financial_constraints.pdf))

Bozanic, Z., Roulstone, D. T. and Van Buskirk, A. (2015). Management Earnings Forecasts and Other Forward-Looking Statements. Online Article (<https://ssrn.com/abstract=2130145>, accessed February 2017).

Cambria, E. and Hussain, A., (2015). Sentic Computing. A Common-Sense Framework for Concept-Level Sentiment Analysis. Springer, Switzerland.

Cassidy, C. (2013). Between the hedges: A Computational analysis of sentiment and linguistic hedging in financial documents. MSC Thesis. Online Article ([https://www.ai.uga.edu/sites/default/files/theses/cassidy\\_caitlin\\_n\\_201505\\_ms.pdf](https://www.ai.uga.edu/sites/default/files/theses/cassidy_caitlin_n_201505_ms.pdf), accessed January 2017).

Chan, A.W.K. and Cong, M.W.C. (2016). Sentiment analysis in financial texts. *Decision Support Systems*. <http://dx.doi.org/10.1016/j.dd.2016.10.006>

Cleverley, P.H. and Burnett, S., (2015). Creating Sparks: Comparing Search Results Using Discriminatory Search Term Word Co-Occurrence to Facilitate Serendipity in the Enterprise. *Journal of Information and Knowledge Management*, (JIKM) 14(1), pp. 1-27.

Deloitte (2015). Deloitte UK Annual Report Insights. Online Article (<https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/audit/deloitte-uk-annual-report-insights-2015-full-survey.pdf>).

Dury, R., (2000). A brief Glossary of Modality. Online Article (<http://dinamico2.unibg.it/anglistica/slin/modgloss.htm>, accessed January 2017).

El-Haj, M., (2014). Analysing UK Annual Report Narratives using Text Analysis and Natural Language Processing. Online Article (<http://www.lancaster.ac.uk/staff/elhaj/docs/GlasgowTalk.pdf>, accessed January 2017).

EOI (2012). Modal Verbs. Online Article (<http://www.eoisabi.org/wp-content/uploads/2012/06/modality-modal-verbs.pdf>, accessed January 2017).

Faith, A. (2011). Linguistically Training Automatic Indexing Software for Complex Taxonomies. Semantic Technology & Business Conference June 14-18 2013, San Jose, CA, USA.

Fisher, I.E. et al., (2008). The Role of Text Analytics and Information Retrieval in the Accounting Domain. *Journal of Emerging Technologies in Accounting*, 7, pp. 1-24.

Grimes, S., (2010). Expert Analysis: Is Sentiment Analysis an 80% Solution? InformationWeek Online Article (<http://www.informationweek.com/software/information-management/expert-analysis-is-sentiment-analysis-an-80--solution/d/d-id/1087919>, accessed February 2017).

Grimes, S., (2012). From Sentiment Analysis to Enterprise Applications. Greenbook Online Article (<http://www.greenbookblog.org/2012/01/02/from-sentiment-analysis-to-enterprise-applications/>)

Gunning, R. (1969). The Fog Index After Twenty Years. *International Journal of Business Communication*, 6(2).

- Gupta, A and Liu, H., (2017). Addressing the Risk Culture Challenge in Banking using Text Analytics. European Financial Management Association (EFMA), Annual Meeting June 28-July 1 Athens, Greece.
- Hajek, P. and Olej, V., (2013). Evaluating Sentiment in Annual Reports for Financial Stress Prediction using Neural Networks and Support Vector Machines. EANN 2013, Part II, CCIS 384, pp. 1-10.
- Hedden, H. (2013). Taxonomies for Auto-Tagging Unstructured Content. Text Analytics World, October 1 2013, Boston USA.
- Hildebrandt, H.W. and Snyder, R.D. (1981). The Pollyanna Hypothesis in Business Writing: Initial Results, Suggestions for Research. *The Journal of Business Communication*, 18(1), pp. 5-15.
- Hinkel, E. (1995). The Use of Modal Verbs as a Reflection of Cultural Values. In *TESOL Quarterly*. Ed McKay et al. 29(2), pp. 100-118.
- Humpherys, S.L. et al., (2011). Identification of fraudulent financial statements using linguistic credibility analysis. *Decision Support Systems*, 50(3), pp. 585-594.
- Hykes, J.M. (2000). A comparison of the use of modal verbs in research articles by professionals and non-native speaking graduate students. MA Thesis Online Article (<http://lib.dr.iastate.edu/cgi/viewcontent.cgi?article=8928&context=rtd>, accessed January 2017).
- Jaime, A and Perez-Guillot, C. (2015). A comparison analysis of modal auxiliary verbs in Technical and General English. *Procedia – Social and Behavioural Sciences*, 212, pp. 292-297
- Jurka, T.P., Collingwood, L., Boydston, A.E., Grossman, E., van Atteveldt, W. (2013). RTextTools: A Supervisory Learning Package for Text Classification. *The R Journal*, 5(1), pp. 6-12.
- Kets De Vries, M., and Miller, D., (1987). "Interpreting Organizational Texts," *Journal Of Management Studies*, 24(3), pp. 233-247.
- Kloptchenko, A. et al., (2004). Mining Textual Contents of Financial Reports. *The International Journal of Digital Accounting Research*, 4(7), pp. 1-29.
- Kruschwitz, U. and Hull, C., (2017). "Searching the Enterprise", *Foundations and Trends in Information Retrieval*, 11(1), pp 1-142. <http://dx.doi.org/10.1561/15000000053>
- Levin, D. and Edwards, T. (2007). The leader coach: Exposing your soul. Authorhouse, USA.
- Lexalytics (2017). What is Sentiment Scoring? White Paper. Online Article (<https://www.lexalytics.com/content/whitepapers/Lexalytics-WP-Sentiment-Scoring.pdf>)
- Li, F., (2006). Do stock market investors understand the risk sentiment of corporate annual reports? Working paper. Online Article (<http://www.cis.upenn.edu/~mkearns/finread/sentiment.pdf>, accessed January 2017).
- Li, F., (2010). The Information Content of Forward-Looking Statements in Corporate Filings – A Naïve Bayesian Machine Learning Approach. *Journal of Accounting Research*, 48(5), pp. 1049-1102.

Loughran, T. and McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries and 10-K's. *The Journal of Finance*, 66(1), pp. 35-65.

Magnuson, D. (2014). Auto Classification and the Holy Grail for Records Managers. IBM Presentation as the Association of Records Managers and Administrators (ARMA), Houston, USA.

Malhotra, A. et al., (2013). 'HypothesisFinder:' A strategy for the Detection of Speculative Statements in Scientific Text. *PLoS Computational Biology*, 9(7). Doi [10.1371/journal.pcbi.1003117](https://doi.org/10.1371/journal.pcbi.1003117)

Manning, C.D., Raghavan, P. and Schütze, H. (2008). Introduction to Information Retrieval. Cambridge University Press.

Michel, J-B., (2011). Quantitative analysis of culture using millions of digitized books. *Science*, 331(6014), pp. 176-182.

Milbank (2009). Client Alert. Online Article (<https://www.milbank.com/images/content/6/3/634/012609-Changes-to-SEC-Form-20F-Reporting-Obligations.pdf>, accessed January 2009).

Millar, N. (2009). Modal verbs in TIME: Frequency changes 1923-2000. *International Journal of Corpus Linguistics*, 14(2), pp. 191-220.

Miller, D. (2014). Just the facts: Auto-classification and Taxonomies. ConceptSearching Webinar, Online Article (<https://www.conceptsearching.com/just-the-facts-auto-classification-and-taxonomies/>).

Minhas, S. and Hussain, A. (2016). From Spin to Swindle: Identifying Falsification in Financial Text. *Cognitive Computing*, 8, pp. 729-745.

Moore, D.S., Notz, W.I. and Flinger, M.A. (2013). The basic practice of statistics. 6th Ed. New York. W.H. Freeman and Company.

Moran, S. and Kral, R., (2013). Board of Directors and Audit Committee guide to Fiduciary responsibilities. AMACOM Division American Management Association.

Muslu, V. et al., (2012). Forward-Looking Disclosures and the Information Environment. *Management Science*, 61(5), pp. 931-948.

Narayanan, R., Liu, B. and Choudhary, A. (2009). Sentiment analysis of conditional sentences. Proceedings of the Conference on Empirical Methods in Natural Language Processing, August 6-7 2009: Singapore, 1(1), pp. 180-189.

Nathan, P.B. (2010). A Genre-based study of Pedagogical Business Case Reports. PhD Thesis Online Article (<http://etheses.bham.ac.uk/711/1/Nathan10PhD.pdf>, accessed January 2017).

Pang, B. and Lee, L., (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2), pp. 1-135.

Peet, R.K., (1974). The Measurement of Species Diversity. *Annual Review of Ecology and Systematics*, 5(1-463). Online Article

(<http://www.annualreviews.org/doi/abs/10.1146/annurev.es.05.110174.001441?journalCode=ecolsys.1>, accessed January 2017).

Piotti, S. (2014). Exploring Corporate Rhetoric in English: Hedging in Company Annual Reports: A corpus assisted analysis. EDUCatt: Milan, Italy.

Pique-Angordans, Posteguillo and Andreu-Beso (2002). Epistemic and Deontic Modality: A Linguistic Indicator or Disciplinary Variation in Academic English. *LSP & Professional Communication*, 2(2), pp. 49-65.

Pulman, S. (2014). Multi-Dimensional Sentiment Analysis. University of Oxford and TheySay. Online Article:

(<http://www.theysay.io/wp-content/uploads/2016/05/SGPulman-LT-Accelerate-Brussels2014.pdf>)

Reagan, A.J., Tivnan, B., Williams, J.R. et al., (2015). Benchmarking sentiment analysis methods for large-scale texts: a case study for using continuum-scored words and word shift graphs. Cornell University Library. Online Article <https://arxiv.org/abs/1512.00531>

Roose, M.O. and Sikström, S., (2014). Strategic change: a journey towards new meaning? Semantic analysis of corporate communication. *International Journal of Business Science and Applied Management*, 9(2), pp. 51-64.

Rutherford, B.A., (2005). *Journal of Business Communication*, 42(4), pp. 349-378.

Sasaki, Y. (2008). Automatic Text Classification. University of Manchester. Online Article (<http://www.nactem.ac.uk/dtc/DTC-Sasaki.pdf>).

Seeger, M.W., Sellnow, T.L. and Ulmer, R.R., (1998). Communication, Organization and Crisis. *Annals of the International Communication Association*, 21, pp. 231-276.

SentiWordNet (2010). Online Article (<http://sentiwordnet.isti.cnr.it/>, accessed January 2017).

Shiri, A.A., Revie, C.W. and Chowdhury, G., (2002). Thesaurus-assisted search term selection and query expansion: a review of user-centred studies. *Knowledge Organization*, 29(1), pp. 1-19.

Siegel, J.A., Saukko, P.J. and Houck, M.M. (2013). Encyclopaedia of Forensic Sciences, 2<sup>nd</sup> Edition. Elsevier, USA.

Smiraglia, R.P. and Van Den Heuvel, C., (2011) Idea Collider: From a Theory of Knowledge Organization to a Theory of Knowledge Interaction. *Bulletin of the American Society for Information Science and Technology*, April/May 2011, 37(4), pp. 43-47.

Taboada M., et al., (2011). Lexicon-based methods for sentiment analysis. *Computational Linguist.* 37, pp. 267–307.

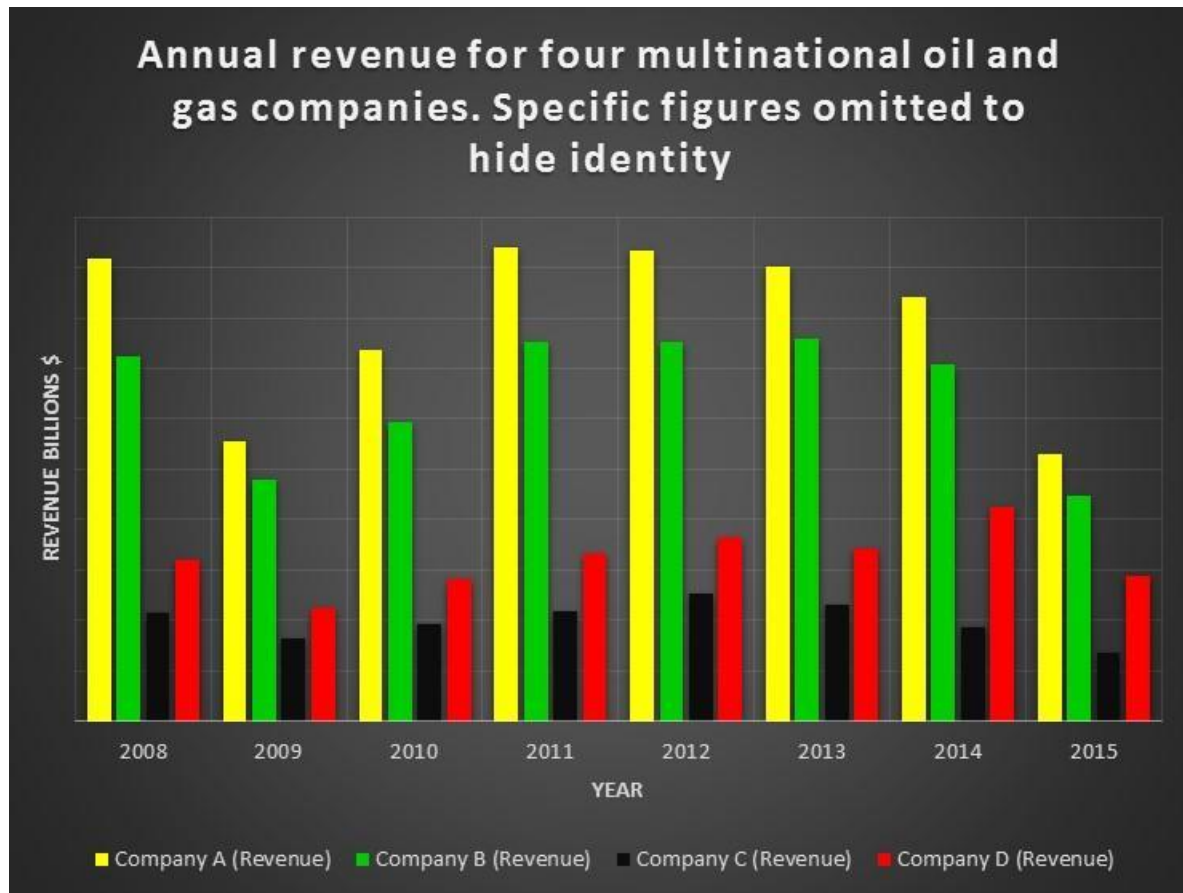
Taboada, M., (2015). Sentiment Analysis: An Overview from Linguistics. *Annual Review of Linguistics*. 2(1-425). Published online  
<http://www.annualreviews.org/doi/full/10.1146/annurev-linguistics-011415-040518>

Tausczik, Y.R. and Pennebaker, J.W. (2010). The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology*, 29(1), pp. 24-54.

- TeachIT (2016). Modal verbs and Adverbs. Online Article ([www.teachit.co.uk](http://www.teachit.co.uk), accessed January 2017)
- Tennis, J.T., (2009). Fringe Types and KOS Systematics: Examining the limits of the population perspective of knowledge organization systems. *Advanced in Classification Research Online*, 20(1).
- Tetlock, P.C., (2007). Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance*, 62, pp. 1139–1168.
- Thompson, P. et al (2008). Categorising Modality in Biomedical Texts. Online Article ([http://personalpages.manchester.ac.uk/staff/paul.thompson/papers/LREC\\_modality\\_2008\\_final.pdf](http://personalpages.manchester.ac.uk/staff/paul.thompson/papers/LREC_modality_2008_final.pdf) )
- Ulmer, R.R., Sellnow, T.L. and Seeger, M.W., (2011). *Effective Crisis Communication: Moving from Crisis to Opportunity*, 2<sup>nd</sup> Edition. Sage, USA.
- UNC Chapel Hill (2014). Online Article (<http://writingcenter.unc.edu/handouts/modals/>, accessed January 2017)
- Van Boeyen, S. (2014). Why Sentiment analysis engines need customization. Online Article (<http://www.techradar.com/news/software/business-software/why-sentiment-analysis-engines-need-customization-1256701>, accessed July 2017).
- Vold, E.T. (2006). Epistemic modality markers in research articles: a cross linguistic and cross-disciplinary study. *International Journal of Applied Linguistics*, 16(1), pp. 61-87.
- Wang, C-J et al., (2013). Financial Sentiment Analysis for Risk Prediction. International Joint Conference on Natural Language Processing, Nagoya, Japan 14-18 October 2013, pp. 802-808.
- Wilson, T., Wiebe, J. and Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing, Oct 6-8 2005, Vancouver, Canada, pp. 347-354.
- WordNet (2010). Princeton University. Online Article (<http://wordnet.princeton.edu>, accessed January 2017).
- Yasumasa, S., (2008). Modal Verbs and Their Semantic Functions in Business English. *Aoyama Journal of Business* (Aoyama Keiei Ronshu), 44(3).
- Yu, K., Zhang, J., Chen, M., Xu, X., Suzuki, A., Ilic, K., Tong, W. (2014). Mining hidden knowledge for drug safety assessment: topic modelling of LiverTox as a case study. *BMC Bioinformatics*, 15.
- Yuet-yung, S., (2014). Forward-looking statements in annual reports: how is futurity expressed? MA Thesis (<https://hub.hku.hk/bitstream/10722/207135/1/FullText.pdf>).
- Zeng, M. et al., (2007). Knowledge organization systems (KOS) standards. Proceedings of the Association for Information Science and Technology, 44(1), pp. 1-3.

## **APPENDIX I**

Trends in revenue for Company A, B, C and D.





## APPENDIX II

Average oil price (U.S. Dollars per Barrel) 2008-2016 (Staistica 2017)

<https://www.statista.com/statistics/262858/change-in-opec-crude-oil-prices-since-1960/>

